Describing Health and Medical Costs, and the Economic Evaluation of Health Care: applications in injuries and cervical cancer

Willem Jan Meerding

Voor Margreet, Rinko, en Loek

Describing Health and Medical Costs, and the Economic Evaluation of Health Care: applications in injuries and cervical cancer

Het beschrijven van gezondheid en medische kosten, en de economische evaluatie van gezondheidszorg: toepassingen bij ongevalsletsels en baarmoederhalskanker

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus Prof.dr. S.W.J. Lamberts
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 29 september 2004 om 15.45 uur

door

Willem Johannes Meerding

geboren te Gouda

Promotiecommissie

Promotoren: Prof.dr.ir. J.D.F. Habbema
                    Prof.dr. F.F.H. Rutten

Overige leden: Prof.dr. G.J. Bonsel
                    Prof.dr. L. Leenen
                    Prof.dr. Th.J.M. Helmerhorst

Copromotor: Dr. E.F. van Beeck

Table of contents

Introduction 1

1.1 Scope of the thesis

Since the late nineteenth century, population's health has improved tremendously, at least in wealthy nations. Life expectancy at birth in the Netherlands has increased at an unprecedented pace from less than 40 years in 1860 to currently 76 years for males and 81 years for females. It is widely accepted that better hygiene, housing and nutrition have contributed to this increase through a reduction in communicable diseases, rather than medical care. The small overlap between the period with the major declines in mortality and the period of major advances in medicine suggests that the contribution of medical care to population health is at least limited, and has even been doubted [170]. However, it is now commonly believed that many preventive and curative interventions have contributed to improved health [153]. Diseases that were killers in the past can now be cured or prevented.

   The improvement of population health comes at a price. Particularly from the 1970s onwards, health care spending has increased fast, and currently accounts for about 10% of the gross national product. The increase in health care spending is, apart from wage and price increases, generally attributed to two phenomena: the development of medical technology and, more recently, the ageing of society. It is too commonplace to claim that medical innovation is cost increasing in general. Some innovations will increase costs, whereas others will be cost saving. In fact, the economic consequences of technologies can only be assessed for specific technologies related to specific indications [91]. For example, preventive interventions such as cervical and breast cancer screening are effectively reducing cancer incidence but costs have shown to outweigh savings [270 280]. In contrast, treatment of stroke patients in specialized stroke units results in more favourable health outcomes and cuts costs compared to treatment in general wards [122]. However, there are strong indications that on an aggregate level medical technology leads to increased health care spending [219].

   An objective of health care policy is to provide health care in such manner that population's health is maximized with an efficient use of resources. The current situation has put pressure upon policy makers to contain health care costs, and to restrict access to health care interventions that are evidence-

based, with proven effectiveness, and with acceptable cost-effectiveness. However, health (care) policies that maximize the overall health benefit at contained costs can only be successful when data are available on a population's health and its determinants, on health care needs, and on the effectiveness and efficiency of existing and newly developed medical technology.

In this thesis we present these data for two quite different health problems: injury and cervical cancer. For injury, population based estimates are presented of health care costs and post-injury functional outcome, subdivided by injury diagnoses. We also investigated determinants of injury related health care costs and functional outcome. For cervical cancer, we evaluated several options to improve the efficiency of population based screening and follow-up. The thesis starts with a chapter on the distribution of total health care costs by diseases and injuries, and by basic demographic indicators.

1.2 Injuries

Acute physical injuries account for 12% of the burden of disease in established market economies, and for even higher shares in other global regions [194]. It is known as a heterogeneous health problem, ranging from high frequency, minor injuries (e.g. superficial injury) to low frequency, major injuries (e.g. polytrauma patients). A general distinction is between unintentional injury – home and leisure, occupational and traffic accidents – and intentional injury – violence, self-inflicted injury.

The consequences of injuries can be estimated by the resulting health care utilization, with costs as a single index. Previous estimates of the cost of injury in the Netherlands were made for broadly defined accident categories, and include medical costs and production loss due to work absence, permanent disability and death, also known as 'indirect costs' [275]. About 60% of medical costs were attributed to home injuries, with high costs among elderly females. Traffic injury accounted for another one-fifth. So far, international studies have been occasional and fragmented, and have mainly focussed on per patient medical costs and its determinants. For instance, high medical costs per patient have been reported for lower extremity injury including hip fracture [10 159 305] and for head injury [167].

However, considerable knowledge gaps remain. First, more detailed estimates are needed on the health care costs of specific injuries, because injury control measures and trauma care are often specific for narrowly defined injury groups. Examples are poisonings in children, or severe traumatic brain injury in motor cyclists. Second, cost estimates should preferrably be comprehensive and encompass all injuries to make comparisons among injury groups, identify

previously unrecognized injury groups, and put other injuries into perspective. Third, injuries are a dynamic public health area. History has shown that exposure to newly emerging risks leads to an increasing incidence in specific injury categories. As a result, there is a need for detailed monitoring of injury related health care demands together with incidence and mortality. The first research question will therefore be:

1. *How are medical costs of injury at national level distributed by type of injury and health care sector, and what are their major determinants?*

Apart from data on health care consumption and costs, epidemiological indicators are very important. Since the beginning of the 1970s the overall injury mortality rate has shown a considerable decline (crude and age-adjusted), following an increase in previous decades. The decline in mortality is partly due to a decreasing incidence in specific areas, e.g. traffic injury, and partly due to improved survival rates (all injury categories), reflecting the success of several preventive interventions, such as mandatory helmet use, and improved trauma care, respectively [274].

The decline in mortality has contributed to a growing attention for the disability component in the burden of injury. In the Netherlands little quantitative information on injury related disability is available. Estimates of the burden of injury by broadly defined types of injury were made for the Population Health Forecasts in 1997, with traffic injury and suicide ranking highest in terms of disability adjusted life years lost (DALY) [232]. However, these estimates were based on expert guesses of the prevalence and severity of permanent disability.

Since then, empirical data have been collected on the functional outcome of major trauma [283 291], hip fractures [13] and tibial fractures [112]. So far, international studies have concentrated on functional limitations in patients with high energy injuries that require hospitalization, including polytrauma patients [5 109 123 154 159 289]. Others considered specific major injuries, such as vertebral fractures [11 148], pelvic ring fractures [208 278], tibial shaft fractures [97] and ankle fractures [224]. Nevertheless, there are indications that a considerable share of total disability is attributable to patients that have never been hospitalized [169 217 292]. Because studies in these patient groups hardly exist, research on functional outcome is urgently needed in comprehensive populations of minor and major trauma patients.

Typical for injuries are their heterogeneous functional sequelae and recovery patterns. Injuries can struck any body region, and multiple mechanisms (fall, fire, chemical substance, etcetera) lead to evenly multiple

types of injury (fracture, strains, burns, etcetera). A uniform and systematic measurement of functional outcome can therefore make an important contribution to the comparison of functional outcome among different injury diagnoses. The second research question is therefore:

2. *How is injury related disability at national level distributed by type of injury, and what are its major determinants?*

Although the description of the functional and economic consequences of injury are two separate scientific fields, it is likely that they are related. Total medical costs are by definition incidence multiplied with cost per patient, whereas cost per patient will in many cases be related to the level and duration of disability, apart from several socio-demographic characteristics that are known as determinants of health service use (e.g. age, sex, living alone, socio-economic status) [205 281].

1.3 Cervical cancer

Cervical cancer is the second most common cancer among women around the world, but the incidence and mortality varies among global regions. In developed countries the incidence is lower. In the Netherlands the chance that a woman gets cervical cancer during her life is about 1-2% [290]. Yearly about 700 women develop cervical cancer, of whom about 235 die of the disease. The low incidence of invasive cancers is partly due to population based screening. In the Netherlands Pap smear screening was introduced in the 1960s. For many years women were screened on their own demand. This socalled "spontaneous" screening practice led to irregular screening intervals whereas many women were not screened at all. Organized screening, by which women in the target age range are invited according to a fixed schedule, has taken place from 1988 onwards in some regions, and nationwide from 1995 onwards.

 The impact of cervical cancer screening on mortality reduction has never been determined in a controlled experiment. By the time the effectiveness was doubted, screening had already been widely disseminated. This made an assessment in a controlled experiment impossible. But in the past decades convincing indirect observational evidence has been collected primarily through the analysis of screening data, supported by advanced modelling techniques [140 267]. By now the evidence that cervical cancer screening is effective is so strong, that a controlled experiment is deemed unethical.

 There are several reasons for the remaining cervical cancer incidence and mortality despite widespread screening. Invasive cancers can still be found in age groups uncovered by screening. In the age group that is invited for

screening (30 to 60 years in the Netherlands) invasive cancers may be due to non-participation (about 60%) or insufficient screening (about 10%) [30]. About 30% are 'interval cancers' detected inbetween two subsequent screening rounds, which might partly be missed in the previous screening round due to a false negative smear.

For decades the Pap smear has been the screening test. The technique works as follows: cell material is scraped from the cervix uteri, and subsequently stained on a glass slide, which can be assessed microscopically. Depending on the degree of abnormality of the screening smear, women will follow the screening schedule, will have a repeat smear, or will be referred to the gynecologist for further diagnosis and treatment.

Ever since its introduction, there have been worries about the accuracy of the Pap smear [6 135 197]. In systematic reviews its sensitivity has been estimated at about 50-80% [82 197]. These estimates depend on the cut-off value for smear abnormality, the study population, the reference test used, and other characteristics of study design. Also, it is essential to distinguish the sensitivity of a single Pap smear from the 'programme sensitivity' in organized screening. Due to the long duration of the pre-invasive stages, estimated at 16 years on average [287], missed cases will have a considerable chance to be detected at subsequent screening rounds. Nevertheless, a low sensitivity of the Pap smear compromises its ability to detect cervical lesions at an early stage. Particularly in the US, the debate on Pap smear accuracy is enforced by lawsuits of women with invasive cancer, whose diagnoses have probably been missed in the pathologic laboratory [236]. In this respect the high workloads in commercial laboratories have been mentioned as a possible cause.

In recent years several cytologic technologies have been developed that claim to faciltate the screening process and reduce the number of missed cases. Some of these are based on thinlayer technology that makes the smear easier to interpret (ThinPrep™, SurePath™). Others are based on automation of the screening, by which slides can be triaged into low and high risk slides. This accelerates the throughput of screening smears, most of which can be assessed as normal, whereas the high risk slides can be assessed with increased alertness. In some countries these technologies are already applied in routine practice. In the Netherlands some laboratories have already converted, despite clinical guidelines that prescribe the use of the conventional Pap smear. The third research question will therefore be:

3. *What are the test characteristics of newly developed cytologic technologies for cervical cancer screening, and how (cost-)effective are these technologies compared to screening with Pap smears?*

Cervical cancer screening is effective in preventing invasive cancers and mortality, but it has unfavourable side-effects as well. One of the main unfavourable effects is that currently about 3% of screened women have an abnormal smear, of whom only a small part would develop invasive cancer in the absence of screening. In some countries the proportion of abnormals may be as high as 10%, as in the Netherlands until recently. These women will get follow-up by a repeat smear or will be referred for colposcopy and treated if necessary, whereas only few of them will actually benefit. A considerable part of pre-invasive stages of cervical cancer will spontaneously regress to normal, and would never have been detected without screening. For the women concerned follow-up implies uncertainty and discomfort, which may last for more than a year (according to the recent guidelines).

These unfavourable side effects of screening might be reduced by testing women with an abnormal smear for the presence of human papillomavirus (HPV). HPV testing can be used as a diagnostic tool to triage women for further management [52]. Since the 1980s there is growing molecular and epidemiological evidence that some HPV types act as causal agents for the development of cervical cancer [137 190]. These high-risk types (hr-HPV) have been detected in over 95% of invasive cancers, and in 50-80% of the pre-invasive stage cervical intraepithelial neoplasia (CIN). Prospective studies have shown that progression of pre-invasive stages was only observed in case of 'persistent HPV', in women that tested positive for hr-HPV in the screening smear and repeat smears [204]. In women without hr-HPV, progression does not occur. However, regression is also observed in women with hr-HPV [202]. It therefore remains to be determined whether HPV testing is a valuable diagnostic tool in women with abnormal smears, with the aim to reduce anxiety and discomfort induced by screening. The fourth research question will therefore be:

4. *Can the follow-up of women with abnormal Pap smears be made more efficient by human papillomavirus testing?*

HPV testing may also be applied as a primary screening tool and as a surveillance instrument in women that have been treated for CIN. These possible applications have been evaluated elsewhere, and will not be further investigated in this thesis [203 267].

Figure 1.1
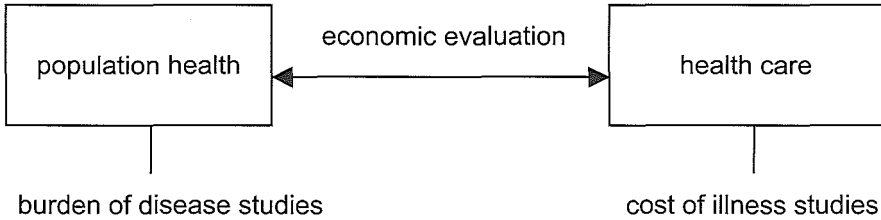
1.4 Analytical tools

For the analysis of the questions posed so far, we used three existing analytical tools that are known as burden of disease (BOD) studies, cost of illness (COI) studies and economic evaluation studies. The relationship between these tools and the two principal entities they relate to, population health and health care, are presented in figure 1.1. BOD and COI studies provide a comprehensive and coherent description of population health and health care, respectively, and of their distribution across diseases and injuries, risk factors, and other population variables. They may be used for comparative purposes to identify differences in health and costs (e.g. between groups) or changes over time. Summary measures play a key function in BOD studies. Summary measures of population health combine information on mortality and non-fatal health outcomes to represent the health of a particular population, using time as the common numerator [88]. Examples are the quality adjusted life year (QALY) and disability adjusted life year (DALY) [194 261]. In COI studies, costs can be regarded as a summary measure of health care. BOD and COI studies are complementary in measuring the societal burden of disease and injury, and they could be usefully combined to translate expected dynamics in population health into future health care needs. In contrast with BOD and COI studies, economic evaluations focus on the dynamic relationship between health care and population health, and are a tool to assess the implications of a change in health care for population health.

Burden of disease studies

Population health can be distinguished into mortality and disability (non-fatal health outcomes). Frequently used indicators of mortality are (standardized) mortality rates, the number of life years lost, survival rates, and life expectancy. Disability is a more complex, multidimensional concept, similar to its positive counterpart: health. A conceptual framework for measuring disability is provided by the International Classification of Functioning, Disability and Health (ICF), which distinguishes between functioning and contextual factors

[297]. Functioning encompasses the level of body functions, activities and participation in life situations. The impact of disability may be determined by contextual factors, the social and physical environment, in addition to personal factors, specifically in terms of limitations in social participation such as the ability to engage in work or social activities.

The data requirements of BOD studies are large, particularly for the construction of summary measures, and encompass disease-specific epidemiological frequency data (incidence, prevalence, mortality), data on duration of disease, and disease-specific health status valuations. Complete and consistent epidemiological frequency data are often not readily available, but are a prerequisite for collecting data on disease severity. Data on disease severity are of no use when frequency data with which they must be combined are incomplete or inconsistent [78]. Some of the problems with epidemiological frequency data may be solved with modelling [139].

Severity of disease can be measured with specific instruments for health status measurement. Many instruments (read: questionnaires) have been developed. A general distinction is between generic instruments and disease- or domain-specific instruments. Generic instruments include items on all three domains of health: physical, mental and social functioning. These instruments are applicable to all diagnoses, and are therefore useful to make comparisons among diseases or injuries that may be quite different. Disease-specific instrument have a more detailed focus on functional consequences that are specific for certain diseases or injuries. As a result they are more sensitive to specific changes in health, but cannot be used to make comparisons across very different diagnoses. Domain-specific instruments focus on specific types of functioning, e.g. pain or depression.

Despite their large differences, all instruments share common characteristics. They capture a number of health items (e.g. mobility, pain) that can be associated with specific body functions, activities and social roles, and each item can be scored in several response categories. Instruments may differ in the items that they include and in the level of detail in response categories. An instrument with many items and many response categories may well be able to discriminate among health states, but this should be weighed against their complexity.

A general shortcoming of health measurement instruments is that they do not enable a judgement which health status is to be preferred, in case one health status scores worse on one health domain but better on another domain compared to another health status. This requires value judgements on the relative severity of health states. These valuations (scores, weights) can then be combined with information on the duration of functional sequelae and with

other epidemiological frequency data to arrive at a single metric of population health.

Health status valuations can be derived in two ways. The first is by describing the functional consequences of all possible diseases and injuries and their different disease stages, and subsequent valuation of these descriptions by one or several valuation techniques. Considering the wide spectrum of possible diagnoses, even within the field of injury, this would at least be time consuming. A second, more efficient approach is by describing disease stages using a generic measurement instrument, and subsequently converting these descriptions into valuations by using existing algorithms. For instance, such algorithms are available for the EuroQol and SF-36 generic instruments, based on statistical modelling of empirical valuations of a set of key health state descriptions [34 65]. This second approach has been followed in this thesis for describing non-fatal health outcomes in injury patients. Apart from being more efficient, an advantage compared to the first approach is that it provides information on the prevalence of restrictions on specific underlying health domains, which facilitates the interpretation of health status valuations and of metrics that summarize non-fatal health outcomes.

Cost of illness studies
Health care can be quantitatively described by numbers of inputs, such as labour and equipment, and by outputs, such as hospital bed days. Similar to summary measures of population health, costs can be regarded a summary measure of health care. In contrast to economic evaluations, COI studies provide a cross-sectional description of costs by diseases and injury. Because of their cross-sectional design, they do not provide insight into the relationship between changes in input (health care) and output (health), and statements on the efficiency of health care are therefore not possible.

There is a considerable variety in COI studies [219]. They may be restricted to health care costs, or also include costs to patients and other economic sectors (e.g. productivity losses). They may be disease-specific, describing the costs of specific diseases, or generic, giving a comprehensive overview of costs of the entire spectrum of diseases. A related distinction is between top-down and bottom-up studies. Total (health care) costs in a specified period can be attributed top-down to specific diseases, injuries and risk factors. This has the advantage that a uniform methodology is applied for all diseases, and facilitates the comparison of costs. Each euro can be attributed only once and double counting of costs is therefore avoided. In a bottom-up approach, often used in disease-specific studies, lifetime health care consumption (and other costs) of individual patients are aggregated into a total

estimate. Disease-specific COI studies may use detailed data sets that are not available for generic COI studies, and therefore may give more detailed and reliable results. In theory, and assuming a stable population with no changes in demography or epidemiology, both approaches should produce the same results. Differences often occur because of differences in how costs are attributed in patients with two ore more conditions (comorbidity).

In this thesis generic and disease-specific COI studies have been conducted. Total health care costs by diseases and injuries were described using a top-down approach, whereas health care costs of injury were estimated bottom-up.

Economic evaluation studies
In contrast to COI studies, that provide a static description of the relationship between population health and health care, economic evaluation studies provide insight into the (potential) changes in costs and population health as a result of a particular intervention. In other words, they evaluate the impact of an intervention in terms of monetary costs and savings, and positive and negative health effects. These studies are a powerful tool for health care rationing. A more efficient use of resources results when priority is given to those interventions with the most favourable balance between costs and health effects ('cost-effectiveness ratio'). The findings of economic evaluations are often implemented in clinical practice [282]. However, economic evaluations are not yet systematically embedded in decisions concerning the financing and implementation of health care technologies [239].

Different types of economic evaluation exist, namely cost-effectiveness, cost-utility, and cost-benefit analysis, depending on whether the effects are counted in natural units (e.g. life years lost), quality of life (valuations or utilities), or monetary units, respectively. Evaluations of interventions without positive or negative health effects are cost-minimization analyses. Because most health interventions have multiple health effects, there has been an increasing tendency to translate outcomes into summary measures such as quality adjusted life years (QALYs). These summary measures facilitate the comparison of different interventions in different areas of population health, but should not draw too much attention at the expense of important differences on underlying health components. In practice, evaluative studies that use utilities as outcome measure are often called cost-effectiveness analyses. Cost-benefit analyses (CBA) are more strongly rooted into welfare economic theory, but their application has not been widespread in the evaluation of health care, and their share has also decreased over time [75]. This may be partly because of the empirical difficulties that are experienced in deriving monetary values of health

benefits through willingness to pay methods (WTP), and because of equity concerns [209]. In this thesis, several interventions for the optimization of cervical cancer screening have been evaluated with application of cost-effectiveness analysis.

Depending on the decision authority, economic evaluations may be conducted from different perspectives. Most economic evaluations support national policy decisions and use the societal perspective as a general rule. This implies that costs and benefits of all societal parties are accounted for, including health care resources, costs to patients (out of pocket expenses, travel costs), and costs to other economic sectors, including production losses due to work absence, permanent disability, and premature death. In general, non-medical costs are more difficult to measure and also their valuation may be contentious. This applies for example to time costs of informal care and production losses [38 131]. Other perspectives from which an evaluation is conducted are a health care provider perspective or company perspective. An economic evaluation then includes only those components that are relevant from the perspective of the decision maker.

In summary, the three analytical tools as described here – BOD, COI, and economic evaluation studies – are strongly related to each other and are complementary with respect to informing health care policy and planning. Nevertheless, questions remain about their relative contribution, of which this thesis presents some examples. This brings us to the fifth and last research question:

5. *To what extent do burden of disease studies, cost of illness studies and economic evaluation studies provide helpful information for the prioritization of health care?*

1.5   Reading guidance

The research questions are dealt with in this thesis as follows: chapter 2 presents a generic COI study for the Netherlands, based on health care expenditures in 1994. After this broad picture of where all the money goes, we focus on injuries in part I, and start with a detailed, bottom-up cost of injury study in chapter 3 (question 1). The basis is formed by a costing model linked to an injury surveillance system. Particular attention is paid to the distribution by injury diagnosis, demography and type of health care. We continue in chapter 4 with a review of existing cost of injury studies, in order to determine whether the picture that is drawn by cost of injury studies is sufficiently transparent to be useful for health care policy and planning. In chapter 5 we shift the focus from costs to functional outcome after trauma in a comprehensive population of injury patients, including minor and major trauma. This chapter describes how

levels of functioning differ by type of injury, and how functioning is determined by socio-demographic factors and injury severity (question 2).

In part II two major evaluative studies of cervical cancer screening are described. In chapters 6 and 7 the costs and cost-effectiveness of newly developed cytologic tests are determined (question 3). Test sensitivity and specificity, and costs per test are the key dimensions in a decision analytic framework that is designed with use of a microsimulation model. Current evidence on test sensitivity and specificity is confronted with this framework, to judge whether these tests can be expected to provide enough 'value for money'. In chapter 8 we evaluate the possible role of HPV testing in the triage of women with abnormal smears, with help of a decision analysis (question 4).

In chapter 9 the main conclusions are summarized and discussed, including the relative contribution of BOD and COI studies and of economic evaluation to the prioritization of health care is discussed (question 5). We conclude with a concise answer on each of the research questions and define the possible implications for further research and health care policy.

Abstract

Objectives The debate on cost containment in health care mainly concentrates on the supply side. The objective of this study is to present data on the demand side: the epidemiologic and demographic causes of health care use.

Design Information on health care use was obtained from all (22) health care sectors of the Netherlands. Most important sectors (hospitals, nursing homes, inpatient psychiatric care, institutions for the mentally retarded) have registries with nation-wide coverage. Total expenditures in a sector are subdivided into 21 age groups, sex, and 34 diagnostic groups.

Results After the first year of life, costs per person drop to their lowest levels in youth. They rise slowly throughout adult life, and increase exponentially from age 50 onwards till the oldest age group (95+). The top 5 causes of health care costs are mental retardation, musculoskeletal disease (predominantly joint disease and dorsopathy), dementia, a heterogeneous group of other mental disorders, and ill-defined conditions. Stroke, all cancers combined, and coronary heart disease, the main causes of death, rank 7, 8 and 10, respectively.

Conclusions The main determinants of health care use in the Netherlands are old age and disabling conditions, particularly mental disability. A large share of the health care budget is spent on long-term nursing care, which will inevitably increase further in an ageing population. Aspecific cost containment measures may endanger the quality of care of the old and mentally disabled.

Demographic and epidemiological determinants of health care costs in Netherlands: cost of illness study

2.1 Introduction

The debate on cost containment in health care is mainly focused on the supply side and the financing of health care [1]. Changes in population health status as another important determinant of costs play a minor role in the discussion. One reason is that the relation between diseases and costs is not straightforward, and relevant data are often lacking. This study connects the supply and demand side by subdividing total health care costs by health care sector, diagnosis, age and sex. Analyzing the Dutch health care budget offers good opportunities for this purpose; the country is small, more than 99% of its population has full health insurance coverage and, because of a long-standing administrative tradition, most health care sectors have excellent registries, of which the most important are nation-wide. The completeness of available Dutch health care data allows for a comprehensive description of epidemiologic and demographic determinants of health care costs. This means that not only the acute care sectors are represented, but also those sectors which deliver long-term care to the disabled. These are rarely included in other studies [9 150 186 201], which as a consequence underestimate the high costs of disabling disease.

2.2 Methods

Table 2.1 gives an overview of the health care costs in 1994 for each health care sector as presented annually by the Ministry of Health [182]. Additional personal expenditures, such as over-the-counter medication and spectacles (6% of all costs) are not included.

For the purpose of this article, the diagnoses of the International Classification of Diseases (ICD, 9th revision) [296] were clustered into 34 diagnostic groups, which can be regrouped into the 17 chapters of the ICD (see table 2.2). We defined groups of diagnoses in order to minimise misclassification between diagnostic groups and in order that each group be large enough to efficiently describe a sufficiently large proportion of health care costs. Conditions that could not be related to a specific diagnostic group but that are unambiguously related to a specific functional system (cardiovascular, respiratory, mental, etc.) were assigned to the remainder group of that specific

ICD chapter. Ill-defined conditions which could not be related to a specific ICD chapter were classified as 'Symptoms and ill-defined conditions' (ICD chapter 16). Particularly in primary health care this is a relevant category, as patients present with problems, not diagnoses. To avoid double counting, we have considered only primary diagnoses. Of all health care costs 8.1% could not be allocated to any diagnostic group because of insufficient information from some smaller health care sectors. Of all health care costs 5.3% are due to health care administration, and are not related to specific health problems. Together with the living costs in homes for the elderly, these latter costs were assigned to aspecific health care costs.

Table 2.1 Percentage of health care budget spent on different sectors of health care in Netherlands, 1994.

Health care sector	% of total*
Hospital care	32.1
Nursing homes	8.9
Old people's homes	
- medical costs	3.7
- living costs	5.4
Psychiatric care	7.1
Institutions for mentally and physically disabled people	8.6
Primary medical and paramedical services (excluding dental care)	5.7
Dental care	4.0
Pharmaceutical care	8.8
Home care and other small sectors	10.4
Health care administration	5.3

* Health care spending in 1994 was 59.5 billion guilders ($32.7 bn, £21.3 bn), 9.7% of gross national product.

For each health care sector, we identified key variables that are representative for health care use in that sector, such as days of stay for nursing costs in hospitals and nursing homes, or outpatient visits for costs of outpatient hospital care. For a specific sector the distribution of costs by 2 sexes, 21 age groups (0, 1-4, 5-9, 10-14,....., 95+ years) and 34 diagnostic clusters in 1428 (2 x 21 x 34) cells is considered equal to the distribution of the key variable for that sector. Thus, for each health care sector, costs for each combination of age, sex and diagnostic group are equal to the fraction of the key variable in that cell times the total costs for that sector. The summation of these cells over all health care sectors yields the data presented here.

The probability distribution of key variables was derived from sector-specific registries and sample surveys. Detailed information about the registries and the key variables used is available elsewhere [219].

Table 2.2 Diagnostic groups used in study and corresponding ICD 9 code [296].

ICD chapter		Diagnostic group	ICD codes
I	Infectious and parasitic diseases	Infection	1-139
II	Neoplasms	Cancer	140-208
		Benign neoplasms	210-239
III	Endocrine, metabolic and nutritional diseases	Diabetes	250
		Other endocrine diseases	240-279
IV	Blood and blood-forming organs	Blood diseases	280-289
V	Mental disorders	Dementia	290
		Schizophrenia	295
		Depression/anxiety	296, 300
		Alcohol/drugs	291-292, 303-305
		Mental retardation, Down's syndrome *	317-319, 758.0 *
		Other mental disorders	remainder 290-316
VIa	Nervous system	Neurologic disorders	320-359
VIb	Sense organs	Eye disorders	360-379
		Ear disorders	380-389
VII	Circulatory system	Hypertension	401-405
		Coronary heart diseases	410-414
		Heart failure	428-429
		Stroke	430-438
		Other circulatory diseases	remainder 390-459
VIII	Respiratory system	Asthma & COPD	490-496
		Other respiratory diseases	460-489, 497-519
IX	Digestive system	Dental diseases	520-529
		Gastro-intestinal diseases	531-569
		Liver, gall, pancreas diseases	570-579
Xa	Urinary system	Urinary disorders	580-599
Xb	Genital organs	Genital disorders	600-629
XI	Pregnancy & childbirth †	Pregnancy †	630-676
XII	Skin diseases	Skin diseases	680-709
XIII	Musculoskeletal system	Musculoskeletal diseases	710-739
XIV/XV	Perinatal/congenital conditions	Perinatal/congenital conditions	740-779
XVI	Symptoms, signs and ill-defined conditions	Ill-defined conditions	780-899
XVII	Accidents	Falls	E880-888
		Other accidents	E800-879, E890-999
		Not allocated	
		Non-specific ‡	

* Down's syndrome is classified in ICD chapter XV, code 758.0.
† Hospital costs of healthy babies (boys and girls) after childbirth were assigned to pregnancy and childbirth (women).
‡ Costs of health care administration and living costs in homes for the elderly.

2.3 Results

Total health care costs, representing 9.7% of the Dutch gross national product, were $2,124 per capita in 1994, $2,481 for women and $1,760 for men. The distribution is strongly age-dependent (figure 2.1). Costs are relatively high in the first year of life, reflecting the high costs of perinatal and infant care, but

than drop to the lowest levels in youth. During adulthood costs increase slowly and after age 50 they start to increase exponentially up to the highest age group (95+). The higher share in total costs of women (59%) is predominantly caused by their longer life expectancy, the higher prevalence of women in nursing homes and homes for the elderly, and the high costs of reproduction (including contraconception and diseases of the genital organs).

Figure 2.1 Total and per capita health care costs by age and sex in the Netherlands, 1994. Long-term care includes nursing homes, old people's homes, institutional care for disabled people, and appliances to assist disabled people. In 1994 $1=Dfl 1.82.
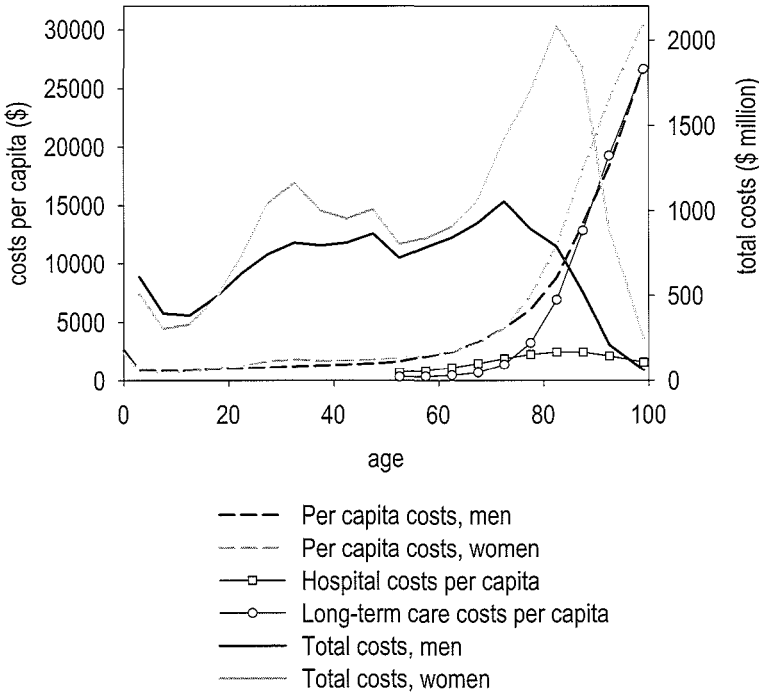
Tables 2.3 and 2.4 show the share in total costs of diagnostic groups by sex (table 2.3) and by age (table 2.4). Table 2.3 shows the high proportion of health care costs caused by mental disorders. Mental retardation ranks 1, dementia ranks 3, depression and anxiety ranks 15, schizophrenia 23, alcohol and drug abuse 31, and the heterogeneous remainder group of mental disorders ranks 4. All mental disorders together cover 28.4% of the health care budget that could be allocated to diagnostic groups. Ill-defined conditions, covering among others many psychosomatic problems, rank 5. Musculoskeletal diseases (predominantly all types of arthritis) rank 2. Dental diseases (predominantly

dentists' costs) rank 6. The main causes of death, i.e. stroke, all cancers combined, and coronary heart disease, rank 7, 8 and 10, respectively. Among women, costs of reproduction rank 6.

Table 2.3 Health care costs by diagnostic group and sex, the Netherlands 1994, ranked by share (in % of total health care costs).

Rank	Diagnostic group*	Men	Women	Total
1	Mental retardation, Down's syndrome	11.0	6.0	8.1
2	Musculoskeletal diseases	5.4	6.4	6.0
3	Dementia	2.9	7.4	5.6
4	Other mental disorders	5.4	4.7	5.0
5	Ill-defined conditions	4.6	5.0	4.8
6	Dental diseases	4.9	3.8	4.2
7	Stroke	3.0	3.4	3.2
8	Cancers	3.7	2.8	3.2
9	Pregnancy	0.0	4.3	2.6
10	Coronary heart diseases	3.9	1.5	2.5
11	Neurologic disorders	2.6	2.3	2.4
12	Other circulatory diseases	2.8	2.1	2.4
13	Other respiratory diseases	2.9	1.9	2.3
14	Other accidents	2.8	1.9	2.3
15	Depression and anxiety	1.8	2.6	2.3
16	Falls	1.3	2.4	2.0
17	Gastro-intestinal diseases	2.4	1.6	1.9
18	Asthma & COPD	2.4	1.2	1.7
19	Eye disorders	1.7	1.7	1.7
20	Liver, gall, and pancreas diseases	1.7	1.6	1.7
21	Skin dis4eases	1.7	1.6	1.6
22	Genital disorders	0.9	1.9	1.5
23	Schizophrenia	2.1	1.0	1.4
24	Urinary disorders	1.3	1.3	1.3
25	Infections	1.5	1.2	1.3
26	Hypertension	1.3	1.3	1.3
27	Diabetes	1.1	1.4	1.2
28	Ear disorders	1.4	0.9	1.1
29	Heart failure	1.1	1.1	1.1
30	Perinatal / congenital conditions	1.4	0.9	1.1
31	Alcohol/drugs	1.4	0.4	0.8
32	Benign neoplasms	0.5	0.9	0.7
33	Other endocrine diseases	0.4	0.8	0.6
34	Blood diseases	0.3	0.3	0.3
	Not allocated	7.2	8.8	8.1
	Non-specific†	9.1	11.7	10.7
Share in total costs (%)		41.0	59.0	100.0

* For ICD codes of all diagnostic groups, see table 2.2.
† Costs of health care administration and living costs in homes for the elderly

Table 2.4 shows the top 15 diagnostic categories for 5 age groups. In all age groups either mental retardation or dementia is the leading cause of health care costs. In youth, cognitive disability ranks second but congenital diseases also cover many mental disabling conditions. Among younger adults (age 15-44) the heterogeneous remainder group of mental disorders is second. Schizophrenia, depression, and alcohol and drug-related problems all rank among the top 15. Musculoskeletal diseases rank among the top 5 in all age groups after age 14, and ill-defined conditions rank among the top 6 in all age groups. Among the oldest age group (85+) stroke is second and accidental falls (predominantly hip fractures) is third. All cancers reach the top 5 only in the 65-84 age group and coronary heart disease only in middle age (age 45-64).

Table 2.4 Fifteen diagnostic groups* accounting for highest percentage of health care costs for five age groups, Netherlands 1994.

	age 0-14		age 15-44		age 45-64		age 65-84		age 85+	
Rank	Diagnostic group	%	Diagnostic group	%	Diagnostic group	%	Diagnostic group	%	Diagnostic group	%
1	Perinatal / congenital conditions	10.2	Mental retardation, Down's syndrome	16.5	Mental retardation, Down's syndrome	9.4	Dementia	9.5	Dementia	22.2
2	Mental retardation, Down's syndrome	9.7	Other mental disorders	8.6	Musculoskeletal diseases	8.3	Stroke	6.7	Stroke	6.6
3	Other respiratory diseases	6.3	Pregnancy	8.5	Dental diseases	6.3	Musculoskeletal diseases	5.8	Falls	5.9
4	Other mental disorders	6.0	Dental diseases	6.6	Ill-defined conditions	5.8	Cancer	5.6	Musculoskeletal diseases	4.3
5	Ill-defined conditions	5.5	Musculoskeletal diseases	6.3	Coronary heart diseases	5.0	Ill-defined conditions	4.6	Ill-defined conditions	3.7
6	Ear disorders	5.2	Ill-defined conditions	4.7	Other mental disorders	4.9	Coronary heart diseases	4.0	Heart failure	2.9
7	Dental disorders	4.6	Schizophrenia	3.5	Cancer	4.6	Other circulatory diseases	3.9	Cancer	2.1
8	Infection	4.0	Depression/anxiety	3.4	Depression/anxiety	3.4	Neurologic disorders	2.9	Other respiratory diseases	2.1
9	Neurologic disorders	2.8	Other accidents	3.1	Other circulatory diseases	3.3	Other mental disorders	2.6	Neurologic disorders	2.0
10	Other accidents	2.3	Genital disorders	2.3	Gastro-intestinal diseases	2.7	Falls	2.5	Other circulatory diseases	1.7
11	Eye disorders	2.2	Skin diseases	2.2	Neurologic disorders	2.7	Asthma & COPD	2.5	Other mental disorders	1.5
12	Asthma & COPD	2.3	Other respiratory diseases	2.0	Liver, gall and pancreas diseases	2.5	Eye disorders	2.3	Liver, gall and pancreas diseases	1.3
13	Musculoskeletal diseases	1.9	Neurologic disorders	2.0	Hypertension	2.5	Diabetes	2.2	Eye disorders	1.2
14	Gastro-intestinal diseases	1.6	Alcohol/drugs	1.6	Asthma & COPD	2.2	Gastro-intestinal diseases	2.2	Urinary disorders	1.2
15	Skin diseases	1.6	Gastro-intestinal diseases	1.6	Other accidents	2.2	Heart failure	2.1	Other accidents	1.1
Share of age groups in total costs	7.9		29.3		20.7		30.6		11.6	
Share of age groups in population	18.4		46.0		22.5		11.8		1.3	

* See table 2.2 for ICD codes of all diagnostic groups.

2.4 Discussion

In the Netherlands, health care costs are dominated by old age and by disability, particularly mental disability and musculoskeletal diseases. The share in the health care budget of the main fatal diseases is relatively modest: all cardiovascular diseases and all cancers, together 67% of all causes of death, cover 17% of all health care costs that could be allocated to a diagnostic group.

Obviously, these results have to be interpreted with caution. The exact share of each separate diagnostic group is less trustworthy than the patterns of distribution which emerge from this data. Firstly, the key variables used to break down costs are generally not collected for epidemiological purposes, but in the Netherlands there is no financial incentive to register one diagnosis rather than another. Only primary diagnoses are taken into account. It is beyond the limits of the method used to assign costs appropriately to the primary as well as each secondary diagnosis. Valid information about secondary diagnoses is generally lacking or incomplete. As a result, costs of diagnoses that are more often registered as secondary or tertiary, such as diabetes, are slightly underestimated. However, the registered primary diagnosis is generally the more important diagnosis for the health care sector concerned, and the main reason why health care is needed: e.g. what the internist calls osteoporosis, is for the surgeon a hip fracture, for the ambulance an accidental fall, and for the nursing home a demented patient. The obvious advantage of the used method is that each guilder is allocated to only one combination of age, sex and diagnostic group, avoiding double counting.

Secondly, the key variables used to break down costs for each health care sector do not represent exactly equal amounts of resources. Not all days of stay in hospitals or nursing homes are equally expensive, some hours of care are more labour intensive than others, and outpatient visits or primary care consultations can vary in length. As a result, costs of some diagnoses may be biased. For example, because hospital nursing costs are broken down by bed days without any differentiation, costs of diagnoses for which relatively more days are spent in intensive care will be slightly underestimated and vice versa.

These limitations, however, will not affect the major findings of this study, such as the exponential increase by age or the heavy health care burden of mental disorders.

The major strength of the present study is its comprehensiveness. This explains why our results seem at variance with a USA-based (Medicare) study that shows decreasing costs at the oldest ages [150]. This latter study does not include long-term home care for the elderly, and care in elderly homes or nursing homes. It is exactly these costs which cause the exponential increase at old age. Our findings agree with the American study, as we found that costs for acute admissions in hospital decrease at the oldest ages (figure 2.1). Most of these patients are already admitted to a nursing home or a home for the elderly, and/or are too old or too ill to consider hospital admission useful. A Swedish study, which is older and less complete, shows the same results [146].

Our findings correspond to a large extent with those of our earlier study for the year 1988 [132 133]. Studies that are more or less comparable have

been published for England [201], Australia [9] and Canada [186]. These studies show basically similar cost patterns, but with lower shares particularly for mental retardation and dementia. However, they either did not consider all health care, particularly long-term (psychiatric) care [9 201], or could not assign these costs to diagnoses [186]. Apart from the degree of comprehensiveness, many other methodological and country-specific issues may cause differences in cost distributions. A serious international comparison of cost-of-illness distributions would require specifically designed cross-national studies.

The present study only considers medical costs; costs of informal care are not included. It has been estimated for the Netherlands that if informal care is entirely substituted by professional care, this would generate costs that are comparable to the current costs of professional home care [98]. Informal care mainly substitutes for simple forms of professional care. If these costs had been included, the total costs of chronic, disabling conditions (e.g. dementia, musculoskeletal disease) would be even more dominant, thus strengthening our conclusions.

It is not surprising that the share of fatal diseases is relatively limited: care stops at death. Disability is the main reason why ill people use health care. The pattern of epidemiological causes of costs found by us is remarkably consistent with the main causes of disability as estimated by Murray and Lopez [192 194]. In 1990, in the developed world, they estimated that mental disorders (including dementia and hereditary disorders of the central nervous system) accounted for 35.5% of life years lived with disability. In the present study, the same disorders, including congenital anomalies, caused 28.4% of all health care costs that could be allocated to diagnostic groups. Musculoskeletal diseases, including arthritis and dorsopathy, caused 7.3% of the allocated health care costs, while Murray and Lopez estimated that osteoarthritis covered 6.1% of the life years lived with disability in the developed world.

The costs presented here are grouped cross-sectional figures. Each age group mixes persons with low or no costs, and persons with high costs due to costly interventions, severe disability or impending death. In higher age groups this mixture shifts towards the latter, causing costs per person to rise. Any lifetime expected costs that are derived from these data only, assume that someone alive today is 'exposed' in the future to currently observed age-specific health care costs.

The cost distribution by age is notifying especially for societies that face a further ageing of the population. Because the distribution of costs is determined by the current prevalence of disease and disability, future health care costs will depend (among others) on the evolution of the risk of disability and death by age. If it were possible to delay the senescent process as cause of

both disability and death, senescence-related costs would be postponed, and perhaps curtailed by death. However, as long as the main disabling diseases of old age, such as dementia, osteoarthritis and hip fractures, remain more or less resistant to prevention and therapy, increasing life expectancy can only result in a steep increase in health care needs.

We conclude that health care costs in the Netherlands are strongly determined by old age and disability. In the future, the ageing of the society will undoubtedly increase health care needs. When talking about cost containment in health care, we should not forget that large shares of the budgets are not spent on 'cure', but on 'care'. Long-term care of the old, the frail and the mentally disabled will always be labour intensive and expensive, but is the hallmark of a civilized society.

Part I Injuries

Abstract

Objectives To describe health care costs of injury by its medical and demographic determinants.

Design An incidence based cost model was developed to estimate the lifetime costs of injury occurring in a specific period. We defined patient groups that are homogeneous in terms of health service use. Health service use per patient group was estimated with data from national health care registers and a prospective follow-up among 5,755 injury patients.

Setting Netherlands, 1998.

Subjects Injury patients presenting at an Emergency Department.

Measures Health care costs.

Results Total health care costs due to injury in 1998 are 1.1 billion euro, or 3.4% of the total health care budget. Health care costs of injury shows two major age peaks: one among males between age 15 and 44 due to high numbers of injury, and the second among among females from age 65 onwards due to high costs per patient. Costs per injury patient rise linearly up to age 60 and rise exponentially thereafter. From age 25 onwards, females account for higher costs per patient than males. Hip fracture (21.3%), superficial injury (13.5%), open wounds (6.1%) and skull-brain injury (6.0%) have the highest total costs. Superficial injuries rank first among the health care costs of injury up to age 65, and is dominated only by hip fracture beyond age 65.

Conclusions Minor injuries without need for hospitalization account for a substantial share of health care costs.

Health care costs of injury in the Netherlands

3

3.1 Introduction

Injuries account for a considerable share in the global burden of disease, estimated at 12% for the established market economies and even higher shares for other global regions [194]. In addition to their impact on public health, injuries are a major cause of health care costs, comparable to the costs of cancer and stroke (chapter 2). Because injuries have a very heterogeneous origin, more detailed information on health care costs by type of injury may help to identify previously unnoticed health problems within this field. Also, such information may be a first step in identifying existing inefficiencies in health care and direct the development of preventive policies and trauma care. Being a unidimensional measure, costs enable rapid comparisons among types of injury that differ with respect to severity and health care need [159].

Previous studies have identified substantial resources going to patients with lower extremity fractures, including hip fracture [10 305]. So far, cost of injury studies have been occasional and limited to specific injuries, health care sectors and age groups [155 159 167 231], or did not distinguish among injury diagnoses [275] (see also chapter 2). Therefore it is largely unknown which types of injury contribute most to the high medical costs of injury.

To fill this knowledge gap, we estimated health care costs of injury by an incidence-based model linked to an injury surveillance system. With this cost model health care costs can be described by type of injury, health sector, basic socio-demographic indicators, and multiple external causes. It is comprehensive for it covers all injuries and all health care sectors including long-term health services. The Netherlands is an ideal setting for such a study, because almost 100% of the population is covered by health insurance, and most important health care sectors have data registries with national coverage.

3.2 Methods

Model description

We developed an incidence-based model [231] to measure and describe the lifetime health care costs of injury occurring in a specified period. The present paper contains results for 1998. A full description of the model is available elsewhere [172]. We considered all injuries of chapter 17 of the international

classification of diseases (ICD, 9th revision) [296], except injury due to medical adverse events (ICD-9 995-999, E870-E879, E930-E949), early complications of trauma (ICD-9 958), late effects of injury (ICD-9 905-909), and injuries occurring in hospitalized patients. Incidence was restricted to patients who attend a hospital Emergency Department (ED), so patients who are fully treated by general practitioners or at the injury scene were excluded.

Table 3.1 Diagnostic groups* used in study and corresponding ICD 9 codes [296].

	Diagnostic group	ICD codes
1	skull-brain injury	800-01, 803-04, 850-54, 950-51
2	facial injury	802, 870-71, 918
3	vertebral column, spinal cord injury	805-06, 839.0-5, 846-47, 952
4	injury to internal organs	860-69, 900-02, 926, 929
5	fractured ribs / sternum	807.0-3, 809
6	fractured collar bone / shoulder	810-11
7	fractured upper arm	812.0-3
8	fractured elbow / lower arm	812.4-5, 813.0-3, 813.8-9
9	fractured wrist	813.4-5, 814
10	fractured hand / finger	815-17
11	dislocation / strain / sprain upper extremities	831-34, 840-42
12	traumatic amputation / crushing injury upper extremities	880.2, 881.2, 882.2, 883.2, 884.2, 885-87, 903, 927
13	pelvis fracture	808
14	hip fracture	820
15	fractured shaft of femur	821.0-1
16	fractured knee / lower leg	821.2-3, 822-23
17	fractured ankle	824
18	fractured foot / toes	825-26
19	dislocation / strain / sprain lower extremities	835-38, 843-45
20	traumatic amputation / crushing injury lower extremities	890.2, 891.2, 892.2, 893.2, 894.2, 895-97, 904, 928
21	superficial injury (incl. contusions)	910-17, 919-24
22	open wound	872-84 (excl. 880.2, 881.2, 882.2, 883.2, 884.2), 890-94 (excl. 890.2, 891.2, 892.2, 893.2, 894.2)
23	burns	940-49
24	poisoning	960-89
25	other and non-specified injury †	807.4-6, 818-19, 827-30, 839.6-9, 848, 925, 930-39, 953-57, 959, 990-95

* Excluded are: late consequences of trauma (ICD-9 905-09), early complications of trauma (ICD-9 958), and injuries due to medical adverse events (ICD-9 996-99).
† Includes: other fractures, other strains and sprains, injury to peripheral nerves, injury due to foreign body, other injury.

We included all health services that are relevant for the treatment and rehabilitation of injury patients, except for dental care, aids and appliances, and institutions for mentally and physically disabled persons, due to lack of information on the cause of injury.

We calculated lifetime health care costs of injury as a multiplication of incidence, transition probabilities (e.g. chance of nursing home admission), health care volumes (e.g. length of stay) and unit costs (e.g. costs per day in nursing home). Incidence, transition probabilities and health care volumes were subdivided by patient groups that are homogeneous in terms of health service use. We tested known determinants of health service use: age, sex, location and type of the injury, and indicators of injury severity[4 159 179], and patient groups were defined accordingly. Injuries were classified by location and type into 39 groups (table 3.1 presents an aggregation into 25 groups) after consultations with experts in traumatology, orthopedics and rehabilitation. We considered hospitalization, number of injuries, and motor vehicle involvement to be indicators of injury severity.

Data sources
Injury incidence was extracted from the Dutch Injury Surveillance System (LIS) for non-hospitalized cases and the hospital discharge register for hospitalized cases. LIS is a continuous monitoring system which records all unintentional and intentional injury treated at 17 ED's in the Netherlands, resulting in a representative 12% sample. The hospital discharge register has national coverage.

For inpatient hospital care, medical procedures, nursing homes and rehabilitative services, we estimated health service use (transition probabilities, health care volumes) from sector-specific data systems with national coverage [198 206 230]. The selection and classification of injury patients from national data systems was based on the registered primary diagnosis. In case of multiple injuries, we determined the primary injury in LIS by application of an algorithm derived from the literature [159].

For emergency services and GP services preceding ED treatment we used data recorded in LIS.

We performed a patient follow-up among a sample of 5,755 injury patients who attended one of the hospitals of LIS between July 14, 1997 and October 18, 1998 in order to collect data on other health services used: intensive care, outpatient visits, GP visits after the ED treatment, outpatient physical therapy, home care, medication, and aids and appliances. The sample contained an overrepresentation of hospitalized patients and severe, less common injuries, such as injuries to the vertebrae and spine and skull-brain injury. Victims from

self-inflicted injury were excluded. Postal questionnaires were sent two, five and nine months after the injury occurrence. As a result, health service use estimated from the questionnaires is up to nine months, while health service use derived from national data systems can be considered lifetime.

Data analysis health service use
For each health care sector for which national data were available, determinants of individual health service use were derived by crosstable analysis, and patient groups were defined accordingly. Length of stay in nursing homes was adjusted for the presence of comorbidity, so days that are not attributable to injury are excluded.

Because the response rates of the patient follow-up were 41.4%, 77.5% and 64.2% for the first, second and third questionnaire respectively, data were adjusted for non-response using socio-demographic and injury-related information from the patient sample. For each type of health service for which data from the questionnaires were used, multivariate logistic regression was used to estimate the probability of health service use (response variable), and to test which determinants of health service use were significantly predictive. We used logistic regression because health service utilization appeared to be very skewed. Only significant ($p<0.05$) determinants were included in the final models and were used to classify patient groups. The estimated probability of health service use multiplied with the average health service use given this probability, results in the average health service use specific for each patient group and health care sector.

Unit costs
For each health care sector we determined costs per volume unit that reflect real resource use. All unit costs were estimated according to national guidelines for health care costing [210]. We assumed that health care fees were representative of real resource use for GP consultations, inpatient medical procedures, home care, and rehabilitative treatment. Unit costs of emergency and ordered transport, inpatient hospital days (excluding medical procedures), outpatient visits, nursing home days, other rehabilitative services, physical therapy, and pharmaceuticals were calculated from national production and cost statistics. All costs are expressed in 1998 Euros.

Costs of ED visits were decomposed and estimated as follows. Visit duration as recorded in LIS was considered as an indicator of nursing costs. Visit duration was entered as a response variable in multivariate linear regression analysis, and determinants of health service use were tested as predictors. Costs (labour time) of physicians by injury group were determined

by expert guesses from two ED physicians. National data on hospital costs were used to calibrate the estimated labour costs, and to calculate material, diagnostic and overhead costs of ED visits.

Figure 3.1 Numbers of injury (ED visits), total health care costs of injury (€1,000) and costs per patient (€) by age and sex, Netherlands 1998. Costs of care include nursing homes and home care.
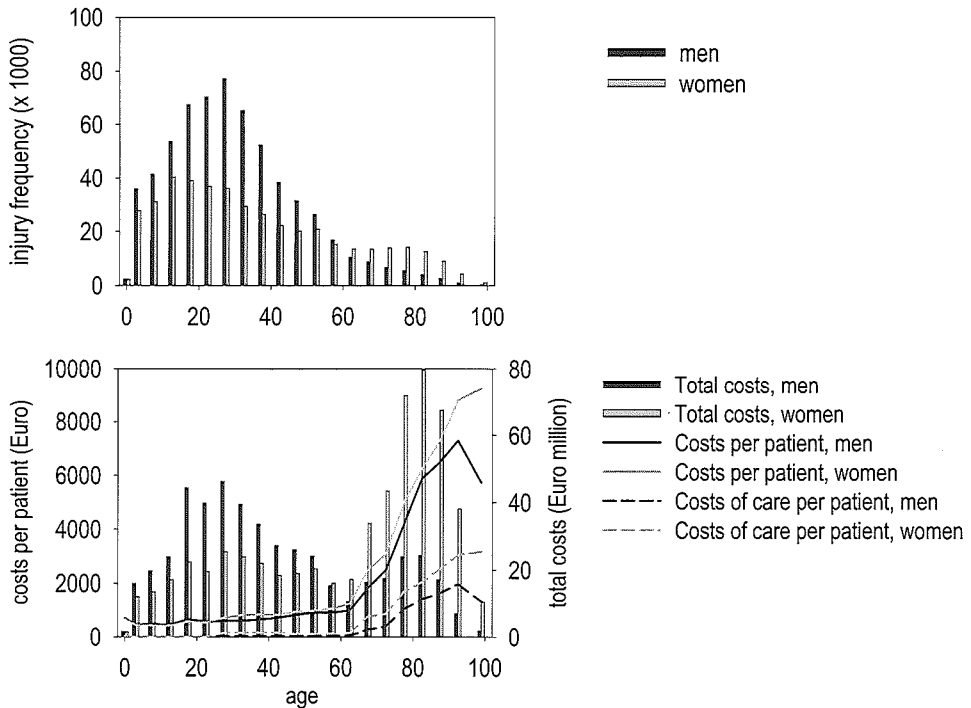
3.3 Results

Total costs of injury were 1.1 billion euro or 3.4% of total health care costs. Costs per capita were Euro 62 for males and Euro 75 for females, and costs per patient were Euro 769 for males and Euro 1,380 for females. Total health care costs of injury reach a first peak among males aged 15 to 44 due to a high number of injuries. An even higher second peak is found among females beyond age 65 due to increasing costs per injured patient with age and their longer life expectancy compared to males (figure 3.1).

Costs per patient show a linear increase from childhood until age 60, and rise exponentially after this age. From age 25 onwards females show higher costs per patient than males, which is largely due to more intensive use of home care and nursing homes.

Table 3.2 Health care costs of injuries by sex, ranked by share (in % of health care costs by sex), and injury frequency, ranked by share in total frequency, Netherlands 1998.

	injury group	Costs				Incidence	
		males	females	total	rank	total	rank
14	hip fracture	11.4	29.2	21.3	1	1.2	17
21	superficial injury	14.6	12.7	13.5	2	34.6	1
22	open wounds	9.2	3.7	6.1	3	17.1	2
1	skull-brain injury	7.9	4.4	6.0	4	2.1	11
16	fractured knee / lower leg	5.7	6.0	5.9	5	1.3	16
19	lower extremity strain / sprain	5.6	4.4	4.9	6	7.6	3
17	fractured ankle	3.6	4.4	4.0	7	1.6	13
9	fractured wrist	2.7	3.9	3.4	8	4.0	7
8	fractured elbow / lower arm	3.0	3.3	3.2	9	2.3	10
3	vertebral column / spinal cord	3.7	2.7	3.2	10	0.6	21
24	poisoning	2.7	3.4	3.1	11	1.6	12
25	other injury	3.5	2.3	2.8	12	5.1	4
13	pelvis fracture	1.7	3.3	2.6	13	0.3	25
10	fractured hand / finger	3.8	1.6	2.6	14	4.1	6
15	fractured shaft of femur	2.3	2.7	2.5	15	0.4	24
7	fractured upper arm	1.2	3.2	2.3	16	0.9	20
2	facial injury	3.2	0.9	2.0	17	4.2	5
4	organ injury	2.5	1.3	1.8	18	0.4	23
11	upper extremity strain / sprain	1.9	1.5	1.7	19	2.9	8
18	fractured foot / toes	2.1	1.3	1.6	20	2.6	9
23	burns	1.7	1.0	1.3	21	1.5	14
5	fractured ribs / sternum	1.5	1.1	1.2	22	0.4	22
12	upper extremity traumatic amputation/crushing injury	2.0	0.5	1.2	23	0.9	19
6	fractured clavicle / shoulder	1.5	0.9	1.1	24	1.3	15
20	lower extremity traumatic amputation / crushing injury	1.0	0.4	0.7	25	1.0	18
	all injury by sex	44.5	55.5	100.0			

Table 3.2 shows the share of injuries in incidence and total health care costs by sex. The four injuries with the highest health care costs are hip fracture, superficial injury (mainly bruises and abrasions), open wounds and skull-brain injury. Hip fracture and skull-brain injury rank 17 and 11 respectively in terms of numbers, but result in high medical costs per patient. The high frequency of superficial injury and open wounds result in a large proportion of health care costs. Of the seven injury groups with the highest total costs, four are lower extremity injury. All fractures combined are responsible for 52.8% of total costs and 21.3% of total incidence.

Table 3.3 Six injury groups accounting for highest percentage of health care costs for selected age groups, Netherlands 1998.

	age 0-14	%	age 15-24	%	age 45-64	%	age 75+	%
1	superficial injury		21.2 superficial injury		21.5 superficial injury		11.7 hip fracture	52.6
2	fractured elbow / lower arm		10.4 lower extremity strain / sprain		9.5 hip fracture		8.0 superficial injury	6.0
3	open wounds		9.2 open wounds		9.4 skull-brain injury		7.9 fractured knee / lower leg	5.8
4	fractured wrist		8.3 skull-brain injury		8.1 fractured knee / lower leg		7.4 pelvis fracture	4.9
5	skull-brain injury		7.8 fractured knee / lower leg		6.0 fractured ankle		6.9 fractured shaft of femur	3.8
6	fractured hand / finger		4.5 fractured hand / finger		4.9 open wounds		6.8 fractured upper arm	3.7

Superficial injuries dominate costs up to age 65. Beyond this age, hip fracture has the highest costs (table 3.3). Up to age 75, open wounds and skull-brain injury (including concussion) are among the six injuries with the highest costs. The importance of injuries to the upper extremities in terms of health care costs is relatively high during childhood, but from age 15 onwards injuries to the lower extremities (e.g. knee and lower leg fractures, hip fractures) increasingly dominate health care costs.

Table 3.4 Six injury groups accounting for highest percentage of health care costs , and share of non-admitted and admitted patients in health care costs of these injuries, Netherlands 1998.

		total costs (€ mln)	admitted patients	nonadmitted patients
1	hip fracture	227.9	99.7%	0.3%
2	superficial injury	145.2	14.5%	85.5%
3	open wounds	65.6	21.2%	78.8%
4	skull-brain injury	64.0	92.6%	7.4%
5	fractured knee / lower leg	63.2	91.3%	8.7%
6	lower extremity strain / sprain	52.7	26.3%	73.7%
	Total	1072.2	64.2%	35.8%

The skewed distribution of health care costs is reflected by the high cost share (approximately two-third) of admitted patients, who account for only 9.4% of total incidence (table 3.4). Their share in costs rises from 37% in age group 0-14 to 94% beyond age 85. This pattern reflects decreasing personal independence and slower recovery with age giving rise to higher hospitalization rates and length of stay. The share of admitted patients in health care costs varies among types of injury, and is relatively high among disabling conditions such as skull-brain injury and lower extremity fractures.

Table 3.5 shows the distribution of health care costs by sector of all injuries together, and of three injuries that are typical for childhood (fracture of

elbow/lower arm), adolescence (skull-brain injury) and old age (hip fracture).
Hospital costs dominate total health care costs of injury with a share of 68.7%,
followed by home care (8.8%) and nursing homes (7.7%). The high cost shares
for ED and outpatient care are typical for fractures of the elbow / lower arm, but
also in general for childhood injury. Nursing homes, home care and inpatient
hospital care are the major cost components in hip fracture, but also in general
among elderly. The share in costs of rehabilitation hospitals that is observed in
skull-brain injury (9.9%) is only higher for injuries of the spinal cord and
vertebral column (14.0%), and for crushing injury and traumatic amputations of
the lower limb (12.1%).

Table 3.5 Total costs (mln Euro), costs per capita and costs per patient (Euro), and share of health care sectors in total health care costs, for selected injury groups and all injuries, Netherlands 1998.

	skull-brain injury	hip fracture	fractured elbow / lower arm	all injuries
Hospital care	67.6	70.9	70.5	68.7
- nursing	53.1	59.4	24.1	35.4
- operations	1.3	7.2	11.5	4.6
- outpatient	5.5	2.5	19.2	11.2
- Emergency Department	7.7	1.8	15.7	17.6
Rehabilitation hospitals	9.9	0.2	0.5	1.8
Nursing homes	6.1	14.9	6.4	7.7
General practitioners	1.1	0.3	1.2	1.5
Ambulance services	5.8	2.5	4.2	5.7
Physical therapy	2.7	2.3	9.6	5.5
Home care	6.5	8.8	7.1	8.8
Pharmaceuticals	0.2	0.1	0.3	0.2
	100.0	100.0	100.0	100.0
Costs per capita (€)	4.1	14.6	2.2	68.5
Costs per patient (€)	2,951	14536	1,382	1,019
Total costs (€ mln)	64.0	227.9	34.1	1072.2

3.4 Discussion

Substantial parts of health care costs of injury are due to high numbers of
injury, such as in males between age 15 and 44, superficial injury and open
wounds, or are due to high costs per patient, such as in females beyond age 65,
hip fracture and skull-brain injury. Minor injuries without need for
hospitalization account for more than one-third of health care costs of injury,
which is almost twice the costs of hip fracture. Costs per patient are higher for
females than for males from age 25 onwards due to higher costs of care.

The biggest strength of our study is that it presents comprehensive estimates of health care costs by injury that are fully comparable across all output dimensions, and that include most relevant health care sectors and both major and minor, intentional and unintentional injury. For those health care sectors that are most important for injuries in terms of health care use - hospital inpatient care, medical procedures, rehabilitation clinics and nursing homes - we used registries with national coverage. Nevertheless, one should take into account that we included only injuries that are treated at an ED and may be hospitalized thereafter. In the Netherlands the total number of injuries on an ED is about 1 million per year (6% of the population) and an additional number of about 1.3 million are fully treated by a GP or other primary health care providers [62]. The vast majority of this second group are patients with minor injuries: cuts, abrasions, superficial injuries, dislocations, strains, sprains, small burns and poisonings. The associated health care costs will add at most 10% to our cost estimate.

Secondly, for outpatient and primary health services we recorded consumption up to nine months after the injury event, and excluded institutions for permanently disabled persons. As a result, we underestimated lifetime consumption that is particularly relevant for injuries with long-term needs, such as permanent brain injury and spinal cord injury. However, for the vast majority of injuries all health care needs are in the first year post-injury [134].

Thirdly, the response on the patient follow-up was 41% which could have biased estimates of outpatient and primary health care. However, systematic response bias was accounted for as far as socio-demographic, injury and treatment related factors were associated with response.

Fourthly, to estimate injury-specific health service use, we adjusted for comorbidity in nursing home costs by using data on length of stay of patients without other disabling conditions. In the patient survey we only asked for health care use related to the injury. Hospital costs include additional days because of other chronic conditions and complications, but this may be justified because the injury was the cause for admission and the additional costs would not have occurred without the injury.

Finally, about 5% of ED patients had multiple injuries. The main injury was then identified by an algorithm that was derived from other studies [159 179], giving priority to spinal cord injury, skull-brain injury and lower extremity injury above injury in other body parts, and to fractures above other types of injury. Because costs were attributed to the main injury, costs of injuries that often occur in combination with more severe injuries were underestimated. For hospitalized patients, we used the registered primary diagnosis in the hospital discharge register.

The comprehensive macro-level approach of the present study makes comparisons with other studies difficult. Most cost of injury studies apply to specific injuries [167], are restricted to hospitalized patients or specific age groups [155 159 161], report only micro-level results or describe costs by injury cause [134]. Compared to the study reported in chapter 2, the present study estimates higher costs for home care, and higher costs for injuries that do not need hospitalization (e.g. superficial injury, upper extremity fractures) which is due to the use of an ED-based injury surveillance system and a separate operationalization of ED costs. In a classical study from the US [231], medical costs of injury were estimated at about $250 per capita (adjusted for inflation up to 1998) or about three times the estimate in the present study. Medical costs per patient of about $1,030 per patient are similar to our estimate of Euro 1,019. However, the US study included all injuries, also those not treated in an ED. When this is accounted for, costs per patient in the Netherlands will be about half the US estimate. In addition, our study reports a much higher share in costs for home care and emergency services, and a lower share for pharmaceuticals. Because many other methodogical and country-specific issues may cause differences, a full comparison will need a specific study.

The cost of injury estimates in the present study have shown the impact on health care of injuries with a high incidence (e.g. superficial injury) compared to injuries that occur far less frequently but have large health care needs (e.g. hip fracture). Health care costs by injury not only represent their economic impact, but also reflect their impact on population health. This shows the potential of cost estimates as a composite population health measure similar to disability adjusted life years (DALY). Reasoning that health care costs are to some extent the product of injury incidence, degree of disability and the duration of this disability, they will rather be associated with the morbidity component than the mortality component of the injury burden. Likewise, the future distribution of health care costs will among others be determined by trends in injury risk, related disability and survival. More research on this matter however has to wait for forthcoming estimates of this burden [194].

Although the present study demonstrates the substantial impact on health care costs of non-hospitalized injury, costs are still unevenly distributed with less than 10% of patients (those hospitalized) accounting for almost two-third of total costs. Due to population aging, and the positive relationship between age and hospitalization independent of type of injury, the share of admitted injury patients in total costs of injury will be probably enhanced in the future.

Cost of illness studies have been criticized for they do not provide enough information to identify health care inefficiency, and they are no aid for

prioritizing health care because they do not give information on the effectiveness, costs and savings of interventions [45 53]. However, without comprehensive burden of disease information, with health care costs (current or projected) being just one health measure, the search for cost-effective interventions will be a blind search. This search is often susceptible to single disease advocacy, whereas comprehensive information on costs of illness (injury) will put specific diseases or injuries into perspective and may highlight other health problems that receive insufficient policy attention. Moreover, cost of illness (injury) studies produce a starting point for cost-effectiveness studies by raising insight into health care costs by types of injury, health sectors and basic demographic indicators, and therefore where costs might potentially be saved or not. For example, when all patients at an ED with minor injury (dislocations, sprains, strains, superficial injury, open wounds, small burns, poisonings, foreign body injury) who did not need hospitalization and who were not referred by their GP were instead treated by their GP, this would reduce the number of ED patients by more than 50% and save about 7% of total health care costs of injury.

We did not conduct estimates of production loss due to injury. Because of the relatively low hospitalization rates of persons in the productive phase, and assumed that minor injuries will often lead to one or more work days lost, we hypothesize that the contribution of non-hospitalized patients to production losses will be even larger than to health care costs. Future research should verify this hypothesis.

44

Abstract

Objective To compare published cost of injury studies from different countries, and increase their usefulness for setting priorities in injury prevention and trauma care.

Design We selected 17 cost of injury studies from Pubmed and our own files, that reported on population based estimates of costs of all injuries combined or transport injuries. The studies were from 6 countries. We assessed their methodology in-depth, and calculated basic economic figures by which results could be compared.

Setting Review of published studies.

Patients All injuries combined, transport injury.

Main outcome measures Costs per capita, costs per patient.

Results Per capita health care costs of all injuries combined ranged from $35-275, and of transport injury from $2-116 (in PPP 2000 US dollars). These differences could partly be attributed to differences in injury incidence, cost items and patient groups included. When these factors were accounted for, considerable differences in costs per patient remained between countries, with high costs in the US and Australia, intermediate costs in the Netherlands, and low costs in Sweden, Norway and New Zealand. Productivity costs of injury were estimated consistently at about three times the medical costs.

Conclusions Reported costs of injury could only partially be made comparable by accounting for differences in methodology, demography and injury incidence. Guidelines for conducting and reporting cost of injury studies are urgently needed to better inform health policy and planning, and enable meaningful comparisons among injury groups, countries, and time periods.

Meerding WJ, Mulder S, van Beeck EF.
Submitted for publication.

Cost of injury studies: do they bring us more than confusion?

4

4.1 Introduction

Injuries are an often neglected though important public health problem. It is by far the most important cause of mortality among persons between 15 and 44 years, and injuries account for 12% of the burden of disease in established market economies, and even higher shares in other global regions [194]. Information on injury-related health care consumption and costs can be complementary to epidemiological data in identifying existing or emerging risks and health problems. It might also be a useful tool to prioritize health policy and improve trauma care [189]. Being a unidimensional measure, costs can be a useful indicator for comparative analysis in a heterogeneous problem field ranging from high frequency minor injuries (e.g. superficial injuries) to low frequency severe injuries (e.g. polytrauma patients).

Several population-based cost of injury studies have been conducted [177 231 275 293]. They all highlight the economic burden of (specific) injuries. However, studies differ notoriously with respect to their methodology, including comprehensiveness, matters of definition and classification, and the way they measure and value costs. They may consider different patient groups, and some studies only consider medical costs whereas others include other societal costs and 'human costs' as well [177]. Although such differences may be justified by the study objectives, they may create confusion among users such as policy makers and health care professionals. They may even attract unjustified attention to (specific) injuries at the expense of other health problems, and complicate the development of a coherent and efficient health policy. Also, artificial differences in cost estimates obscure the genuine causes underlying cost differences that are relevant for policy development.

In this paper, we discuss the most important methodological issues concerning cost of injury studies. In a review of studies from different countries we assessed their comparability and traced the possible reasons for observed differences.

4.2 Methods

Selection of literature

We conducted a Pubmed search for studies on population-based costs of all injuries combined or transport injuries. We only selected English language publications from 1995 onwards. To this set we added older key studies, and studies from participants of the Burden of Injury Conferences in 2000 and 2002 that have partly focussed on the economic burden of injury [151 262]. The resulting set contained generic cost of illness studies, injury-specific studies with population based cost estimates, and studies that primarily focussed on costs per injury patient but extrapolated these costs to population level.

Assessment of studies

We assessed all studies on pre-specified methodological issues. First, we determined the comprehensiveness of studies with respect to included cost items: medical cost items, and other societal costs such as material damage, costs of the legal system, and productivity losses due to work incapacity. Some studies may also monetize premature mortality and lost quality of life ('human costs').

Second, we determined the case definition in each study, particularly regarding external causes and injury diagnoses included, the extent to which non-hospitalized patients were considered, and age criteria.

Third, studies may use either a top-down or bottom-up estimation of costs. In a top-down approach total health care costs are broken down by health care sector and by (injury) diagnoses, with key variables that more or less represent equal amounts of health care resources (e.g. hospital days, outpatient visits). In a bottom-up approach costs per injury patient are multiplied with the number of injuries. Ideally both approaches should yield similar results, but some sources of divergence are known, usually leading to bottom-up costs exceeding top-down costs. One of these sources is comorbidity. In top-down studies, all costs are attributed to the recorded principal diagnosis, whereas in bottom-up studies costs related to comorbid conditions may be attributed to the injury when health care consumption do not discriminate sufficiently among diagnoses. An important limitation of top-down studies is that key variables used to break down total costs may not represent equal amounts of health care resources (e.g. normal versus intensive care days).

Fourth, incidence-based cost studies consider the lifetime costs of injuries that have occurred in a given year, whereas in prevalence-based cost studies health care consumption in a given year is attributed to injuries that have occurred in this or previous years. Again, both approaches may lead to differences in outcomes, for instance due to dynamics in the injury

epidemiology, the practice of discounting future costs, and incomplete data on long-term costs.

Fifth, costs can be calculated as charges or as real costs, the opportunity costs of using resources. Charges may differ from actual resource use. For instance, in the US charges are deliberately set at a higher level because they partly cover the medical care of uninsured persons [180]. How costs should be calculated depends on the study perspective (e.g. societal, insurance company).

Sixth, we assessed the most important data sources that were used. These may be data from health surveys, administrative data (e.g. hospital discharge registers), follow-up interviews of injury patients, etcetera. The possible limitations of these data are numerous, and are more extensively described elsewhere [44 235]. Surveys and follow-up interviews may suffer from recall bias, may not distinguish injury-specific health care consumption, may exclude specific populations (institutionalized persons, severely impaired patients, etc.), and may have incomplete information on long-term health care. Emergency department based systems may not be representative and may have less than optimal practices of data collection and codification [256].

A seventh issue concerns the measurement and valuation of productivity losses due to work days lost. Here two approaches prevail that conflict with regard to their account of long-term labour incapacity and death. According to the human capital method (HCM), the costs of work absence, disability and death are equal to the stream of future production that would have been generated without the disease. The friction cost method (FCM) assumes that in a situation of unemployment sick workers will be eventually replaced by others, thereby limiting the long-term productivity losses [131]. Any losses are limited to the friction period, i.e. the period up to the replacement of the sick worker by a new person, and include also the costs of recruiting and training the new worker.

Data analysis
We have standardized the results of studies by distinguishing among medical costs and other reported costs. Medical costs were calculated per capita, per injury patient, per hospitalized injury patient and per non-hospitalized injury patient. We converted all costs to year 2000 US dollars, first accounting for inflation with use of nominal price indices, and than converting national currencies to US dollars by common exchange rates and by exchange rates adjusted for international differences in purchasing power (PPP). Nominal price indices, exchange rates and PPP adjusted exchange rates were taken from the OECD Health Data [207]. If possible, costs were age-adjusted by direct standardization to the US population of 1997.

4.3 Results

We included 17 studies that were published between 1980 and 2002, of which 11 considered all injuries, 2 considered unintentional injuries, and 13 reported costs of transport injuries (table 4.1). Seven studies were from the US. Half the studies included all patients irrespective of where they were treated, four studies were limited to ED patients [134 147 218] (and chapter 3), and four studies considered hospitalized patients only [105 141 155 158]. The majority of studies adopted a bottom-up, incidence based approach.

Six studies included productivity costs in addition to medical costs [22 104 147 231 275 293], and four studies included one or more of the following non-medical cost items: home modifications, vocational rehabilitation, legal costs, and administrative costs for health and car insurance [22 104 231 293]. As for medical costs, six studies included hospital costs only [105 134 141 147 155 218]. Of the remaining studies, two excluded nursing home care [103 178], one excluded community health services and ambulance services [164], and two excluded aids and appliances [164] (and chapter 3). In addition, one study regarded administrative costs for public and private health insurance as part of medical costs [164], whereas four studies mentioned earlier classified these under non-medical costs.

The basic quantitative results have been summarized in table 4.2. We maximized the comparability by presenting health care costs per capita and per patient in US dollars of 2000. In general, accounting for price level differences had a larger impact than demographic standardization. For example, health care costs of injury in Australia are $111 per capita, $145 when adjusted for price level differences, and $147 when subsequently standardized for population demographics. In the following, we will concentrate on the unstandardized results in $PPP.

Per capita health care costs of all injuries ranged from $35-275. Between country differences are larger than within country differences. Costs per capita are highest in the US [178 231], which is about double the costs in Australia [164 293], and about triple the costs in the Netherlands [275] (see also chapters 2 and 3). The high costs in the US are partly explained by a high incidence (e.g. 1.6 times the incidence in the Netherlands). However, costs per patient in Australia are comparable [164] or even higher ($1,390) [293] than in the US ($1,150) [231].

Interestingly, the Australian bottom-up study [293] generates about 40% higher cost estimates compared to the top-down study [164], whereas the definitions of cases and costs are comparable. Within the US, per capita costs of injury are highest in the Rice-study. The studies of Miller and Harlan do not include institutionalized persons (e.g. nursing homes) and do not fully capture long-term health spending. The Harlan-study is rather outdated and

adjustment for price level could be insufficient. The studies of MacKenzie have low estimates but consider only hospitalized patients [155 158].

The estimates for Sweden [147] and Norway [134] exclude intentional injury and, together with the study of New Zealand [218], include hospital costs only. Even then cost estimates in these countries can be considered low: costs per ED patient are almost $300 (New Zealand), $500 (Norway) and $600 (Sweden), which is far below the estimate for the Netherlands (total costs $1,180, hospital costs $840, see chapter 3).

Table 4.3 Proportions of medical costs, productivity costs and other costs in total costs of injury.

Study	hospital costs[a] as % of medical costs	medical costs as % of total costs	other direct costs as % of total costs	productivity costs as % total costs
Harlan *et al.* (1990)	96	100	0	0
Rice *et al.* (1989)	63	25	4 [c]	72
Miller and Lestina (1996)	73	100	0	0
van Beeck *et al.* (1997)	67 [b]	22 (HCM) 58 (FCM)	0	78 (HCM) 42 (FCM)
Meerding *et al.* (1998)	59	100	0	0
Watson *et al.* (1997)	69	29	1 [d]	71
Mathers *et al.* (1998)	71	100	0	0
Phillips *et al.* (1993)	100	100	0	0
Meerding *et al.* (1999)	71	100	0	0
Lindqvist *et al.* (1996)	100	23	0	77
Kopjar (1997)	100	100	0	0
MacKenzie *et al.* (1988)	99	100	0	0
MacKenzie *et al.* (1990)	100	100	0	0
Hartunian *et al.* (1980) [g]	--	26	7 [e]	67
NHTSA (2002) [g]	--	22	22 [f]	56
Hendrie *et al.* (1994) [g]	100	100	0	0
Langley *et al.* (1993) [g]	100	100	0	0

HCM = human capital method, FCM = friction cost method.
a Includes emergency department services, physicians' hospital services and inpatient rehabilitation.
b Hospital costs are exclusive inpatient rehabilitation.
c Included are home modifications, vocational rehabilitation, and administrative costs for health and car insurance.
d Included are among others income-support.
e Included are administrative costs for insurance and legal costs.
f Included are administrative costs for insurance, legal costs, police and fire services, and vocational rehabilitation.
g Transport injuries only (see table 4.1).

The majority of studies concerned transport injuries (whether or not in public roads) and a minority focussed on motor vehicle crashes only (table 4.1). The international cost pattern is similar to that of all injuries combined. By far the highest costs per capita ($116) were reported by NHTSA [22], a 50% higher

estimate than in Rice's study that ranks second with $75 per capita [231]. We could not find an explanation for this large difference. Hartunian [104] reported much lower results for the US, but these could be too outdated. Again, the high per capita costs in the US are partly due to a relatively high incidence and partly due to high costs per patient.

Compared to the US estimate in Rice's study, per capita costs of transport injury are 60% lower in Australia [293], and 80% lower in the Netherlands [275] (see also chapters 2 and 3). The difference with Australia can largely be explained by a 50% lower incidence in Australia, unadjusted for case definition. The incidence in the Netherlands is 25% lower than the US incidence (1,700 vs 2,200 per 100,000 person years), and the largest part of the difference in per capita cost is therefore accounted for by much lower costs per patient in the Netherlands. The lowest per capita costs were reported for Norway ($2) [134] and Sweden ($11) [147]. Both studies included hospital costs only, but even then the estimates can be regarded low. Costs per patient in Sweden is half the estimate in the Netherlands and doubles the Norwegian estimate, despite similar case definitions. Per capita costs in New Zealand [141] are higher than that reported for Australia [293] considered that the former study was restricted to hospitalized patients and hospital costs. The higher costs in New Zealand are due to a much higher incidence of hospitalized patients (550 and 160 per 100,000 person years in New Zealand and Australia respectively) that outweighs the lower costs per hospitalized patient ($5,450 versus $12,500). Late consequences of injury were not included by Phillips, which largely explains the difference between both studies from New Zealand [141 218].

In studies that aimed to include all medical costs, the proportion of hospital costs in all medical costs ranged from 59% to 99% (table 4.3). The variation could only partly be explained by differences in casemix.

Some studies added non-medical direct costs, e.g. vocational rehabilitation and administrative costs for insurance, and productivity costs. The former represent 1-7% of total costs (here including medical and non-medical direct costs, and productivity costs), except for the NHTSA-study in which the non-medical direct costs (inclusive administrative costs for insurance, legal costs, police and fire services, and vocational rehabilitation) equal the medical costs. Productivity costs about triple the medical costs of injury in studies that adopt the HCM, but are less than the medical costs when estimated according to the FCM [275].

4.4   Discussion

Summary of results

The international comparability of cost of injury studies is poor, largely due to differences in cost items and patient groups included, and to a lesser extent due to differences in approaches for estimating costs: bottom-up or top-down, incidence- or prevalence-based. After accounting for the major methodological differences, substantial differences in per capita and per patient costs remained between countries, and to a lesser extent between studies within the same country. It was however impossible to estimate the proportions of variance explained by specific causes.

Added value of cost of injury studies

Cost estimates provide a composite measure of health care demand that enables rapid comparison among very different types of injury, such as minor injuries with a high incidence (e.g contusions, open wounds) and less frequent injuries with substantial health care needs (e.g. hip fracture, skull-brain injury). In addition to epidemiological indicators, cost estimates may help to identify specific injuries as candidates for intervention.

International comparisons of health care costs by injury potentially provide opportunities for identifying underlying determinants of cost differences. Apart from injury incidence, case mix, and demography, particularly differences in health systems may be an important factor. Comparative studies on total health spending have shown that countries with primary care "gatekeepers" and with capitation systems instead of fee-for-service payment systems for physicians had lower costs [93]. Similarly, international comparisons of costs of injury may help to identify system characteristics that improve the efficiency and effectiveness of trauma care. However, the current lack of transparency obstructs this development. E.g. the high costs per ED patient in the Netherlands compared to Norway and Sweden could be due to more comprehensive cost data, but also to differences in injury severity and health system efficiency.

Cost of illness (COI) studies have often been criticized as being of little relevance to health policy [45 53]. Because they do not give information on the incremental costs and health effects of interventions, they would be useless for resource allocation decisions. However, similar to other indicators of the burden of injury, costs are important for the health intelligence function of governments. Comprehensive cost estimates subdivided by type of health care, injury diagnoses and external causes show at a glance where costs might potentially be saved or where interventions are most needed. Without such comprehensive estimates the search for cost-effective interventions will be a

blind search. Moreover, these estimates may highlight previously unidentified health problems and risks, whereas others are put into perspective, and so help to avoid unjustified single disease advocacy. In other words, COI studies are part of the 'public health accounts' similar to what the national accounts are for macro-economic policy, and nobody would question the utility of the latter. Some opponents argue that COI studies might shift resources towards health problems with the largest resource use ('circularity problem'), but nobody has ever advocated such a practice. Cost-effectiveness information is critical for such resource allocation decisions.

Some cost of injury studies included estimates of production losses. These are a contentious issue because they may exceed medical costs by far, and available methods for measuring these costs need further validation. The human capital method has been argued to overestimate production losses from a societal perspective (see Methods) [131]. Empirical research is needed to measure the actual production loss in case of work absence, and should consider possible compensation of work loss by colleagues or after return to work [243], and the occurrence of catastrophic events with high costs in case of unexpected sick leave. Apart from validation, including productivity costs introduces an equity problem, giving priority to injuries occurring in the working population. It should be investigated whether this is in line with societal values and beliefs.

Two studies considered the monetary value of lost quantity and quality of life due to injury [22 177]. Such values have been derived from willingness to pay (WTP) estimates for health care or reductions in health risks. Although the WTP method is more firmly rooted into the standard welfare economic framework than population health measures, there are concerns about its validity. A major concern is its insensitivity to the size of the good that is valued, i.e. the phenomenon that respondents are unwilling to pay more for (much) larger health gains [209]. This compromises the use of WTP values as a descriptive measure of population health. Also, WTP estimates appear to be very sensitive to study design and framing [92 209], and often exceed by far the costs per QALY thresholds that are commonly used as rules of thumb in health care rationing [108]. Others put forward that WTP values will be positively related to the wealth of the individual and negatively to remaining life expectancy [66].

Towards comparable cost of injury studies
The policy relevance of COI studies as discussed above is compromised by lack of standardization and transparency. The between and within country comparability of results can be increased by the development of guidelines for

conducting and reporting research. These guidelines should particularly focus on the cost items to be included (medical and non-medical), the classification of injuries, the measurement of productivity losses, and reporting. Comparability will already be greatly enhanced when detailed data would be reported for medical costs only, with crosstabulations of costs by type of health care, injury diagnoses and/or external causes. More detailed results could be put on a website in addition to the published key figures. In addition, some basic demographic and epidemiologic indicators should be added to nationwide cost estimates to facilitate interpretation and comparability, e.g. age-specific costs per capita, and population rates for ED visits and hospitalizations. Such practice would facilitate greatly the research into the underlying causes of international differences in health care costs of injury, that are of interest for health policy making.

Table 4.1 Selected cost of injury studies: major characteristics.

Study	External causes	Patient group [5]	Country	Year	Method	Productivity costs
All injuries						
Harlan *et al.* (1990)	all [1]	all	US	1980	BU, prevalence-based	no
Rice *et al.* (1989)	all	all	US	1985	BU, incidence-based	yes
Miller *et al.* (1996)	all [1]	all	US	1987	BU, prevalence-based	no
van Beeck *et al.* (1997)	all	all	Netherlands	1988	TD, prevalence-based	yes
Meerding *et al.* (1998)	all	all	Netherlands	1994	TD, prevalence-based	no
Watson *et al.* (1997)	all	all	Australia: Victoria	1994	BU, incidence-based	yes
Mathers *et al.* (1998)	all	all	Australia	1994	TD, prevalence-based	no
Phillips *et al.* (1993)	all	ED	N-Zealand: Dunedin	1990	BU, incidence-based	no
Meerding *et al.* (2000)	all [2]	ED	Netherlands	1998	BU, incidence-based	no
Lindqvist *et al.* (1996)	unintentional	ED	Sweden: Motala	1983	BU, incidence-based	yes
Kopjar (1997)	unintentional [2]	ED	Norway: Stavanger	1992	BU, incidence-based	no
MacKenzie *et al.* (1988)	all	hospitalized [3], age 16-45	US: Maryland	1983	BU, incidence-based	no
MacKenzie *et al.* (1990)	all [2]	hospitalized	US	1985	BU, incidence-based	no
Transport injuries						
Hartunian *et al.* (1980)	mva	all	US	1975	BU, incidence-based	yes
Rice *et al.* (1989)	mva	all	US	1987	BU, incidence-based	yes
NHTSA (2002)	mva	all	US	2000	BU, incidence-based	yes
van Beeck *et al.* (1997)	transport	all	Netherlands	1988	TD, prevalence-based	yes
Meerding *et al.* (1998)	transport	all	Netherlands	1994	TD, prevalence-based	no
Watson *et al.* (1997)	transport	all	Australia: Victoria	1994	BU, incidence-based	yes
Mathers *et al.* (1998)	transport	all	Australia	1994	TD, prevalence-based	no
Phillips *et al.* (1993)	transport [2]	ED	N-Zealand: Dunedin	1990	BU, incidence-based	no
Meerding *et al.* (2000)	transport [2]	ED	Netherlands	1998	BU, incidence-based	no

Table 4.1 Selected cost of injury studies: major characteristics.

Study	External causes	Patient group [5]	Country	Year	Method	Productivity costs
Lindqvist et al. (1996)	transport [4]	ED	Sweden: Motala	1983	BU, incidence-based	yes
Kopjar (1997)	road traffic [2]	ED	Norway: Stavanger	1992	BU, incidence-based	no
Hendrie et al. (1994)	road traffic	hospitalized	W. Australia	1988	BU, incidence-based	no
Langley et al. (1993)	mvta	hospitalized	N-Zealand: Dunedin	1989	BU, incidence-based	no

mva = motor vehicle accidents, mvta = motor vehicle traffic accidents

[1] Excluding institutionalized
[2] Excluding late consequences of injury.
[3] Includes surviving patients who have been hospitalized for >1 night, and who have not been transferred to another acute care facility.
[4] Traffic injuries during work time are not included.
[5] All = patients treated in primary care facility or ED; ED = patients treated in ED.

Table 4.2 Cross-national comparison of incidence (per 100,000 person years) and medical costs of injury (year 2000 US$).

Study	Country	Patient group	incidence, standardized [a]	health care costs per capita	health care costs per capita (PPP)	health care costs per capita (PPP), standardized [a]	health care costs per patient (PPP)
All injuries							
Harlan et al. (1990)	US	All	--	139	139	143	--
Rice et al. (1989)	US	All	23,599	275	275	275	1,150
Miller and Lestina (1996)	US	All	--	219	219	226	--
van Beeck et al. (1997)	Netherlands	All	--	65	77	--	--
Meerding et al. (1998)	Netherlands	All	15,100 [62]	62	74	77	481
Watson et al. (1997)	Australia	All	10,320	111	145	147	1,389
Mathers et al. (1998)	Australia	All	--	81	105	107	--
Phillips et al. (1993)	N-Zealand	ED	--	--	--	--	266 [c]
Meerding et al. (2000)	Netherlands	ED	6,983	67	79	80	1,182
Lindqvist et al. (1996)	Sweden	ED	11,889 [b]	73 [c]	70 [c]	--	582 [c]
Kopjar (1997)	Norway	ED	7,137	42 [c]	35 [c]	33 [c]	493 [c]
MacKenzie et al. (1988)	US	Hospitalized [d]	856 [b]	81	81	--	9,478
MacKenzie et al. (1990)	US	Hospitalized	895	70 [c]	70 [c]	71 [c]	7,867 [c]
Transport injuries							
Hartunian et al. (1980)	US	All	1,977 [b]	46	46	--	2,337
Rice et al. (1989)	US	All	2,194	75	75	72	3,355
NHTSA (2002)	US	All	1,887	116	116	--	6,144
van Beeck et al. (1997)	Netherlands	All	--	12	15	--	--
Meerding et al. (1998)	Netherlands	All	1,700 [62]	12	15	15	870
Watson et al. (1997)	Australia	All	1,111	23	30	30	2,645
Mathers et al. (1998)	Australia	All	--	14	18	17	--

Table 4.2 Cross-national comparison of incidence (per 100,000 person years) and medical costs of injury (year 2000 US$).

Study	Country	Patient group	incidence, standardized [a]	health care costs per capita	health care costs per capita (PPP)	health care costs per capita (PPP), standardized [a]	health care costs per patient (PPP)
Phillips *et al.* (1993)	N-Zealand	ED	--	--	--	--	--
Meerding *et al.* (2000)	Netherlands	ED	957	13	15	15	1,569
Lindqvist *et al.* (1996)	Sweden	ED	1,525 [b]	12 [c]	11 [c]	--	727 [c]
Kopjar (1997)	Norway	ED	704	3 [c]	2 [c]	2 [c]	326 [c]
Hendrie *et al.* (1994)	Australia	Hospitalized	--	--	--	--	3,523 [c]
Langley *et al.* (1993)	N-Zealand	Hospitalized	551 [b]	20 [c]	30 [c]	--	5,452 [c]

[a] Figures are age-standardized by using the US 1997 population as the standard population.
[b] Not age-standardized
[c] Includes hospital costs only.
[d] Includes surviving patients between 16-45 years and who have been hospitalized for >1 night.

Abstract

Background Insight into the distribution and determinants of both short- and long-term disability can be used to prioritize the development of prevention policies and to improve trauma care. We report on a large follow-up study in a comprehensive population of injury patients.

Methods We fielded a postal questionaire in a stratified sample of 4,639 nonhospitalized and hospitalized injury patients aged 15 years and older, at 2, 5, and 9 months after injury. We gathered sociodemographic information, data on functional outcome with a generic instrument for health status measurement (EuroQol EQ-5D+), and data on work absence.

Results The response rates were 39%, 75% and 68% after 2, 5, and 9 months, respectively. The reported data were adjusted for response bias and stratification. The 2-month health status of nonhospitalized patients was comparable to the general population's health when measured by the EQ-5D summary score, although considerable prevalences of restrictions in usual activities (24.0%) and pain and discomfort (34.8%) were reported. Hospitalized patients reported higher prevalences of disability in all health domains. Their mean EQ-5D summary score increased from 0.62 at 2 months to 0.74 at 5 months but remained below the population norm at 9 months, particularly for patients with a long hospital stay. Patients with injuries of the spinal cord and vertebral column, hip fracture and other lower extremity fractures reported the worst health status, also when adjusted for age, sex and educational level. Age, sex, type of injury, length of stay (LOS), educational level, motor vehicle injury, intensive care unit admission, medical operation, and number of injuries were all significant predictors of functioning. Nonhospitalized and hospitalized injury patients lost on average 5.2 and 72.1 work days, respectively. Of nonhospitalized patients, 5% had not yet returned to work after 2 months, and 39%, 20% and 10% of hospitalized patients had not yet returned to work after 2, 5, and 9 months, respectively. In a multivariate regression analysis, LOS, type of injury, level of education, and ICU admission appeared to be significant predictors of absence duration and return to work.

Conclusions Injury is a major source of disease burden and work absence. Both hospitalized and nonhospitalized patients contribute significantly to this burden.

Distribution and determinants of health and work status in a comprehensive population of injury patients

5

5.1 Introduction

Injury patients are a heterogeneous population with respect to physical, emotional and social functional sequelae, in both the short term and the long term. A uniform comparison of functional consequences after injury is therefore a difficult and challenging task. Insight into the distribution and predictors of both short and long-term disability can be used to prioritize the development of prevention policies and to improve trauma care. Additional monitoring of changes in health will then be supportive of the evaluation of injury control and trauma care.

Several follow-up studies on levels of functioning or disability have already been performed among more or less comprehensive, broadly defined populations of serious trauma patients [95 109 110 157 159 166] or specific serious injuries [123 154 156 223]. In these studies, many predictors of functioning could be identified, such as age, sex, maximum Abbreviated Injury Scale (AIS), serious extremity injury, spinal cord injury, length of stay (LOS), and intensive care unit (ICU) admission. More recently, posttraumatic stress and depression were found to significantly predict physical and social functioning [109 110 223]. Return to (household) work is a more specific indicator of disability and has appeared to be significantly determined by educational level, job type, income, age, body region affected (particularly extremity injury), presence of a supportive network, and nation- specific income replacement services [95 156].

So far, studies have concentrated on severely injured, hospitalized patients up to age 64 and have used different time intervals and measurement instruments. As a result, existing information on disability is difficult to compare across patients with different types of injury, and little is known with respect to elderly patients and the large number of nonhospitalized injury patients. The latter are likely to have minor disability levels with short durations, but there are large numbers of them, so the total disease burden might still be significant.

To fill these significant gaps, we report on a large follow-up study of a comprehensive population of injury patients aged 15 years and older, both hospitalized and nonhospitalized. We aim to answer the following questions: How are the levels of functioning and work status distributed across patient groups in the first year after the injury? How does the health status of injury patients compare with the general population? What personal, injury and health care factors are predictive for levels of functioning and work status?

5.2 Patients and methods

Survey

We conducted a patient survey in a sample of 4,639 injury patients aged 15 and older, who had visited one of the hospital emergency departments (EDs) of the Dutch Injury Surveillance System (LIS) between July 14, 1997, and October 18, 1998. The LIS is based in 17 hospitals in The Netherlands (approximately 15% coverage), in which all unintentional and intentional injuries are recorded. These hospitals are geographically spread across the country; include both academic and nonacademic hospitals, trauma centers and nontrauma centers; and cover representative amounts of urban and rural populations. As a result, the recorded injury incidence in the LIS is regarded representative for the total population. The sample was stratified such that severe, less common injury groups and hospitalized patients were overrepresented in the survey in order to get high enough numbers of patients to analyze differences in functional outcome by type of injury. On average, hospitalized patients were sampled approximately 10 times more than nonhospitalized patients. Hospitalization is decided by the treating physician on the ED and is based on medical needs for the majority of patients; for a small minority it is also based on social needs (e.g., an elderly woman with a concussion living alone). Persons younger than age 15, victims of self-inflicted injury, and institutionalized persons were excluded.

After a pilot test, postal questionnaires were sent 2, 5, and 9 months after the injury, of which the first was posted by the hospital. All hospitals gave permission for the study before the questionnaires were fielded. Before questionnaires were sent, it was verified whether patients were still alive. Reminders could not be sent. The questionnaire was designed to collect information on functioning and work absenteeism, sociodemographic and injury characteristics, and health care use. Of 4,639 persons addressed, 1,806 (39%) responded on average 2 weeks later, giving an average interval of 2.5 months. Because the majority of nonhospitalized patients have minor injuries that need a relatively short recovery period, only hospitalized responders were sent repeat questionnaires, of which 75% and 68% responded after 5 and 9

months, respectively. The response rates are quite common to similar surveys in The Netherlands.

Functional outcome
We used the EuroQol (EQ-5D) generic instrument for measuring functional outcome [260]. In this instrument, health is defined along five dimensions: mobility, self-care, usual activities (such as work, study, housework, and leisure activities), pain/discomfort, and anxiety/depression. Each dimension has three levels: no problem, moderate problem, or severe problem. To capture consequences of head injury, we added a question on cognitive ability (EQ-5D+) [138]. In the second part of the EuroQol instrument, respondents recorded their health status on a visual analogue scale (VAS), between 0 (worst imaginable health state) and 100 (best imaginable health state). We selected the EQ-5D+ because it covers the main health domains that are affected by injury. It was therefore thought to describe well a heterogeneous injury population and to discriminate among specific injuries. In addition, a scoring algorithm, based on empirical valuations from the UK general population and subsequent statistical modeling is available by which each health state description can be expressed into a summary score [65]. This summary score ranges from 1 for full health to 0 for death, and can be interpreted as a judgement on the relative desirability of a health status compared with perfect health. The validity and reliability of the EuroQol instrument have been extensively tested [33 79 265]. It can well be self-administered and takes only 2 minutes to complete [77]. It has been fielded in the general population in several countries and in many specific patient groups. So far, the EuroQol instrument has not been applied to a comprehensive population of injury patients. We compared our estimates with EuroQol data from the Swedish general population [43].

Work status
We added questions relating to work absence, absence duration and return to work (RTW) only in people with paid jobs, to capture one of the more important socioeconomic consequences of health problems. The questions on work absence and RTW strongly relate to the "usual activities" dimension of the EuroQol instrument but are more detailed.

Sociodemographic, injury, and health care characteristics
From the literature, potential determinants of health and work status were identified [86 95 109 110 156 157 159 168]. These can be grouped into sociodemographic (age, sex, and education) and injury-related characteristics (type of injury, number of injuries, motor vehicle crash, hospitalization, LOS,

admission to the ICU, and medical operation). Injury-related factors can be regarded as proxy indicators of injury severity. Educational level was used as an indicator of socioeconomic status. The type of injury was picked from the LIS surveillance system, in which up to three injuries can be recorded by type and body region. The diagnosis of hospitalized patients was verified with information from the hospital discharge register (according to the International Classification of Diseases, ninth revision). In discordant cases, the hospital discharge diagnosis replaced the ED diagnosis. The principal injury was classified by body region, with additional categories for extremity fractures, superficial injury (abrasions and contusions) and open wounds, burns, and poisonings. In case of multiple injuries, the main injury was determined by an algorithm derived from MacKenzie *et al.* [159]. By this algorithm, priority was given to spinal cord injury over skull/brain injury (except concussion), hip fracture, and other lower extremity fractures, respectively. Age and sex could be drawn from the LIS surveillance system, and these were verified by the questionnaire.

Statistical analysis
A nonresponse analysis was performed by forward stepwise multivariate logistic regression separate for the 2-, 5-, and 9-month measurement. We tested age, sex, socioeconomic status, type of injury, type of injury event, motor vehicle injury, ambulance transport, number of injuries, health status (EQ-5D summary score) and hospitalization as possible determinants of nonresponse. Only significant variables ($p < 0.05$) were used to adjust for response bias. Subsequently, the respondents were weighted with the inverse probability of response resulting from the final model. In addition, the data were adjusted for the sample stratification. The resulting weighted data (adjusted for nonresponse and stratification) were representative for the original patient population with injury presenting at an ED in terms of basic demographics, injury cause and type of injury. Further statistical analyses were performed on the weighted data.

We performed regression analyses on the weighted data of each follow-up measurement with the following response variables: the EuroQol summary score and VAS (continuous variables), probability of work absence (dichotomous variable), number of work days lost (continuous variable), overall RTW and RTW of persons who reported work absence (dichotomous variable), and number of work days lost of persons who had returned to work within the 2-month interval (continuous variable). The number of work days lost was corrected for small differences in the follow-up time by assuming a constant RTW hazard over time. In case the reported work absence duration was longer

than the time interval , the reported work absence was interpreted as calendar days and was subsequently converted to work days. Eleven percent of responders did not report on one or more health domains of the EQ-5D. Because the summary score can only be computed in case of complete information on all health domains, the missing values were estimated by hotdeck imputation in case only one or two domains were not reported, using the reported values of persons with similar scores in the health domains that were reported [238].

 The sociodemographic and injury-related characteristics were tested as significant predictors of functional outcome and work status in forward-step multivariate regression analyses. They were all entered as categorical variables. We included an injury by hospitalization interaction term, in order to test whether the distribution of functioning and work status by type of injury was significantly different between persons who were not hospitalized or were hospitalized for a short (<7 days) or long (≥7 days) time. The extreme unequal weighting of the data due to the nonresponse analysis and adjustment for stratification could influence the identification of significant independent variables. To avoid this we used bootstrap analysis [176 226]. This is a socalled resampling technique by which a specified number of population samples are drawn from the data (iterations), given the distribution of the population across the variables that are tested. The distribution of the drawn populations across the variables provides information about the significance of each variable. We performed 100 iterations to test the significance levels of the independent variables. We calculated overall p values using the covariance matrix resulting from the bootstrap replicas. The most significant variable was entered into the model, and the other variables were subsequently entered. This procedure was repeated until none of the remaining variables was significant. The 95% confidence intervals of the variables in the univariate and final multivariate models were determined by using the 2.5% lowest and highest percentiles of 500 iterations.

5.3 Results

Study population

Because of stratification, severe injuries such as lower extremity fractures (26.5%) and skull/brain injury (9.9%), female sex (45.2%), and persons aged 65 and older (27.4%) were overrepresented in the study sample (table 5.1). The proportion of traffic injury was twice as high (26.1%) as in the Dutch Injury Surveillance System (13.2%). Home and leisure injuries, occupational injuries and intentional injuries represent 66%, 13%, and 6%, respectively, of all ED attendances and were slightly underrepresented in the study population.

Persons with a lower chance to respond were young (15-34 years old) and elderly (75+ years old), males, less educated, nonhospitalized, reported a good health status (EQ-5D summary score) at the previous measurement, victims of home injury and violence, and/or had the following injuries: intoxication, foreign body injury, spinal cord injury, eye injury, concussion, burns, and ankle/knee strain or sprain. These determinants were all significant (p <0.05) in a multivariate analysis.

Table 5.1 Study population by age, sex, injury, and hospitalization.

	Dutch Injury Surveillance System, 1997 n=106,318	Study sample * n=4,639	Respon- dents † n=1,806	Respondents working population† n=896
age	%	%	%	%
15-24	26.8	19.2	18.4	20.4
25-44	42.5	32.8	31.8	50.1
45-64	18.7	20.6	25.0	29.5
65-74	5.4	9.0	11.7	--
75-84	4.4	11.0	9.0	--
85+	2.2	7.3	4.0	--
	100.0	100.0	100.0	100.0
sex				
male	60.2	54.6	50.6	66.0
female	39.8	45.4	49.4	34.0
	100.0	100.0	100.0	100.0
type of injury				
skull/brain	1.9	9.9	8.9	10.2
facial injury	5.3	4.1	2.9	3.6
spine, vertebrae	0.6	5.8	6.8	7.0
internal organs	1.0	6.7	6.9	7.4
upper extremity, fractures	10.9	10.4	12.8	11.6
upper extremity, other injury	4.2	6.4	6.2	8.4
hip fracture	1.5	8.0	6.6	1.5
lower extremity, other fractures	6.3	18.5	21.0	19.4
lower extremity, other injury	10.7	7.5	8.3	10.4
superficial, open wounds	49.4	13.7	13.6	14.0
burns	1.5	2.4	1.8	2.5
poisonings	1.6	3.2	1.6	1.1
other	4.9	3.4	2.8	3.1
	100.0	100.0	100.0	100.0
hospitalization				
no	91.0	37.4	33.1	35.4
yes, <7 days	4.7	32.2	37.2	40.4
yes, >=7 days	4.3	30.4	29.7	24.2
	100.0	100.0	100.0	100.0

* Hospitalized patients and patients with skull/brain injury, injuries to the spine or vertebral column were overrepresented.
† Response to the 2-month questionnaire.

Table 5.2 Prevalence (95% CI) of moderate or severe problems after 2 months in the EQ-5D health domains and cognitive ability (in %), and mean EQ-5D summary score and VAS by key indicators.

	mobility	self-care	usual activities	pain, discomfort	anxiety, depression	cognitive ability	EQ-5D summary score	VAS
Total	17 (14, 22)	7 (5, 11)	28 (22, 35)	37 (31, 45)	12 (9, 16)	5 (3, 9)	0.86 (0.84, 0.88)	82 (79, 84)
hospitalization								
no	14 (10, 20)	5 (3, 9)	24 (17, 32)	35 (27, 43)	10 (7, 15)	4 (1, 7)	0.88 (0.86, 0.90)	83 (80, 85)
yes, <7 days	31 (27, 37)	15 (12, 19)	50 (46, 55)	56 (51, 62)	27 (22, 32)	18 (14, 23)	0.73 (0.70, 0.75)	77 (75, 79)
yes, >=7 days	79 (76, 83)	50 (45, 55)	84 (81, 87)	76 (71, 79)	44 (39, 49)	32 (27, 38)	0.51 (0.48, 0.53)	63 (61, 66)
age								
15-24	9 (5, 16)	10 (3, 21)	18 (9, 31)	29 (17, 43)	12 (5, 27)	3 (1, 6)	0.89 (0.85, 0.94)	86 (82, 90)
25-44	18 (9, 27)	1 (1, 2)	30 (18, 44)	39 (27, 49)	9 (5, 16)	5 (2, 12)	0.88 (0.85, 0.91)	83 (77, 87)
45-65	13 (9, 19)	8 (5, 11)	25 (15, 36)	39 (27, 51)	10 (5, 19)	3 (1, 4)	0.86 (0.82, 0.90)	79 (71, 85)
65-74	35 (19, 51)	18 (10, 32)	38 (23, 63)	43 (28, 70)	14 (9, 26)	12 (4, 25)	0.80 (0.66, 0.87)	78 (68, 86)
75-84	66 (51, 83)	35 (25, 49)	58 (44, 79)	58 (44, 75)	43 (28, 59)	21 (14, 33)	0.63 (0.52, 0.77)	66 (60, 72)
85+	74 (57, 90)	67 (50, 88)	74 (53, 93)	60 (41, 77)	45 (30, 61)	52 (30, 73)	0.46 (0.27, 0.61)	45 (38, 54) ·
sex								
males	12 (7, 19)	6 (3, 12)	24 (16, 35)	29 (21, 38)	4 (3, 6)	3 (2, 5)	0.90 (0.87, 0.92)	84 (80, 87)
females	25 (19, 33)	9 (7, 11)	32 (24, 45)	49 (40, 60)	22 (16, 31)	9 (5, 15)	0.81 (0.76, 0.84)	79 (75, 82)
type of injury								
skull/brain injury	17 (10, 23)	8 (4, 15)	32 (23, 47)	44 (32, 60)	22 (15, 29)	31 (23, 43)	0.80 (0.70, 0.85)	77 (72, 81)
facial fracture, eye injury	1 (0, 5)	0 (0, 0)	8 (1, 27)	9 (3, 20)	3 (1, 8)	2 (0, 5)	0.96 (0.94, 0.98)	83 (76, 90)
spine, vertebrae	41 (31, 51)	25 (16, 35)	56 (44, 67)	69 (60, 79)	29 (18, 42)	21 (13, 31)	0.64 (0.56, 0.73)	68 (61, 73)
internal organ injury	24 (16, 31)	10 (3, 17)	41 (28, 50)	60 (48, 69)	20 (12, 29)	9 (4, 14)	0.77 (0.74, 0.82)	74 (71, 78)
upper extremity fracture	5 (2, 10)	18 (12, 31)	34 (25, 49)	57 (44, 72)	12 (7, 21)	4 (1, 9)	0.78 (0.70, 0.83)	80 (74, 85)
upper extremity, other	2 (0, 4)	19 (11, 27)	41 (29, 53)	54 (42, 66)	19 (10, 29)	6 (1, 12)	0.81 (0.77, 0.84)	77 (73, 82)
hip fracture	90 (84, 94)	68 (58, 77)	92 (87, 96)	74 (65, 83)	45 (33, 56)	40 (29, 52)	0.45 (0.40, 0.50)	56 (51, 61)
lower extremity, other fractures	57 (49, 68)	14 (10, 18)	51 (41, 63)	56 (45, 67)	19 (13, 27)	8 (5, 13)	0.74 (0.69, 0.78)	79 (74, 83)
lower extremity, other injury	44 (33, 63)	5 (1, 13)	39 (25, 54)	54 (38, 70)	19 (9, 33)	0 (0, 1)	0.80 (0.74, 0.85)	83 (79, 87)
superficial injury, open wounds	9 (1, 17)	4 (0, 12)	23 (9, 35)	31 (17, 45)	9 (4, 17)	5 (1, 11)	0.90 (0.87, 0.94)	83 (78, 88)
burns	0 (0, 1)	0 (0, 0)	0 (0, 1)	5 (1, 15)	5 (1, 14)	0 (0, 1)	0.98 (0.96, 0.99)	87 (81, 90)
poisonings	12 (3, 22)	4 (0, 12)	6 (0, 16)	9 (0, 19)	20 (6, 38)	17 (4, 29)	0.84 (0.73, 0.94)	85 (78, 93)
other injury	27 (5, 57)	3 (1, 8)	15 (3, 42)	34 (7, 58)	4 (2, 9)	2 (1, 6)	0.88 (0.81, 0.94)	86 (79, 91)

CI, confidence interval; VAS, visual analogue scale; EQ-5D, EuroQol questionnaire.

Figure 5.1 EQ-5D summary score (95% CI) by age and hospital admission. The injury patients data are 2 months after the injury occurred for patients 15 years and older. Data are adjusted for selective nonresponse and stratification. The injury patients data are from the present study, the data for the Swedish general population are from Burström (2001). Notice that the age groups have unequal lengths, in order to enable comparison with the Swedish population.
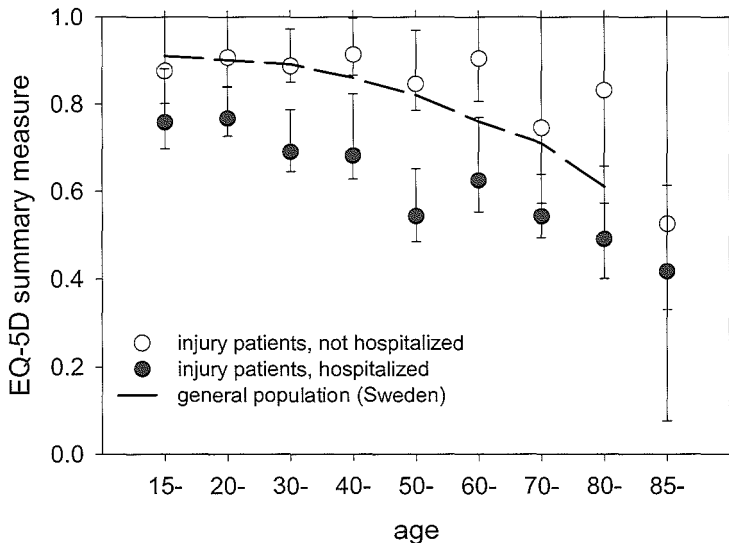
Functional outcome (EuroQol)

The presented results are all adjusted for response bias and sample stratification. Two months after the injury, the health status of nonhospitalized patients was similar to the general population's health when measured by the EQ-5D summary score (figure 5.1). In contrast, the health status of hospitalized patients was lower in all age groups, and significant in most groups. The decrease in health status by age in both hospitalized and nonhospitalized injury patients was similar to what has been observed in the general population.

Nevertheless, a significant proportion of nonhospitalized persons reported restrictions in usual activities (24.0%) and pain and discomfort (34.8%) (table 5.2). Hospitalized patients reported higher prevalences of disability in all health domains than nonhospitalized patients, and even much higher in case of long-term hospital stay. All injury patients combined, the highest prevalences of disability were reported for pain and discomfort (37.4%), restrictions in usual activities (27.6%), and mobility (17.2%).

Patients with injuries of the spinal cord and vertebral column, hip fracture and other lower extremity fractures reported the worst health status as

measured by the EQ-5D summary score. Patients with burns, facial injury (eye injury, fractures), or superficial injury (contusions, wounds, and abrasions) reported the best health status. The health status as measured by the VAS was lower at the higher end of the spectrum compared with the EQ-5D summary score, and higher at the lower end, but the ordering by type of injury was almost similar.

Figure 5.2 EQ-5D summary score by type of injury and hospitalization. The point estimates represent predicted values from a statistical model. The model includes age, sex, injury, hospital admission, a injury times hospital admission interaction term, and education as determinants of health status. The presented figures are for age 15-19, males, and lowest educational level. The size of the dots represent the number of observations.
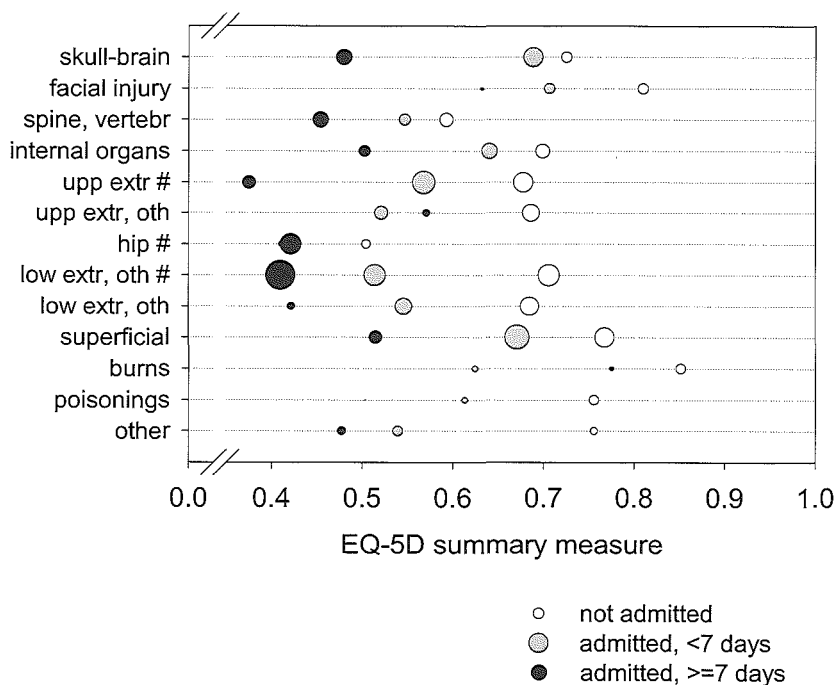
In the multivariate regression analysis, personal background characteristics (age, sex, and educational level), type of injury, injury severity (measured by motor vehicle involvement and number of injuries) and health care-related factors (hospitalization, LOS, and medical operation) were all significant predictors of health status (EQ-5D summary score, table 5.3). All coefficients had the expected sign, among which were a positive correlation between health status and a higher educational level and a better health status

for men compared with women. When adjusted for age, sex, and education, the ranking of type of injury in terms of quality of life remained the same, but the difference between hip fracture and all other injuries became smaller (summary score changes from –0.45 to –0.30 with respect to superficial injury). The distribution of functioning by type of injury differed significantly between nonhospitalized patients and hospitalized patients with a short or long LOS (figure 5.2). For example, among shortly hospitalized patients a relatively favorable health status was found in patients with skull/brain injury and facial injury; among patients with a long LOS, a relatively bad health status was found in patients with extremity fractures and injuries of the spinal cord and vertebral column. The average health status of nonhospitalized patients was similar to the general population, but patients with injury to the vertebral column appeared to be the most unfavorable exception to this rule.

Table 5.3 Determinants of health and work status after 2 months and their p values by multivariate regression analysis.

	EQ-5D summary score	absence probability	absence duration	return to work probability
Age	<0.0001*	0.62	<0.01*	0.40
Sex	<0.001*	0.38	0.19	0.98
Hospital LOS	NA	NA	NA	<0.0001*
Type of injury	NA	NA	NA	<0.0001*
Hospital LOS *times* type of injury	<0.05*	<0.0001*	<0.0001*	0.96
Admittance to ICU	0.07	<0.001*	<0.01*	<0.05*
Medical operation	<0.05*	0.32	0.64	0.32
Education	<0.01*	0.71	<0.0001*	<0.01*
Motor vehicle involvement	<0.05*	0.54	0.61	0.50
Number of injuries	<0.05*	<0.0001*	0.52	0.43

NA, not applicable; LOS, length of stay.
* significant. Significance levels are for models including all other significant variables.

We specifically addressed the impact of a second injury on level of functioning, as was observed in 4.3% of nonhospitalized patients and 18.4% of hospitalized patients. A second or third injury had a relatively large negative influence in patients with lower extremity fractures (except hip fracture), upper extremity fractures, facial fractures, and internal organ injury (table 5.4). For hospitalized patients, the mean EQ-5D summary score increased from 0.62 at 2 months to 0.74 at 5 months and remained stable up to 9 months, whereas the mean VAS score was 70 at 2 months, and increased to 74 and 76 at 5 and 9 months, respectively (table 5.5). Particularly in patients with a long LOS (≥7 days) health status remained below general population norms [43], as well as in patients below age 60 (data not shown).

Table 5.4 Functional outcome after 2 months by type of injury as single injury or with multiple injury.*

	single injury		multiple injury	
skull/brain injury	-0.090	(-0.260, 0.021)	-0.092	(-0.173, -0.028)
facial fracture, eye injury	0.041	(-0.006, 0.097)	-0.139	(-0.232, 0.003)
spine, vertebrae	-0.198	(-0.295, -0.092)	-0.258	(-0.455, -0.065)
internal organs	-0.084	(-0.147, -0.021)	-0.224	(-0.334, -0.114)
upper extremity fracture	-0.083	(-0.160, -0.015)	-0.212	(-0.336, -0.123)
other upper extremity injury	-0.089	(-0.159, -0.026)	-0.176	(-0.271, -0.106)
hip fracture	-0.306	(-0.401, -0.199)	-0.306	(-0.444, -0.168)
other lower extremity fracture	-0.115	(-0.166, -0.064)	-0.407	(-0.480, -0.328)
other lower extremity injury	-0.095	(-0.170, -0.023)	0.003	(-0.722, 0.169)
superficial injury, open wounds	0.0		-0.091	(-0.217, 0.046)
burns	0.071	(0.012, 0.124)	0.097	(-0.107, 0.184)
poisonings	-0.064	(-0.183, 0.043)	0.061	(-0.008, 0.156)
other injury	-0.040	(-0.129, 0.036)	-0.271	(-0.405, -0.126)

* Coefficients (95% CI) from a multivariate linear regression model with the EuroQol-5D summary score as dependent variable, adjusted for age, sex and educational level.

Table 5.5 Health and work status in the first year after injury: mean (95% CI) EQ-5D summary score, VAS, work absence and return-to-work rate after 2, 5, and 9 months.

	N*	2 months	5 months	9 months	Work days lost
		EQ-5D summary score			
all injury	1806	0.86 (0.84, 0.88)	NA	NA	
not hospitalized	598	0.88 (0.86, 0.90)	NA	NA	
hospitalized	1408	0.63 (0.61, 0.66)	0.74 (0.70, 0.77)	0.74 (0.67, 0.78)	
<7 days	671	0.73 (0.70, 0.75)	0.82 (0.80, 0.84)	0.85 (0.80, 0.90)	
≥7 days	537	0.51 (0.48, 0.53)	0.65 (0.59, 0.70)	0.62 (0.54, 0.69)	
		VAS			
all injury	1806	82 (79, 84)	NA	NA	
not hospitalized	598	83 (80, 85)	NA	NA	
hospitalized	1408	71 (69, 73)	74 (72, 75)	76 (73, 78)	
<7 days	671	77 (75, 79)	77 (75, 79)	82 (80, 84)	
≥7 days	537	63 (61, 66)	71 (68, 73)	70 (65, 74)	
		Return to work rates			
all injury	896	0.93 (0.88, 0.96)	NA	NA	11.2
not hospitalized	317	0.95 (0.89, 0.99)	NA	NA	5.2
hospitalized	579	0.63 (0.58, 0.67)	0.80 (0.76, 0.85)	0.90 (0.86, 0.94)	72.1
<7 days	362	0.76 (0.71, 0.82)	0.92 (0.88, 0.96)	0.96 (0.91, 0.99)	52.6
≥7 days	217	0.37 (0.30, 0.44)	0.61 (0.51, 0.69)	0.84 (0.74, 0.90)	101.4

CI, confidence interval; VAS, visual analogue scale; EQ-5D, EuroQol questionnaire; NA, not applicable.
* Response to the 2-month questionnaire.

In a multivariate regression analysis, age, sex (women), and hospital LOS appeared to have a significant (p < 0.05) negative association with health status in hospitalized patients at each interval, and type of injury appeared a significant (p < 0.0001) predictor of short-term health status only (table 5.6). In contrast to the analysis of health status in nonhospitalized and hospitalized

patients combined (table 5.3), ICU admittance appeared to have a significant (p = 0.04) negative correllation with health status after 2 months, and offset medical operation, motor vehicle crash, and number of injuries as significant predictors of health status.

Table 5.6 *P* values resulting from multivariate regression analysis of health status (EQ-5D summary score) at 2, 5, and 9 months.

	2 months	5 months	9 months
age	<0.0001*	<0.0001*	<0.01*
sex	<0.001*	<0.01*	0.03*
education	<0.01*	0.08	<0.01*
hospital LOS	<0.0001*	<0.0001*	<0.0001*
type of injury	<0.0001*	0.29	0.66
admittance to ICU	0.04*	0.61	0.14
medical operation	0.34	0.14	0.17
motor vehicle involvement	0.07	0.34	0.08
number of injuries	0.15	1.00	0.16

LOS, length of stay.
* Significant. Significance levels are for models including all other significant variables.

Table 5.7 Functioning and work status after 2 months for patients aged 15 to 64 with skull/brain injury.

	N	EQ-5D summary score mean	VAS mean	cognitive disability * %	return to work † N	%
concussion	71	0.83	79	26	48	96
not hospitalized	14	0.85	79	18	10	100
hospitalized	57	0.81	79	39	38	90
<7 days	50	0.82	80	39	33	93
≥7 days	7	0.74	73	41	5	69
skull fractures, intracranial injury	64	0.75	73	38	43	59
not hospitalized	11	0.85	78	20	8	72
hospitalized	53	0.68	71	51	35	50
<7 days	18	0.84	77	31	12	89
≥7 days	35	0.59	68	61	23	30
all injury	1359	0.88	83	4	896	93
not hospitalized	495	0.89	83	3	317	95
hospitalized	864	0.70	76	17	579	63
<7 days	549	0.75	78	16	362	76
≥7 days	315	0.59	70	20	217	37

* Minor or severe limitations in cognitive function.
† Including nonabsentees.

Table 5.8 Mean (95% CI) rates of work absence and return to work within 2.5 months by key indicators.

	Absence probability (rate)	Return to work (rate) *	Return to work (rate) †
Total	0.57 (0.47, 0.68)	0.88 (0.80, 0.94)	0.93 (0.88, 0.96)
age			
15-24	0.56 (0.33, 0.75)	0.76 (0.47, 0.94)	0.86 (0.66, 0.97)
25-44	0.55 (0.42, 0.71)	0.93 (0.87, 0.96)	0.96 (0.93, 0.98)
45-65	0.62 (0.45, 0.80)	0.89 (0.80, 0.94)	0.94 (0.87, 0.97)
sex			
males	0.53 (0.40, 0.67)	0.88 (0.75, 0.95)	0.94 (0.85, 0.98)
females	0.63 (0.45, 0.79)	0.89 (0.81, 0.95)	0.93 (0.87, 0.97)
type of injury			
skull/brain injury	0.80 (0.64, 0.99)	0.86 (0.78, 0.93)	0.88 (0.81, 0.94)
facial fracture, eye injury	0.65 (0.26, 0.93)	1.00 (0.97, 1.00)	1.00 (0.99, 1.00)
spine, vertebrae	0.79 (0.65, 0.90)	0.57 (0.40, 0.71)	0.67 (0.54, 0.80)
internal organs	0.88 (0.74, 0.98)	0.71 (0.60, 0.84)	0.75 (0.62, 0.84)
upper extremity fracture	0.85 (0.68, 1.00)	0.78 (0.60, 0.95)	0.82 (0.65, 0.95)
other upper extremity injury	0.69 (0.52, 0.84)	0.87 (0.77, 0.96)	0.91 (0.85, 0.96)
hip fracture	0.99 (0.99, 1.00)	0.42 (0.18, 0.70)	0.42 (0.15, 0.74)
other lower extremity fracture	0.75 (0.62, 0.90)	0.73 (0.60, 0.82)	0.80 (0.70, 0.87)
other lower extremity injury	0.72 (0.52, 0.90)	0.92 (0.81, 0.99)	0.94 (0.85, 0.99)
superficial injury, open wounds	0.44 (0.33, 0.64)	0.90 (0.69, 1.00)	0.96 (0.85, 1.00)
burns	0.36 (0.16, 0.64)	1.00 (0.97, 1.00)	1.00 (0.99, 1.00)
poisonings	0.60 (0.27, 1.00)	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)
other injury	0.64 (0.29, 0.94)	0.94 (0.82, 0.98)	0.96 (0.90, 0.99)
hospitalization			
No	0.54 (0.44, 0.67)	0.91 (0.83, 0.97)	0.95 (0.89, 0.99)
Yes	0.95 (0.92, 0.98)	0.61 (0.56, 0.65)	0.63 (0.58, 0.67)
<7 days	0.93 (0.88, 0.98)	0.75 (0.68, 0.79)	0.76 (0.71, 0.82)
≥7 days	0.99 (0.97, 1.00)	0.36 (0.30, 0.42)	0.37 (0.30, 0.44)

* Return to work rates of those who reported work days lost.
† Return to work rates including those who did not report work days lost.

General and cognitive functioning of patients with skull/brain injury

For patients with skull/brain injury, we specifically analyzed both general and cognitive functioning. In this analysis, we excluded the elderly to eliminate cognitive dysfunction resulting from other causes. Two months postinjury, 26% of patients between the ages of 15 and 64 with concussion and 38% of those with skull fractures and/or intracranial injury had minor or severe cognitive limitations, compared with 4% on average (table 5.7). Nonhospitalized patients with skull/brain injury performed slightly worse than average in terms of overall functioning (EQ-5D summary score and VAS), and cognitive limitations were much more prevalent (18-20%) than average (3%). In contrast, although hospitalized patients with skull/brain injury had much higher prevalences of

cognitive limitations, they performed at least average in terms of overall functioning and return to work. When stratified for hospitalization and LOS, no significant differences exist between patients with concussion and those with more severe skull/brain injury in terms of health and work status, except for patients with skull fractures and/or intracranial injury with a long LOS, who had much lower levels of general and cognitive functioning and a worse return to work rate.

Figure 5.3 Absence duration by injury and hospital admission. The point estimates represent predicted values from a statistical model. The model includes injury, hospital admission, a injury times hospital admission interaction term, and education as determinants of work absence. The presented figures are for the lowest educational level. The size of the dots represent the number of observations.
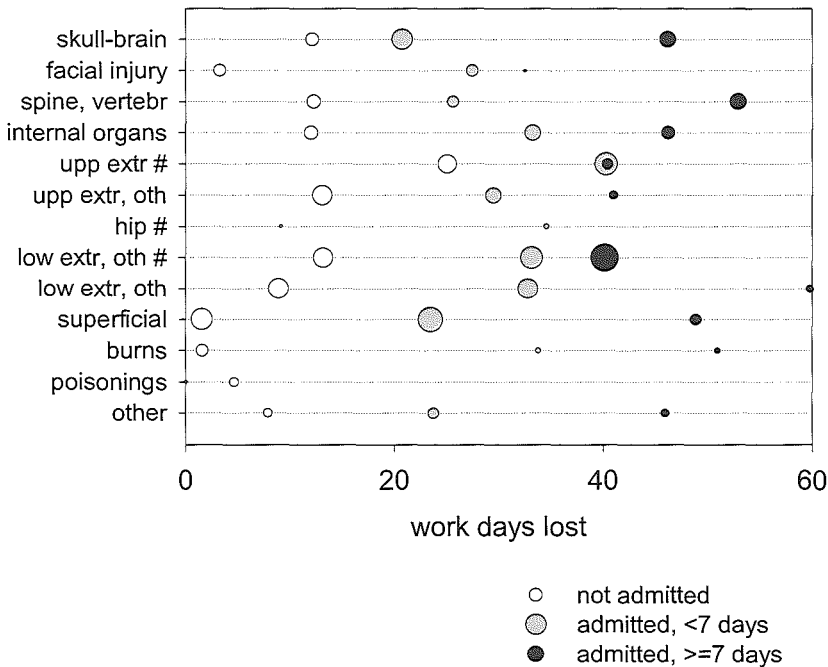
Work status
Among patients with paid jobs, an average of 11.2 lost work days were reported to have been caused by the injury: 5.2 work days among nonhospitalized patients and 72.1 work days among hospitalized patients (table 5.5). Because nonhospitalized patients account for approximately 90% of all injury patients, the total number of work days lost is comparable among both patient groups. Work absence was reported by 54% of nonhospitalized patients and 95% of

hospitalized patients (table 5.8). After 2 months the proportion of hospitalized patients who had returned to work was much lower (61%) compared with nonhospitalized patients (95%). For hospitalized patients, this proportion increased to 80% and 90% after 5 and 9 months, respectively (table 5.5). Patients with a long LOS had considerably worse return-to-work rates than those with a short LOS. Hospitalized patients with lower extremity fracture (excluding hip fracture) had the lowest return-to-work rates: 38%, 64%, and 83% had returned to work within 2, 5, and 9 months, respectively.

Table 5.9 Odds ratio's (95% CI) by type of injury for work absence and return to work within 2.5 months after the injury, adjusted for educational level.

type of injury	absence duration (work days) *		absence probability (rate)		return to work (rate) †
skull/brain injury	9.5	(5.5, 16.0)	5.4	(1.8, 156.9)	0.7 (0.3, 1.3)
facial fracture, eye injury	1.9	(0.8, 4.3)	2.6	(0.6, 26.1)	2.7 (0.8, 9.1)
spine, vertebrae	13.6	(7.6, 22.6)	5.4	(2.0, 28.3)	0.3 (0.1, 0.6)
internal organs	11.6	(6.6, 19.9)	10.3	(3.5, 99.3)	0.5 (0.2, 0.8)
upper extremity fracture	13.0	(7.2, 22.5)	9.8	(3.4, >1000)	0.5 (0.2, 1.3)
other upper extremity injury	7.3	(4.0, 12.3)	2.8	(0.9, 12.4)	0.8 (0.3, 1.5)
hip fracture	26.2	(13.2, 49.3)	178.4	(118.2, >1000)	0.2 (0.0, 0.5)
other lower extremity fracture	9.5	(5.6, 15.6)	4.2	(1.5, 17.4)	0.4 (0.2, 0.8)
other lower extremity injury	5.0	(2.4, 9.5)	3.9	(1.3, 23.2)	1.1 (0.3, 2.6)
superficial injury, open wounds	1.0		1.0		1.0
burns	1.2	(0.4, 2.9)	0.6	(0.1, 3.0)	2.5 (0.7, 16.5)
poisonings	2.1	(0.5, 5.6)	2.2	(0.4, 23.1)	3.9 (1.5, 11.6)
other injury	5.0	(1.4, 10.8)	2.5	(0.5, 30.0)	1.3 (0.4, 3.1)

* An odds of 2.0 means an absence duration two times the base category.
† Return-to-work rates of those who reported work days lost.

In a multivariate analysis, lower educational level, hospitalization and LOS, ICU admittance, and specific types of injury were positively associated with work absence (duration and probability), and negatively with RTW (table 5.3). Age was significantly predictive for work absence duration, but a linear age pattern could not be identified (data not shown). Long-term work absence in hospitalized patients was predominantly associated with educational level and hospital LOS. The highest numbers of lost work days were reported by patients with hip fracture, injuries of the spinal cord and vertebral column, and upper extremity fractures, also when adjusted for educational level (table 5.9). We adjusted for educational level because this is an indicator of job-related factors that mediate RTW, and it appeared to be significantly related to absence duration and RTW. Patients with lower extremity fractures had a lower RTW probability within 2 months than those with upper extremity fractures, but among the latter group a larger share reported work absence. For work absence duration and probability, the hospitalization times type of injury interaction

term appeared statistically significant, indicating that there were significant differences among types of injury within nonhospitalized and hospitalized patient groups. This is illustrated in figure 5.3. For instance, among nonhospitalized and shortly hospitalized patients by far the most work days lost were observed in upper extremity injuries, whereas among patients with a long hospital LOS the highest number of work days lost were reported by patients with injuries to the vertebral column and spinal cord.

5.4   Discussion

It can be concluded that 2 months after injury, the EQ-5D summary score of nonhospitalized injury patients was almost similar to general population norms, and 95% of those with paid work had returned to work. The main unfavorable exceptions were patients with injuries to the vertebral column and to the extremities. In contrast, hospitalized patients had a significantly lower EQ-5D summary score in all age groups, especially patients with hip fracture, injuries of the vertebral column and spinal cord and other lower extremity fractures (adjusted for age and sex). Mean health status of hospitalized injury patients improved up to 5 months, but stabilized thereafter at suboptimal levels, predominantly in patients with long LOS. Of all hospitalized patients working before the injury, significant portions (39%, 20%, and 10%) had not yet returned to work after 2, 5, and 9 months, respectively.

Methodological issues
A major limitation of our study was the relatively low response rate (39% overall, 42% in hospitalized patients on the 2-month questionnaire). This is largely due to the use of postal questionnaires and the very heterogeneous population, and because it was not feasible to use reminders. However, relevant information on the nonrespondents was available from the Injury Surveillance System, which enabled us to perform an extensive nonresponse analysis in which the role of personal, injury related and health care factors was investigated. As a consequence, we were able to adjust for major determinants of nonresponse. Particularly young adults, elderly and patients with minor injuries showed significantly lower response rates. It was not possible to consider socioeconomic status and work status prior to the injury as determinants of nonresponse to the 2-month questionnaire. These are generally found to be positively related to response, and appeared to be positively related to health status. This might have caused our results to be relatively favorable. On the other hand, we found that injury severity was negatively associated with response. It could be that we have insufficiently adjusted for injury severity, for instance when the relatively less severe injuries within an injury

group were less responding. This might have caused our results to be relatively unfavorable, particularly the 9-month data.

One of the major issues in the measurement of functional outcome is to determine the appropriate time intervals. The measurement moment should be representative for a specific stage within the recovery process. The number and length of these stages will differ between specific injuries, and so will the appropriate measurement intervals, but uniform guidelines do not exist. Although the time interval in this study was the shortest feasible, it appears to be too long for nonhospitalized patients (90% of the original population), because the majority of these patients had returned to normal health. Because these injuries can be temporarily quite disabling, it can be concluded that follow-up of these minor injuries should take place a few weeks after the injury.

In addition, because no further increase in the EQ-5D summary score was observed after 5 months, we may conclude that the 2- and 5-month interval were sufficiently remote for measuring an improvement in health status, also in specific age and injury groups, whereas the 5- and 9-month measurements could have been too close. On the other hand, there are other possible explanations for the stabilizing aggregate health status after 5 months. First, a significant number of the patients might have achieved their maximum level of functioning, leaving no room for further improvements. Second, we observed a non-significant decrease in health status in elderly beyond age 85, who constitute about 10% of hospitalized injury patients. Although this might be due to remaining nonresponse bias after the adjustments we made, a deteriorating health status in the long term is generally observed in the very elderly, particularly after specific injuries [90]. Third, we used a generic, non-disease specific instrument for measuring health status (EuroQol) that might have been insufficiently sensitive for aggregate changes in health after 5 months. However, its validity and reliability has been extensively tested and are prerequisites for the measurement of functioning. Moreover, the short completion time, the possibilities for making comparisons with other patient groups, and the possibilities for computing a summary score are major advantages.

The EQ-5D+ appeared to be a feasible and valid instrument for the measurement of functioning and disability in injury patients. The domains discriminated well among different patient groups. Hospitalized patients with skull/brain injury scored relatively favorable on the EQ-5D summary score, although 55% reported minor or severe cognitive limitations after 2 months. This indicates that the EQ-5D captures the specific health consequences of skull/brain injury insufficiently, and underlines the previously indicated usefulness of the additional question on cognitive ability [138].

We compared the EQ-5D summary score with the averages in the general Swedish population [43], which revealed a striking similarity of the age gradient. The age gradient in the general population is a reflection of the prevalence of disability and chronic conditions. For the purpose of the present study, it therefore seems a valid way to adjust for age-specific comorbidity, to identify the injury-specific contribution to loss of health-related quality of life. However, this hypothesis should be confirmed with other research in which direct information on comorbidity is available.

The rank ordering of patient groups was similar between the EQ-5D summary score and the VAS. However, patients with relatively high levels of functioning scored systematically lower on the VAS scale than on the EQ-5D summary score, and vice versa. This socalled end aversion bias is a general observation in visual analogue scales [257].

Other issues and comparison with other studies
The present study has clearly shown the explanatory value of the sociodemographic factors age, sex, and socioeconomic status with respect to the variance in functional outcome, in addition to the type of injury and severity. The role of other factors next to injury-related impairments has been assumed in earlier studies on post-injury functioning [154], and particularly age has been shown a significant term in later studies on long-term disability [5 109]. Sociodemographic factors can be regarded as distal factors, to be distinguished from more proximal factors that are more directly related to disability, and that (partly) mediate the relation between distal factors and disability. An increased understanding of these proximal factors may help improve trauma care and patient recovery, especially those factors that may be amenable to intervention. Examples of proximal factors are comorbidity including mental health, social support, propensity to complain, and the perception of disease and disability. Specific job characteristics will mediate the relationship between educational level and return to work. The importance of preinjury health and social support has been indicated in studies on long-term disability [5 183], and particularly (acquired) poor mental health, post-traumatic depression, and stress, receive increasing attention and appear to have a major impact on overall health status [109 175 223].

Our observation that educational level is positively related to RTW confirms findings in other studies [95 156]. We did not identify a relationship with age, contrary to what has been found in these studies, particulary for persons aged 45-64 years [35 95 156]. However, other factors that have been associated with faster RTW were not investigated in the present study, such as job characteristics, the presence of a social network, and income. In the

Netherlands, work absentees generally receive complete income compensation in the first year after injury, thus neutralizing income-related incentives for return to work. However, much remains unclear about the independent contribution of these factors (e.g., specific job characteristics need to be identified for the development of focussed measures that stimulate persons to resume work).

We found that LOS, motor vehicle involvement, and number of injuries were all independently and significantly predictive for postinjury levels of functioning, even when adjusted for age, hospitalization, and type of injury. ICU admission appeared to be a significant predictor of absence duration and RTW, and functional status in hospitalized patients. These results indicate that these proxy measures for injury severity could be useful for predicting functional recovery and RTW, especially in situations were injury severity is not routinely recorded with classifications such as the AIS and the Injury Severity Score [12]. Particularly the maximum AIS related to extremity and spinal cord injury has been previously identified as a good predictor of functional status and RTW [157 159], whereas Glancy *et al.* found a significant relationship between RTW and the Injury Severity Score [95]. Of the proxy measures in the present study, only LOS and ICU admission have been identified previously as good predictors of functioning and RTW [109 157]. We found that a second injury deteriorates levels of functioning, especially when the principal injury was a lower extremity fracture (excluding hip fracture), upper extremity fracture, internal injury, or facial fracture. Previous research has shown similar findings for patients with lower extremity fractures, especially in combination with head injury [123]. Our findings give rise to future research on this matter, to identify specific combinations of injuries that are particularly disabling.

In previous studies with comprehensive patient populations, spinal cord injury and extremity injury have been identified as being particularly disabling also after several months [109 157 159]. This has given rise to other studies on these particular injuries [123 154]. We found that particularly lower extremity fractures and injuries to the vertebral column and spinal cord had the most negative impact on levels of functioning, also when adjusted for sociodemographic variables.

Because among the injuries to the vertebral column and spinal cord in our sample only few patients (7 of 270) had spinal cord injury, injuries to the vertebral column such as strains, sprains, and fractures without damage to the spinal cord have a significant impact on levels of functioning. In terms of work days lost, upper extremity fractures and skull/brain injury can be added to the injuries with the highest impact. This confirms findings reported previously, although these concern results 1 year after injury [159].

We found that significantly more patients had returned to work after 9 months (90%) than in the study by MacKenzie *et al.* after one year (57%) [159]. Similarly, we found that 38% of hospitalized patients with lower extremity injury (excluding hip fracture) had returned to work after 2 months, which is higher than found elsewhere (26% after three months) [156]. The discrepancies are likely because of differences in case mix, and possibly also because we included persons who had only partly returned to work.

In the present study persons with burn injury showed a relatively good functional status. However, our sample is not representative for burn injuries, because patients with severe burn injury are treated in one of the three burn centres in The Netherlands, and these were not included in the study. Previous research has shown the severe difficulties with community integration for patients with major burns [76].

If we take into account the age composition of hospitalized injury patients in the Netherlands, the observed EQ-5D summary scores after 5 and 9 months (0.74) are still below the population norm of 0.79 calculated with Swedish data [43]. However, our results compare favorably with the average Quality of Well-being scores of 0.63 and 0.67 that were found in hospitalized patients after 6 and 12 months, respectively, by Holbrook *et al.* (population norm of Quality of Well-being score, approximately 0.80) [109]. This is at least partly due to the more severe patient mix in the study by Holbrook *et al.*: 62% were traffic injuries, compared with only 29% in Dutch hospitals.

Until now, studies describing and predicting injury-related disability have concentrated on trauma center patients between 16 and 65 years old. The present study is the first which has been designed to measure levels of functioning and work status after injury in a comprehensive sample of hospitalized and nonhospitalized patients presenting at an ED. In The Netherlands the total number of injuries on an ED is approximately 1 million per year (6% of the population), of whom approximately 10% are hospitalized, and an addional number of approximately 1.3 million injuries are treated only by a general practitioner or other primary health care providers [62]. The vast majority of this group are patients with minor injuries: cuts, abrasions, contusions, strains, sprains, and small burns. Although the duration and level of dysfunction reported by nonhospitalized injury patients are relatively modest, the total burden of injury may still be high because of the large numbers of patients. Similarly, the average number of work days lost was more than 10 times as high for hospitalized patients (72.1 days) compared with nonhospitalized patients (5.2 days up to 2 months after the injury), but the latter constitute 90% of the injuries. This justifies further study on functional outcome and factors determining RTW in injuries that are generally considered minor.

We conclude that injury is a major source of disease burden and work absence. Both hospitalized and nonhospitalized patients contribute significantly to this burden.

Part II Cervical cancer

Abstract

Objective To assess the difference in costs between PAPNET-assisted and conventional microscopy of cervical smears when used as a primary screening tool.

Study design We performed time measurements of the initial screening of smears by four cytotechnicians in one laboratory. Time was measured in 816 conventionally screened smears and in 614 smears with PAPNET-assisted screening. Data were collected on the shares of initial screening, clerical activities and other activities in the total work time of cytotechnicians in the routine situation, and on resource requirements for both techniques.

Results PAPNET saved on average 22% on initial screening time per smear. Due to costs of processing and additional equipment, the costs of PAPNET-assisted screening are estimated to be $2.85 (and at least $1.79) higher per smear than conventional microscopy. The difference in costs is sensitive to the rate of time saving, the possibility to save on quality control procedures and the share of the initial screening time in the total work time of cytotechnicians.

Conclusion Although PAPNET is time saving compared with conventional microscopy, the associated reduction in personnel costs is outweighed by the costs of scanning the slides and additional equipment. This conclusion holds under a variety of assumptions. Using PAPNET in stead of conventional microscopy as a primary screening tool will make cervical cancer screening less cost-effective, unless the costs for PAPNET are considerably reduced and its sensitivity and/or specificity are considerably improved.

Cost analysis of PAPNET-assisted versus conventional Pap smear evaluation in primary screening of cervical smears

6

6.1 Introduction

The PAPNET system [162] is one of various techniques that are being developed to improve on the conventional microscopic examination of Pap smears. It assists the cytotechnician by scanning the slide and selecting through neural network technology those cells on the slide that are most likely to be neoplastic. These cells are represented by 128 images and stored on a digital tape. The images are presented to the cytotechnician on a monitor. If desired, the selected cells can be traced on the original slide under the microscope by use of X/Y coordinates. The automated scanning of slides can be remote or in the review laboratory.

Several studies have assessed the effectiveness [28 69 83 136 237] or have explored the cost implications [99] and the cost-effectiveness [113 143 227 228 241] of the PAPNET system. These studies all refer to a situation where PAPNET is used as a rescreen instrument, and use anecdotal information concerning resource implications. Recently the PAPNET system has been evaluated when used as a primary screening tool. It turned out to have similar diagnostic performance as conventional microscopy, and as a consequence will be equally effective in terms of health benefits [68]. Given this result, PAPNET-assisted screening will only be more cost-effective when costs per smear are lower than with conventional microscopy.

It has been reported that the screen rate is 3-4 times faster with PAPNET compared to conventional microscopy [83 120]. However, these studies compare rescreening in a study situation with routinely used conventional microscopy, and only consider the video image assessment exclusive additional microscopy. In addition, the initial screening of slides is only part of the entire laboratory process of smear examination.

In this paper we present the results of a cost analysis of PAPNET compared with conventional microscopy that was executed parallel to the evaluation of the diagnostic performance [68]. We gathered empirical data on the time involved in screening smears in a study situation using both methods, and on how cytotechnicians spend their total work time in the routine situation.

We also assessed the required resources for conventional microscopy and PAPNET-assisted screening.

6.2 Materials and Methods

PAPNET versus conventional microscopy

On the one hand PAPNET-assisted screening may save screening time, and thereby reduce the number of cytotechnicians, microscopes, and housing costs. On the other hand the PAPNET technology requires additional investments in review stations and involves costs for scanning of slides and for training. We assume that the introduction of PAPNET will have no implications for the amount of work by the pathologist, senior cytotechnician or secretariat. This is reasonable because it has been indicated that the rate of potentially positive smears does not change when PAPNET is used as a primary screening tool [68]. Other studies have suggested that the rate of potentially positive smears increases [28 69 83 136 237]. However, these concern situations in which PAPNET is used as a rescreen instrument, which by definition increases the rate of positive smears.

Cytotechnicians' work time activities

In order to evaluate the impact of PAPNET cytotechnicians' work time, we distinguished three main activities:

1. Initial screening of smears. Using conventional microscopy this includes smear examination, review of the clinical and pathologic history, and reporting of results on paper. Using PAPNET this includes examination of the images, additional microscopy (quick scanning, X/Y coordinates or total screening) in case of technical problems or smear abnormalities following a protocol [68], review of the clinical and pathologic history, and electronic reporting of results.

2. Clerical activities. This includes colouring and filing of slides, data entry of screen results in the computer, and follow-up of women with positive smears. These activities do not change when PAPNET is introduced.

3. Other activities. These are a mixture of quality control activities and aspecific work time (telephone calls, staff meetings and personal time). Quality control comprises multiple screening of high risk smears, screening of marked cells, and rescreening of previous smears in case of positive smears. In routine practice, these activities may happen during initial screening and bilateral and multilateral team discussions among cytotechnicians or with the pathologist. In this study however, we strictly separate initial screening from other activities.

We made this distinction in activities in order to take into account that a reduction in the initial screening time through PAPNET for instance by 50% does not imply that the number of smears processed per cytotechnician can be doubled. Quality control and clerical activities are largely related to the number of smears processed and as a result their shares in the total work time will increase when the number of smears per cytotechnician increases.

Materials
The cost analysis was part of a larger study on the accuracy of the PAPNET-system using archival smears from 1988 [68]. Time measurements were performed on a random sample of these smears. Because the study considered an enriched sample of smears (11.0% were assessed SIL+), we calculated the average screening time according to the mix of negative and positive smears in the Dutch screening programme in the same laboratory in 1988 (1.8% SIL+).
        For PAPNET, smears containing suspicious cells or other specific characteristics, were manually reviewed either rapidly or more detailed according to a protocol [68].

Time measurements
Of all four cytotechnicians participating in the study we measured the initial screening time per smear for each screening technique. All cytotechnicians were experienced in conventional microscopy and had gone through the standard PAPNET training before the start of the study, but they had never worked with PAPNET routinely. In order to capture possible learning effects in PAPNET-assisted screening, the measurements took place on two occasions, three and nine months after the start of the study. The screening for the study took place in a separate room, without the disturbances common to a routine situation, creating equal circumstances for both screening techniques. Time was clocked and registered manually for conventional microscopy and electronically (using the system-integrated time clock) for PAPNET-assisted screening. In PAPNET-assisted screening, any additional microscopy of a smear took place immediately after the video image assessment and was included in the initial screening time measured. The screening times were linked with the cytologic result.
        To know the share of initial screening activities within total work time, we combined the results of this study with a recent time measurement during routine practice in the same laboratory. During a week all cytotechnicians (including the four cytotechnicians in the study) had registered time spent on screening and other activities combined and clerical activities, as well as the number of smears screened. The difference between the total time per smear for

screening and other activities combined in this recent time measurement and the initial screening time per smear as measured in this study generates the time per smear for 'other activities'.

Resource requirements
In our assessment of the resources needed for PAPNET-assisted screening and conventional microscopy we made the following assumptions. For a workload of 40,000 smears per year, process costs for PAPNET are at least $2.82 per smear. This includes only scan station lease costs, technical service, software upgrades, two review stations and training for two cytotechnicians and two scan station operators. When additional review stations and training of cytotechnicians are needed this involves extra investments. We assumed one review station for two cytotechnicians plus one extra review station to handle peak workloads.

For each full time cytotechnician we assumed 200 effective work days per year, one microscope, 10 square metres for housing ($180 per m2 / month), and $1,067 per year for central management, administration and travelling costs. All costs are expressed in 1997 US dollars using an exchange rate of $0.51 per Dutch guilder.

6.3 Results
Time measurements in the study
The crude average initial screening time was 232 seconds (sd = 80 sec) for conventional microscopy and 189 seconds (sd = 109 sec) for PAPNET-assisted screening (table 6.1). The average initial screening time with PAPNET was significantly lower in the second time measurement (t=9 months) than in the first one (t=3 months): 183 vs. 203 seconds (p=.02). We considered the results of the second time measurement representative for a routine situation. As expected, for both screening techniques the average initial screening time increased with the degree of cytologic abnormality (table 6.1). PAPNET was time saving for normal smears only. This reflects the utility of the PAPNET system as an instrument that accelerates the selection of a relatively small sample of abnormal smears from the large pool of screening smears. When corrected for the share of abnormal smears in the Dutch screening programme, the average screening time changes slightly for PAPNET only. This indicates that the sample of smears has been representative for a routine screening situation.

The learning effect with PAPNET as reflected in the reduction of the initial screening time between month 3 and 9 of the study period was mainly related to a reduced usage of the microscope next to the video image

assessment. During the first time measurement only 25% of smears passed without additional microscopy, while this increased to 39% during the second time measurement. Additional microscopy increased the average screening time considerably from 100 seconds per smear screened without microscopy up to 291 seconds per smear with full microscopy.

Table 6.1 Time measurement: average initial screening time in seconds per smear (n), by cytologic result and screen technique.

	Conventional	PAPNET: 1st and 2nd time measurement *	PAPNET: 2nd time measurement *
Unsatisfactory	157 (42)	172 (16)	159 (11)
Negative	219 (620)	169 (499)	164 (365)
ASCUS	303 (82)	279 (41)	304 (20)
SIL	301 (72)	300 (58)	307 (37)
Total	232 (816)	189 (614)	183 (433)
Total, adjusted †	232 (816)	187 (614)	186 (433)

* Including manual microscopy.
† Adjusted to mix of negative and positive smears in the national screening programme in 1988 in the same laboratory: unsatisfactory 0.2%, negative 84.1%, ASCUS 13.9%, SIL 1.8%.

Table 6.2 Time measurements: average time in seconds per smear (n) of initial screening and of initial screening plus 'other activities', by cytotechnician and screen technique.

Cytotechnician	A	B	C	D	Mean
A. Initial screening: study situation					
Conventional *	182 (196)	260 (222)	233 (199)	246 (199)	230
PAPNET * †	116 (100)	235 (122)	178 (102)	192 (109)	180
Reduction by PAPNET	-36%	-9%	-24%	-22%	-22%
B. Initial screening plus 'other activities': routine situation					
Conventional	429 (109)	675 (76)	432 (25)	459 (78)	499
C. Share of initial screening in total work time‡					
Conventional	36%	33%	46%	45%	39%

* Adjusted to mix of negative and positive smears in the national screening programme.
† 2nd measurement only.
‡ Calculation: (A / B) x 85%, because 15% of total work time is spent on clerical activities.

Compared to conventional microscopy, the average initial screening time was reduced by 9% to 36% for PAPNET depending on the cytotechnician (table 6.2). We took the average of 22% as baseline estimate. The relative differences in the average screening time among the cytotechnicians were larger with PAPNET compared to conventional microscopy, but the ranking of the cytotechnicians in terms of speed was the same.

Time measurements in the routine situation
It was observed in the routine situation time measurement that initial screening and 'other activities' taken together accounted for 85% of total work time of cytotechnicians. Per smear this was 499 seconds for the cytotechnicians participating in the study and 621 seconds for all cytotechnicians (40 smears per full work day of 8 hours), so the cytotechnicians participating in the study were 20% faster on average. Combining these findings with the time measurement in this study, it can be concluded for the cytotechnicians participating in the study that the share of initial screening in total work time was 39% on average (range 33% - 46%, table 6.2), leaving 46% for 'other activities', and 15% for clerical activities (114 seconds per smear).

Number of smears per cytotechnician; cost calculations
In table 6.3 the baseline and alternative assumptions are summarized for calculating the number of smears with the PAPNET-system in a routine practice. We were able to capture only the initial screening time of smears for both screening techniques. Our baseline assumption is that PAPNET only saves on the initial screening. In a sensitivity analysis we accounted for the possibility that PAPNET-assisted screening will streamline quality control activities or reduce aspecific work time, which has been translated to a time reduction of 'other activities' analogous to initial screening. In addition we varied the rates of time saving according to the observed differences among the cytotechnicians. A third assumption is a 50% increase in initial screening time for conventional microscopy, holding total work time per smear constant, accounting for the fact that the measured initial screening time per smear might have been biased downward due to the study situation. We assumed PAPNET does not save time on clerical activities.

Table 6.3 Assumptions underlying the calculation of number of smears per cytotechnician and costs per smear.

Assumptions	Baseline	Sensitivity analysis
1. Rate of time saving by PAPNET	22%	9% (low), 36% (high)
2. PAPNET saves time on	initial screening only	initial screening and other activities, but not on clerical activities
3. Initial screening time per smear for conventional microscopy	286 sec	430 sec (+50%)

In table 6.4 the baseline calculations are presented with regard to the increase in number of smears per cytotechnician and the change in costs due to PAPNET. When PAPNET only saves time on initial screening (baseline), this will lead to a 9% increase in smears per cytotechnician per year and a decrease

in cytotechnician's costs by $0.36 per smear. However, due to the PAPNET costs (review stations and scanning), total costs per smear increase by $2.85.

Table 6.4 Baseline results: screen time, smear production and cost per smear for conventional microscopy and PAPNET-assisted screening. *

	Conventional microscopy	PAPNET †
Cytotechnicians' time per smear (sec):		
- initial screentime (sec)	286 (39%) ‡	224
- clerical activities (sec)	114 (15%)	114
- other activities (sec)	335 (46%)¶	335
- total work time (sec)	735 (100%)	673
Smears per cytotechnician/yr	7840	8560
Cytotechnicians	5.1 fte	4.7 fte
Costs per smear:		
- cytotechnicians costs	$4.36	$4.00
- microscopes and housing	$1.32	$1.28
- PAPNET costs	--	$3.25 §
- all other costs	$10.62	$10.62
- total	$16.30	$19.15

* Results have been calculated for a laboratory size of 40,000 smears per year.
† Baseline assumptions: PAPNET only saves on initial screening, and with a rate of -22% with respect to conventional microscopy.
‡ Initial screentime = 230 sec x 1.25 = 286. The factor 1.25 is because all laboratory cytotechnicians taken together were on average 25% slower than the cytotechnicians participating in the study.
¶ Other activities = 621 sec - initial screentime. 621 sec has been measured in the routine situation, and includes both initial screening and other activities, excluding clerical activities.
§ Processing fee (scanning, technical service, software upgrades, two review stations, training for two cytotechnicians and two scan station operators) $2.82, processing time $0.22, extra training $0.06, extra review stations $0.15.

Table 6.5 Results sensitivity analysis (see table 6.3): cost difference per smear using PAPNET relative to conventional microscopy.*

	1. PAPNET reduction in screen time		
	-9%	-22%	-36%
2a. PAPNET saves on initial screening only	+$3.08	+$2.85	+$2.58
2b. PAPNET saves on initial screening and on other activities (excl. clerical activities)	+$2.88	+$2.38	+$1.79
3. Initial screening time per smear contentional microscopy 50% higher	+$2.99	+$2.65	+$2.24

* Results have been calculated for a laboratory size of 40,000 smears per year.

In table 6.5 results are presented for the sensitivity analysis. Results for PAPNET are presented as a difference to the conventional microscopy situation. In case PAPNET also saves on 'other activities' (excluding clerical activities), costs per smear increase less, namely by $2.38 (range $1.79-2.88 depending on the rate of time saving). When initial screening time in a conventional

microscopy situation is 50% higher than observed, savings on cytotechnician's costs per smear are also higher than in the baseline situation, but not enough to outweigh the additional costs for the PAPNET technology.

6.4 Discussion

We conclude that personnel cost savings are likely with PAPNET, but that they are outweighed by the additional investments in review stations and scanning costs under a broad range of assumptions. The costs for PAPNET were assessed for a best case situation in which a scanning station is available in the laboratory, avoiding the packing and unpacking of slides and transportation cost.

How do our results compare with other studies? Only one cytotechnician reached a screen rate that was 2 times faster than with conventional microscopy, which is still considerably lower than rates of 3-4 times faster as published elsewhere [83 120]. Crucial however, is that we included additional microscopy in our time measurements of PAPNET-assisted screening, compared Papnet and conventional microscopy in a comparable study situation, and considered abnormal smears in a proportion observed in a routine screening population. We observed an average initial screening time of 100 seconds if no additional microscopy was required (normal smears), which is in line with other observations of 1-2 minutes [120]. In addition, although the cytotechnicians in the study were relatively unexperienced with PAPNET, we only observed a small (10%) decrease in average initial screening time between month three and nine of the study through learning effects (mainly reduced usage of microscopy). Each cytotechnician had screened several hundreds of smears inbetween these two points of time. In the sensitivity analysis we included a 50% reduction in screening time amongst the four cytotechnicians, but this did not change our results significantly. Above all, we not only included initial screening in our assessment, but all other cytotechnicians' activities as well. In our baseline calculation, intial screening is 40% of their total working time.

PAPNET-assisted screening would be more time saving when the usage of microscopy could be reduced. For instance, in case absence of endocervical cells (ecc-) would not require additional microscopy (quick scanning) as is debated in most countries, PAPNET-assisted screening would be 30% faster (160 seconds on average) per smear than conventional microscopy, leading to additional cost savings of $0.10-0.35 per smear. However, a further reduction in the usage of microscopy, as was already observed during the study, might be compensated by a more extensive and thus longer assessment of the PAPNET video images. Besides, any further shortening of screening time

may rather influence the sensitivity and specificity of PAPNET-assisted screening than increasing the screen rate.

The time measurement for PAPNET was necessarily done in a study situation. Therefore, the time measurement of conventional microscopy was also done in the same study situation, making the initial screening time fully comparable between both techniques. It is however possible that the deviation from routine practice will be larger for conventional microscopy. In reading through a microscope, a lively work environment may cause quite some interruptions, including the need for repetitive reading of smears. This is different for PAPNET-assisted screening: communication and other environmental factors may be less disturbing for screening activities, which increases efficiency. (Note that in this study we classified all aspecific work time during initial screening under 'other activities'). Beyond this, PAPNET might streamline other smear processing activities, including quality control , but in order to capture this fully, both screen techniques have to be compared in a routine situation and preferably in the same laboratory. We doubt however that the rate of time saving due to PAPNET for initial screening and 'other activities' taken together is much larger than the highest rate assumed in the sensitivity analysis (-36%). We calculated that the number of smears per cytotechnician would have to be about threefold the number for conventional microscopy, in order to neutralize the extra costs for equipment and scanning.

We considered a situation were PAPNET-assisted screening would be introduced as a primary screening tool for smears from the national screening programme. Can the conclusions be generalized to repeat smears and other follow-up smears? The assessment of these smears differs from smears in asymptomatic women without a recent positive test. They generally take more time, not only for screening the smear but also due to necessary communication with clinicians. PAPNET will not change the related time-consuming administrative procedures, and as a result will have a smaller impact on working time.

Our calculations relate to a laboratory size of 40,000 smears per year, because we assume that by this size any scale effects favouring either PAPNET or conventional microscopy are fully captured. A smaller laboratory will only be less favourable for PAPNET-assisted screening, among others because of underuse of review stations and transporting slides to and from the scanning station which will then be remote.

In a hypothetical situation where scanning costs of PAPNET are sufficiently reduced and costs of PAPNET-assisted screening are about equal to those of conventional microscopy, additional considerations apply to the decision whether or not PAPNET-assisted screening might replace conventional

microscopy. PAPNET-assisted screening might be preferred to conventional microscopy by cytotechnicians because part of the boring and tedious work (screening of predominantly normal smears) is reduced. When cytotechnicians may be difficult to recruit, as is already occurring in the Netherlands, attractive and varied work circumstances gain importance. On the other hand, cytotechnicians' preferences will not only be based on differences in attractiveness of work, but are also related to their trust in the used technology and therefore to its performance in reducing system error.

In this study it is shown that PAPNET is less cost saving when the average initial screening time per smear for conventional microscopy is low (assuming that PAPNET does not save on clerical activities and 'other activities'). A low initial screening time per smear characterizes a situation of relatively high work load, as in the United States where minimum workloads of 80 smears per day are not uncommon, and more than once provoked by financial pressures [129]. When a negative relationship between screen rate and diagnostic effectiveness is hypothesized, the relative diagnostic performance of PAPNET may be higher in this situation. Conversely, in a situation with relatively low workloads (as in the Netherlands), the possibility to save on personnel costs with PAPNET-assisted screening is higher, but it might be more difficult to improve on the diagnostic effectiveness. This latter has been indicated in our evaluation of the PAPNET system when used as a primary screening tool [68]. It turned out to have similar diagnostic performance as conventional microscopy, and as a consequence will be equally effective in terms of health benefits. Given this performance, costs for using PAPNET (equipment, scanning and training) have to be reduced from the current $3.25 (given an optimal laboratory size) to around $0.50 per slide (baseline) or $1.50 per slide (best case for PAPNET presented here) to reach cost-effectiveness levels similar to conventional microscopy when used as a primary screening tool. Because this seems highly unrealistic, similar cost-effectiveness can only be achieved when the sensitivity and/or specificity of PAPNET is considerably improved.

Abstract

Objective To analyze for which test characteristics and costs new cytologic tests for cervical cancer screening would be at least as cost-effective as the Pap test. To compare the results with evidence about recently developed technologies.
Methods With the validated microsimulation programme MISCAN for the evaluation of cervical cancer screening policies, we calculated the incremental number of cervical intraepithelial neoplasms (CIN) detected, invasive cancers prevented, life years gained, and incremental costs and cost per life year gained for different combinations of test sensitivity, specificity, and unit cost per new screening test, compared with current Pap test screening. We calculated the unit cost threshold of any screening test, given its test characteristics, for which this test would be equally cost-effective as the Pap test. We conducted one-way sensitivity analyses to translate the findings to alternative screening settings. We conducted a literature review to determine the test characteristics and costs of automated screening systems and liquid based cytology (LBC).
Results With cancer incidence, screening policy (5-yearly screening between age 30 and 60) and attendance as observed in the Netherlands, and 80% Pap test sensitivity, the unit cost of a cytologic screening test with 100% sensitivity and specificity can be up to €9.00 more expensive than the Pap test to be at least as cost-effective as the Pap test. In settings with more intensive screening this unit cost threshold is lower, whereas in settings with lower Pap test sensitivity this unit cost threshold is higher. A literature review revealed that there is weak evidence that the ThinPrep™ LBC system, and the automated systems AutoCyte™ SCREEN and AutoPap™ are more sensitive than the Pap test, at the loss of some specificity, whereas they are €2.50 to €7.00 per test more costly than the Pap test.
Conclusion None of the current automated and LBC cytologic screening systems is more cost-effective than the conventional Pap test, except in situations where Pap test accuracy and adequacy is low.

When will new cytologic tests for cervical cancer screening be cost-effective?

7

7.1 Introduction

There is convincing evidence that routine cervical cancer screening with the Pap test has significantly reduced mortality due to cervical cancer [116 140]. However, there are continuous worries and debates on the accuracy of the conventional Pap test [6 135 197]. In the US these debates are intensified by dissatisfaction about the reimbursements by insurers that would put pressure on cytotechnologists' workload, and by legal implications of false negative diagnoses [236].

In recent years, new cytologic technologies have been developed that claim to improve the sensitivity of the Pap test. Automated systems increase the efficiency of the initial screening of slides and the rescreening of normal slides, such as the AutoPap™ 300QC (rescreening) or Primary Screening system, Papnet™ (initial screening and rescreening), and AutoCyte™ SCREEN (only initial screening). In liquid based "monolayer" systems such as ThinPrep™ and SurePath™ (formerly AutoCyte™ PREP), the cell material is prepared and filtered before it is transferred to a glass slide for interpretation. Other systems are designed to support conventional microscopy (Pathfinder™, AcCell™ Series 2000), but these have not been considered in this study. A brief description of the systems is given in the Appendix.

We investigated how much a new improved cytologic test may cost if it should not be less cost-effective than the Pap test. We analyzed the relationship between test sensitivity, specificity and cost-effectiveness for population-based cervical cancer screening, and calculated threshold values for combinations of test sensitivity, specificity and unit cost per test for which any cytologic screening test would be as cost-effective as the conventional Pap test. Also, we conducted an extensive sensitivity analysis to translate the results to different settings. The outcomes can be seen as a decision analytic framework for the assessment of any cytologic test for cervical cancer screening. Subsequently, we assessed published trials of new cytologic technologies and collected available cost information. Tentative conclusions were drawn on the (cost-)effectiveness of considered technologies compared with Pap test screening.

7.2 Materials and methods

Model description

With the MISCAN microsimulation model [102] for the evaluation of screening policies individual life histories are simulated. The course of these life histories is determined by model assumptions on demography (e.g. birth, death from other causes), incidence of preinvasive lesions, the natural history of cervical cancer (e.g. duration of screen-detectable preclinical stages), and the impact of screening (e.g. test sensitivity and specificity, attendance).

The model structure and assumptions are given in Table 7.1. Preclinical disease (i.e., the stage at which the disease is present but not yet detected) is subdivided into four stages: a preinvasive stage (corresponding to cervical intraepithelial neoplasia (CIN), and including carcinoma in situ) and three preclinical invasive stages (International Federation of Gynecology and Obstetrics definition IA, IB, and II+). The duration of these screen-detectable preclinical stages of cervical cancer and the sensitivity of the Pap test have been estimated with screening data from British Columbia [287]. The mean duration and standard deviation of the different stages are 11.8 2.2 years for CIN, 2.0 0.9 years for preclinical invasive stage IA, and 1.9 0.9 years for preclinical invasive stage IB and II+ combined. The variation of the duration of these stages between women is described by a Weibull-distribution. Not all preinvasive lesions progress to cancer. The proportion of lesions that regress to normal decreases with age [287]. The model has been validated by comparison with interval cancer data at different screening intensities [116 286]. We adapted the model to the demographic and epidemiological situation in the Netherlands, and it has been found to satisfactorily reproduce the incidence and mortality of cervical cancer from 1968 to 1992 in this country, a period that includes the start of mass screening [267]. A more elaborate description of the model has been given elsewhere [277].

Simulated screening represents organized screening for women aged 30 to 60 years at 5 year intervals, which is the current policy in the Netherlands and Finland. Screening is assumed to occur between 1993-2020 (28 years). In most western countries centrally organized or spontaneous screening for cervical cancer already occurs, and reduces the effects of current screening rounds. Therefore we assumed previous screening activities before 1993 in the baseline model, that were estimated with Dutch survey data [60 80 267]. Screening attendance is 80%, as observed in the Dutch screening practice. Attenders have a 50% higher probability of attending the next screening round than non-attenders. 10% of the target population are persistent nonattenders and are also at higher risk for cervical cancer. Future costs and health effects are discounted by a annual rate of 3% [96].

Table 7.1 Structure and assumptions of the MISCAN microsimulation model for the evaluation of cervical cancer screening.

	baseline	sensitivity analysis	
Cumulative lifetime incidence of cervical cancer	.0106 - .0235*	.0212 - .0458	
Mean (SD) duration preclinical stages in years		--	
Cervical neoplasia (CIN)	11.8 (2.2)		
Micro invasive (FIGO IA)	2.0 (0.9)		
Macro invasive (FIGO IB)	1.9 (0.9)		
Sensitivity conventional Pap test	CIN 80%	60%	70%
	FIGO I 85%	70%	77.5%
	FIGO II+ 90%	80%	85%
False positive women who are kept under surveillance with repeat smears	14/1,000		
False positive women who are diagnosed normal (no CIN) after referral for colposcopy.	1.2/1,000		
Specificity conventional Pap test	98.5%	99.0%	90.0%
Inadequate screening smears	1.0%		
Screening policy: lifetime screenings (interval in years) age group	7 (5) 30-60 (NL)	10 (5) 20-65 (UK) [200] 17 (3) 18-66 (US) [264] 27 (2) 18-70 (Aus) [63]	
Screening attendance	80%	60%	
Persistent nonattenders	10%	--	
Attendance subsequent round, attenders	94.4%	--	
Attendance subsequent round, nonattenders	44.4%	--	
Discounting costs and health effects	3%	0%, 5%	
Cost per screening (€)	31.53	39.22	
invitation	0.99	--	
sampling	9.14	--	
cytologic examination †	15.38	23.07	
other (patient costs, administration)	6.01	--	
Fixed programme costs per year (€ mln)	3.40	--	
Cost of diagnosis and treatment (€)		--	
cytologic surveillance and no CIN	58		
referral and no CIN	499		
CIN	2,006		
FIGO IA	5,468		
FIGO IB	11,596		
FIGO II+ (screen-detected)‡	10,932		
FIGO II+ (clinically detected)‡	9,991		
Cost of treatment and palliative care for advanced disease (€)		--	
<50 years	31,704		
50-70 years	22,598		
>70 years	9,620		

* Cumulative background incidence for progressive preinvasive cervical cancer develops from 0.0229 for those born between 1889-1918, 0.0235 for those born between 1919-1928, 0.0128 for those born between 1929-1938, 0.0106 for those born between 1939-1948, and 0.0148 for those born from 1949 onwards.

† The costs of cytologic examination decrease with laboratory size, and therefore with the number of screening smears. The figure represents the average costs in the simulated screening period.

‡ The difference in costs between screen-detected and clinically diagnosed stages II+ cervical cancer relates to the less favourable stage-distribution of the latter, with its associated lower number of radical hysterectomies and higher number of (cheaper) radiotherapeutic treatments.

Costs

Costs of screening (organization, invitations, smear taking and review), diagnosis and treatment by disease stage are based on real resource use. Details on the data sources and calculation were presented elsewhere [267 271 272 277]. We collected costing information of liquid based cytology (LBC) and automated screening systems from chapter 6 and the literature [20 40 83 99 113 142 174 187 236 241 252].

Costs are presented in € of the year 2002 (€ = $0.95 in 2002 and $PPP 1.08 in 2001).

Model simulations, test sensitivity, specificity, and adequacy

Test sensitivity is equal to 1 – false negative rate, and is defined as the probability that an underlying preclinical lesion is detected by screening and subsequently treated. We distinguished stage-specific Pap test sensitivities: 80% for the preinvasive stages cervical neoplasia (CIN), 85% for the micro-invasive (FIGO IA and IB) and 90% for the macro-invasive stage (FIGO II+). These sensitivities resulted from an analysis of screening data from British Columbia [287], and are consistent with the findings of a recent meta-analysis [197]. Test specificity is equal to 1 – false positive rate. False positives include women with an abnormal screening smear but ultimately do not need to be referred for colposcopy, and women who are referred for colposcopy and have a normal histological diagnosis. In the Netherlands, these groups constitute 14 and 1.2 out of 1,000 screened women, respectively, corresponding to a 98.5% test specificity (derived from the Dutch Network and National Database for Pathology, PALGA).

We simulated a situation without screening, a situation with Pap test screening, and a screening situation with a new test. We optimized the sensitivity of the new test from 80% to 100% in five equal steps, representing a reduction in false negatives of 20%, 40%, 60%, 80% and 100%. In a separate analysis we optimized the test sensitivity only for the preclinical invasive stages, to account for the finding that new technologies might particularly do better in detecting invasive cancer [130]. The specificity of the Pap test and a new screening test are similar at baseline, and were varied from 99% to 90% in concordance with the proportions of mildly abnormal smears currently observed in several European countries with population based screening [117]. A lower or higher false positive rate of a new screening test will result in lower or higher costs, respectively, because of induced diagnostics, apart from the anxiety and discomfort for false positive women.

For each combination of test characteristics we calculated the incremental number of detected CIN, invasive cancers, life years gained,

screening smears, diagnostic and therapeutic procedures, and incremental costs, compared with current Pap test screening. The health effects are counted until all women who could have benefited from the screening programme have died. Because for each life history the transitions from one disease stage to another are modelled as realizations of a probability distribution, the model outcomes are subject to random fluctuation, that has been reduced by simulating a large population of 45 million women.

Cost-effectiveness thresholds
We calculated the unit cost threshold of any new cytologic test for which the incremental cost-effectiveness (C/E) of screening with the new test compared with Pap test screening would be equal to the incremental C/E of Pap test screening. The incremental C/E of a new test is equal to the difference in costs of screening with the new test compared with Pap test screening, divided by the difference in life years gained [96]. The incremental C/E of Pap test screening was calculated as the difference in costs and life years gained of the baseline screening policy (7 lifetime screenings in women aged 30-60 years at 5 year intervals) compared with a slightly less intensive screening policy with 6 lifetime screenings in the same age group (6 year intervals). The incremental C/E of alternative screening policies in the sensitivity analysis (see also below) was derived from the incremental C/E of efficient screening policies with a similar number of lifetime screenings, published by Van den Akker *et al.* [277]. This study showed that the incremental cost per life year gained increases with the number of lifetime screenings offered. For instance, the incremental C/E of a screening policy with 7 lifetime screenings offered at an efficient schedule is about $17,000 per life year gained compared with 6 lifetime screenings, and is about $35,000 per life year gained for a screening policy with 15 compared with 12 lifetime screenings. These values have been interpreted as an acceptable C/E threshold for implementing a new screening test. Alternatively, we made calculations with arbitrarily chosen C/E thresholds for each simulated screening policy.

Sensitivity analysis
In a one-way sensitivity analysis, we varied a number of crucial parameters that were expected to influence the results, and that describe alternative epidemiological or screening situations, such as cervical cancer incidence, Pap test accuracy, screening policy, and screening attendance. We reduced Pap test sensitivity for CIN to 60% and 70% (baseline 80%), in line with the outcomes of a large meta-analysis of Pap test accuracy [82]. A lower Pap test sensitivity will increase the number of missed cases, that might be detected by screening with a

more sensitive test. Because the Dutch screening policy is conservative compared with most western countries, we also simulated current screening policies and recommendations with more lifetime screenings offered at shorter intervals and in broader age ranges. The potential health benefit of a new screening test is determined by the programme sensitivity, or the ability to detect an abnormal lesion with repeated screenings, rather than by the single test sensitivity. Some missed cases may be detected timely at the next screening round, thereby reducing the unfavourable consequences of a single false negative test. The programme sensitivity will be higher with more intensive screening.

Literature review

We assessed the diagnostic accuracy of available new cytologic screening tests by a literature review. We selected studies that were included in recent meta-analyses on the cost-effectiveness of new cytologic methods [37 40 71 196 197]. Because these meta-analyses included studies until October 1999, we searched Medline for published studies until September 2003 with keywords 'vaginal smears', 'cervical intraepithelial neoplasia' or 'cervix dysplasia', in combination with 'diagnostic errors' or 'sensitivity and specificity', and in combination with 'papnet', 'autopap', 'thinprep', 'autocyte', 'liquid based', 'monolayer' or 'pathfinder'. We also searched the websites of the manufacturers for additional published studies. Duplicate publications with interim results of the same trial and studies that were described in abstract format only were discarded. We included studies in which a) conventional Pap test screening and another screening modality were compared, b) histological follow-up or an expert panel of cytotechnologists or pathologists was used as reference test, and c) the same reference test was used for both screening modalities.

Studies were assessed on the following key characteristics: study population, study design, reference test used, blinding protocol, length of the follow-up period, possible sources of verification bias, and results. In all but 7 studies [16 50 68 85 225 244 245] (a random sample of) negative smears were not verified, and absolute sensitivities and specificities could therefore not be calculated in all studies. Therefore, relative sensitivity (sensitivity new screening test / sensitivity Pap test) and relative specificity (specificity new screening test / specificity Pap test) were calculated if possible, assuming that non-verified smears are negative. We calculated the relative sensitivity and specificity for two diagnostic thresholds of the screening test and the reference test: low-grade (LSIL+ or CIN 1+) and high-grade squamous intraepithelial lesions (HSIL+ or CIN 2+). For trials with a between subjects design, in which the screening tests are applied in different study arms, relative sensitivity was

calculated in two ways if possible: as a ratio of both sensitivities with the respective number of verified cases in the numerator, and as a ratio of the detection rates, with verified test positive cases in the numerator and the respective populations in the denominator. The between subjects design is to be distinguished from the more superior within subjects design, by which both screening modalities are applied to the same women. Both designs can be used for the evaluation of monolayer systems. In case of a within subjects design a split-sample technique is used for material collection, by which the remaining cells are used for a monolayer slide. The within subjects design is common for the evaluation of AutoPap and Papnet, because the same smear can be used for conventional and automated screening.

Table 7.2 Health effects and costs (€) per 1,000,000 women per year of screening: Pap test (baseline assumptions) and optimal test sensitivity. Screening is for women between 30 and 60 years at five year intervals, from 1993-2020. Numbers between () are absolute differences with baseline results.

	Screening with Pap test	Screening with new test			
	baseline assumptions *	100% sensitivity for invasive stages†		100% sensitivity for pre-invasive and invasive stages†	
Effectiveness					
screen-detected CIN	765	765	(+0)	864	(+99)
screen-detected invasives	8.5	9.9	(+1.4)	6.8	(-1.8)
clinically detected invasives	-76.6	-78.0	(-1.4)	-83.1	(-6.5)
life years gained	811	823	(+12)	888	(+78)
life years gained (3% discounting)	243	247	(+4)	266	(+22)
Costs (€ mln)					
screening and surveillance	2.62	2.62	(0.00)‡	2.62	(-0.01)‡
diagnosis and primary treatment CIN	1.53	1.53	(0.00)	1.73	(+0.20)
diagnosis and primary treatment screen-detected invasive cancers	0.07	0.08	(+0.01)	0.05	(-0.01)
diagnosis and primary treatment clinically detected invasive cancers	-0.82	-0.83	(-0.01)	-0.89	(-0.07)
palliative and terminal care	-0.72	-0.73	(-0.01)	-0.78	(-0.07)
total costs	2.69	2.68	(-0.01)	2.73	(+0.04)
total costs (3% discounting)	2.17	2.16	(-0.01)	2.24	(+0.07)

* Pap test sensitivity is 80% for preinvasive stages (CIN), 85% for the micro-invasive stages (FIGO IA and IB) and 90% for the macro-invasive stages (FIGO II+). Baseline results are reported as differences to a situation without screening.
† Test specificity is similar to Pap test.
‡ If the optimal screening test is as costly as the baseline test.

7.3 Results

Incremental costs and health outcomes of increased test sensitivity

An increase in test sensitivity for preinvasive stages prevents much more cancers and gain much more life years than an increase in test sensitivity for invasive stages only (table 7.2). The incremental number of life years gained increases less than proportionate with test sensitivity. In case of optimal sensitivity (100%) for preinvasive stage CIN, test sensitivity increases with 25% (=20/80%), the number of screen detected CIN increases with 13% (+98.6 per 1,000,000 women per year of screening) and the number of life years gained increases with 10% (+78 per 1,000,000 women per year of screening). This is because the programme sensitivity of Pap test screening is higher than the single test sensitivity. The number of women with CIN that are additionally detected by a more sensitive screening test is larger than the number of prevented invasive cancers, mainly because most neoplasms will not progress to cancer. The remaining invasive cancers with optimal test sensitivity are a mixture of cancers in women who are too young or too old for screening, interval cancers (fast growing neoplasms), and cancers in women who did not attend the former screening round or who never attend screening.

Figure 7.1 Combinations of sensitivity, specificity, and extra unit costs for which a new screening test would be as cost-effective as Pap test screening. The false negative rate is decreased for all preclinical stages and for preclinical invasive stages, respectively. Screening is for women aged 30 to 60 at 5 year intervals. Future costs and life years gained have been discounted at 3%. The shaded areas are the observed ranges of sensitivity (LSIL+/HSIL+ threshold, see table 7.5) and unit costs for ThinPrep™ and AutoPap™.
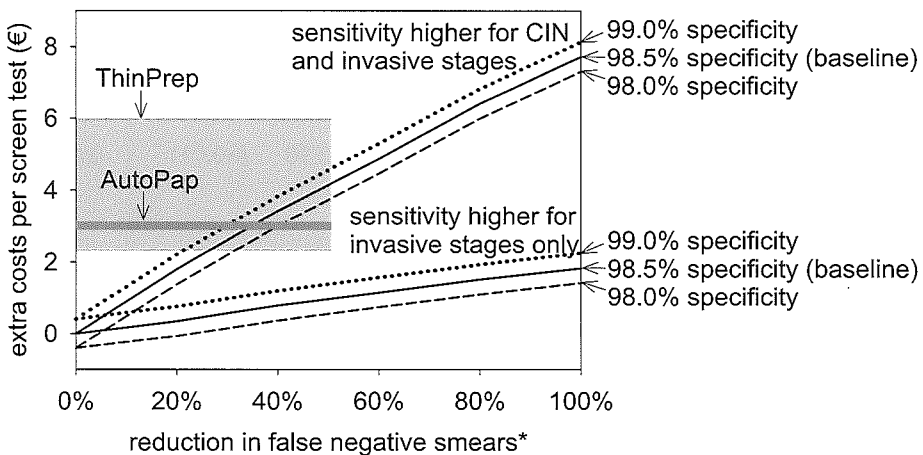
* Corresponds to a test sensitivity for CIN of 80%, 84%, 88%, 92%, 96%, and 100%, respectively.

The cost increase due to the increased detection of preinvasive lesions that need further diagnosis and treatment slightly outweighs the cost savings due to prevented invasive cancers. However, when future costs are discounted the cost increase is much higher, because the costs (management of preclinical stages) precede the savings (prevented cancers) with several years.

Figure 7.2 Combinations of sensitivity, specificity, and various levels of extra unit costs for which a new screening test would be as cost-effective as Pap test screening. The false negative rate is decreased for all preclinical stages. Screening is for women aged 30 to 60 at 5 year intervals. Future costs and life years gained have been discounted at 3%.
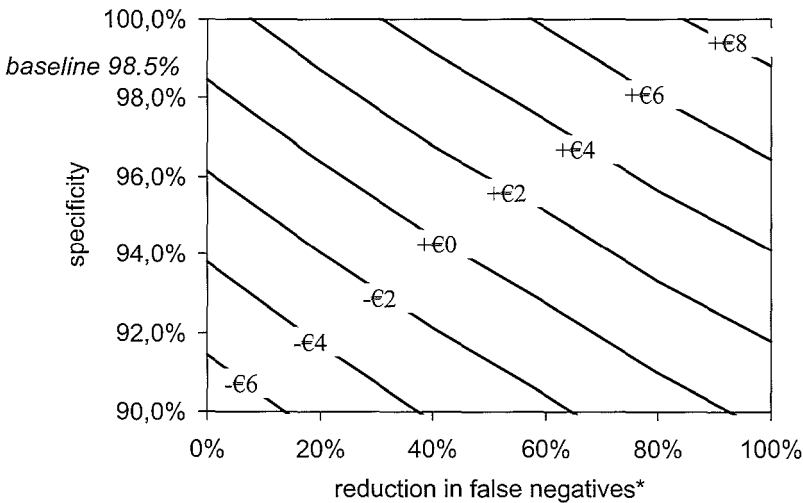
* Corresponds to a test sensitivity for CIN of 80%, 84%, 88%, 92%, 96% and 100%, respectively.

Unit cost thresholds of a new screening test
The incremental cost-effectiveness (C/E) of the baseline screening policy (7 lifetime screenings) is about €19,000 per life year gained compared with a programme with 6 lifetime screenings ((€2.17 - €1.93 million) / (243 - 231 life years gained) = €19,000; figures are per 1,000,000 women per screening year). In figure 7.1 the unit cost threshold of a new screening test is plotted against an increased test sensitivity (reduced false negative rate). In case of 100% sensitivity (0% false negatives) for all preclinical stages, a new screening test is

allowed to be €7.70 more expensive than the Pap test, and €2.80 more expensive if the 100% test sensitivity only applies to preclinical invasive stages.

Figure 7.2 shows isocurves representing combinations of test sensitivity and specificity for which a new cytologic screening test would be as cost-effective as the conventional Pap test, for different unit costs per new screening test. For example, a new test with similar costs but a specificity of 96.0% should compensate this with a 25% reduction of false negative smears in women with preclinical stages of cervical cancer, to be as cost-effective as Pap testing. Each percentage decrease in test specificity would decrease the unit cost threshold of a new screening test with €0.80, and vice versa. So if the current test specificity deteriorates from 98.5% to 90.0% by implementing a new screening test (as in some current situations), this would decrease the unit cost threshold of a new screening test with about €7.

Sensitivity analysis
A number of assumptions were varied in the sensitivity analysis (table 7.3). The potential health impact of a more sensitive screening test decreases with a more intensive screening policy, but the incremental C/E of this policies would be higher. As a result, the unit cost threshold of a new screening test with 100% sensitivity and similar specificity is slightly higher for the UK screening policy compared with the baseline policy, but lower for the more intensive policies in the US and Australia (17 and 27 lifetime examinations, respectively).

In table 7.4, the unit cost thresholds of a new screening test are presented for arbitrary C/E thresholds in stead of the C/E thresholds of current Pap test screening in table 7.3. For any given threshold, the unit cost threshold for a new screening test decreases with the intensity of screening. Similarly as in table 7.4, in figure 7.3 the unit cost thresholds for a new screening test are given for different levels of test sensitivity and incremental cost per life year gained. For instance, at a threshold of €20,000 per life year gained, a new screening test can be up to €3.50 more expensive than the Pap test if the false negative rate can be decreased with 40% for all preclinical stages (88% sensitivity for CIN).

The unit cost threshold for a new screening test increases to €13.80 and €10.60 if the current Pap test sensitivity would be 60% or 70% for CIN, respectively (table 7.3). However, when the current Pap test sensitivity is lower, a reduction in false negatives as used in the calculations corresponds with a higher absolute increase in test sensitivity. All other alternative assumptions have a relatively modest impact on the results.

Table 7.3 Sensitivity analysis. Extra costs of diagnosis and treatment and additional life years gained per 100,000 screenings, and incremental cost per smear for which new cytologic techniques are equally cost-effective as conventional cytology. Future costs and life years gained are discounted at 3%. Baseline and alternative model assumptions (table 7.1).

Model assumptions	Pap test screening [a]	Screening with optimal test [b]		
	Incremental cost per life year gained (€) (1)	Net incremental costs of diagnosis and treatment of (pre)clinical stages (€1000) (2)	Incremental life years gained (3)	Unit cost threshold, difference with Pap test (€) [c]
Baseline assumptions	19,000 [d]	155	49.9	+7.70
Screening 20-65 years at 5-year intervals (UK) [200]	27,100 [e]	157	37.0	+8.20
Screening 18-66 years at 3-year intervals (US) [264]	57,300 [e]	65	12.5	+6.30
Screening 18-70 years at 2-year intervals (Australia) [63]	122,400 [e]	31	5.1	+5.80
Cancer incidence twofold	10,900 [d]	318	99.6	+7.50
Pap test sensitivity 60%	13,900 [d]	343	126.7	+13.80
Pap test sensitivity 70%	15,800 [d]	246	84.5	+10.60
Screening attendance 60%	13,200 [d]	164	65.1	+6.70
Discounting 0%	7,400 [d]	63	119.5	+8.00
Discounting 5%	31,900 [d]	253	30.6	+7.00
Unscreened population	13,600 [d]	190	68.3	+7.20

a Pap test sensitivity is 80% for preinvasive stages (CIN), 85% for the micro-invasive stages (FIGO IA and IB) and 90% for the macro-invasive stages (FIGO II+).

b Test sensitivity 100% for all preclinical (pre)invasive stages, specificity similar to Pap test.

c Calculation: ((1) x (3) - (2) x 1000) / (100,000 x (1 + 0.010 + 0.015 x 1.16)). Proportion of inadequate screening smears = 0.010, proportion of false positive screening smears = 0.015, average number of repeat smears per woman = 1.16.

d Incremental C/E ratio of screening between 30-60 years with 7 lifetime screenings (5 year intervals) compared with screening between 30-60 years with 6 lifetime screenings (6 year intervals).

e Corresponding to the incremental C/E ratio of efficient screening policies with corresponding numbers of lifetime screenings in Van den Akker-van Marle *et al.* [277].

Figure 7.3 Incremental costs per life year gained of a new screening test compared to Pap test screening, for different combinations of test sensitivity and extra costs per screen test (no change in test specificity assumed). Screening is for women aged 30 to 60 at 5 year intervals. Future costs and life years gained have been discounted at 3%.
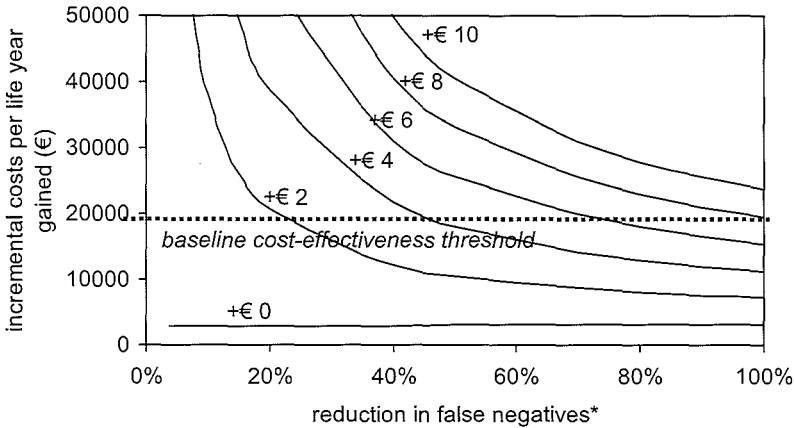
* Corresponds to a test sensitivity for CIN of 80%, 84%, 88%, 92%, 96% and 100%, respectively.

Table 7.4 Unit cost threshold (difference with Pap test) of a new screening test (in €), for different cost-effectiveness thresholds and screening policies.

Screening policy*	Cost per life year threshold		
	€20,000	€50,000	€100,000
100% reduction in false negatives†			
7 (5) 30-60 (NL, Finland)	+8.20	+22.80	+47.00
10 (5) 20-65 (UK) [200]	+5.70	+16.50	+34.50
17 (3) 18-66 (US) [264]	+1.80	+5.50	+11.50
27 (2) 18-70 (Australia) [63]	+0.70	+2.20	+4.70

* Lifetime screenings (interval in years) age group
† 100% sensitivity for all preclinical stages. Pap test sensitivity is 80% for preinvasive stages (CIN), 85% for the micro-invasive stages (FIGO IA and IB) and 90% for the macro-invasive stages (FIGO II+).

Literature review

Of the studies on the accuracy of new cytologic tests that were identified, 15 considered ThinPrep™ (LBC) [25 46 50 64 85 100 114 184 187 213 214 234 244 245 295], 7 considered SurePath™ (LBC) [16 19 21 107 181 187 288], 1 considered AutoCyte™ SCREEN [181], 4 considered AutoPap™ [49 216 255 299], and 8 considered Papnet™ [29 68 120 127 136 225 237 246]. ThinPrep™ slides have a higher sensitivity than conventional Pap tests in most studies, at each diagnostic threshold (table 7.5). The relative sensitivity ranges from 0.95 to 1.22

for the LSIL+/LSIL+ threshold (screening test / reference test), and from 0.97 to 1.12 for the LSIL+/HSIL+ threshold when only split-sample studies are included. The higher test sensitivity should be weighed against a lower specificity as reported in most studies. The relative sensitivity of SurePath™ slides ranges from 0.93 to 1.00 at the LSIL+/HSIL+ diagnostic threshold when only split-sample studies are considered. In studies with a between subjects design the observed relative sensitivity was generally (much) higher. However, in the far majority of LBC studies the results were possibly biased in favour of the monolayer systems due to flaws in the study design. First, in three split-sample studies consensus opinion on cytology was used as a gold standard and not (or partly) histology [107 184 245]. Therefore, it is not known whether the additional detected cytologic abnormalities would have been confirmed by histological assessment. Second, the histological verification was not blinded in all but one study in those with histology as reference test [85]. In one study even the smear assessment was not blinded between both modalities [244]. Third, in none of the studies with a between subjects design were participants randomly allocated to the study arms. The prevalence rates may therefore differ, which may completely explain the higher detection rates by LBC. The incomparability of the study arms is confirmed by the high differences in detection rates in these studies. Also, in those studies that used consensus cytology as reference test, blinding with respect to the screening modality was not possible by definition. Fourth, in studies with histology as reference test, verification bias [229] may have occurred when histological follow-up is only available for a (small) proportion of positive smears, or when the follow-up rate was lower for LBC than for conventional microscopy. Fifth, when cytologic examination with the new technology takes place in a different laboratory, or even in a different country [114], a comparison of results is seriously hampered because of possible differences in laboratory performance. And lastly, in one study the relative specificity could not be calculated, in order to check whether the increased sensitivity was associated with a loss of specificity [245]. Two split-sample studies with sufficient blinding and complete histological follow-up reported a higher [85] and lower test sensitivity [50], respectively, for ThinPrep™. In the UK Pilot, a higher detection rate (+24%) for histology confirmed HSIL+ was observed in regions with ThinPrep™ compared with the previous year. This increased rate was observed in age group 20-34 only, and might be due to increased uptake of previously unscreened women [187].

In several retrospective rescreen studies by which the Papnet™ system was evaluated, the so-called 'rescreen effect' [32] has likely favoured the test performance of Papnet™ relative to conventional screening. [29 120 127 136 237]. The rescreen effect means that the detection of abnormal smears is

increased when the assessment does not have clinical consequences, and/or when samples are used with a relatively high proportion of abnormal smears, resulting in a higher alertness. An evaluation of Papnet™ has indicated that the rescreen effect fully explains the increased detection of abnormal smears [268]. Trials on Papnet™ with designs that excluded rescreen effects reported similar sensitivities [68 225] or even a lower sensitivity [246] for Papnet™ compared with conventional Pap tests.

Three out of four studies reported slightly higher sensivitities for AutoPap™ compared with conventional screening (relative sensitivity up to 1.12, LSIL+/HSIL+ threshold) [49 216 299]. From one of these studies the relative specificity could be calculated (1.005) [299]. In one study only the sensitivity of rescreening normal smears was given [216]. Therefore the relative sensitivity for the initial screening and rescreening procedure combined could not be determined but was greater than one. The three studies were rescreen studies, and the rescreen effect might have biased the results in favour of AutoPap™. In all AutoPap™ trials expert judgement of cytology was used as a reference test instead of histology. In two studies the sensitivity refers to the system sensitivity, which is the proportion of smears that were assessed abnormal by the expert panel and that were subsequently identified by the system [49 216]. This system sensitivity is considered an upper bound for the subsequent manual screening.

One study evaluated the AutoCyte™ SCREEN system, and found a relative sensitivity of 1.44 at the cost of a small loss in specificity [181]. However, also in this study the rescreen effect has likely influenced the results in favour of AutoCyte™ SCREEN.

Costs of new cytologic screening tests
Of the costing data that were collected, the minimum difference in cost per test compared with the Pap test are reported here. For the ThinPrep™ 2000 system €2.70 higher costs per slide have been reported for a large-scale laboratory, including €2.45 for consumables and an assumed 30% increase in slides that can be read per day [174]. The recent UK Pilot reported €6.95 and €6.61 higher lab costs per slide for the ThinPrep™ 2000 and 3000 system, respectively, including savings due to more efficient reading, whereas savings of 1 (€0.92) to almost 5 minutes (€4.34) per slide were reported for smear taking [187]. For SurePath™ $2.33 [19] and €3.10 [187] higher costs per slide have been reported, accounting for more efficient reading of slides. The difference in costs of Papnet™ was estimated at $2.85 when used as a primary screening tool (chapter 6) and at $6.80 when used as rescreening tool and assuming $5.00 for processing per slide [99]. AutoPap™ is $3.30 per slide more expensive when used as rescreening

tool, and $3.15 when used as primary screening tool, including product costs of $4-5 per slide [99]. No cost information was found for the AutoCyte™ SCREEN system.

Integration of results
The observed relative sensitivities up to 1.12 for ThinPrep™ (split-sample studies) and AutoPap™ if we assume an LSIL+/HSIL+ threshold, would correspond to a test sensitivity of up to 90% (=1.12 x 80%). We take this diagnostic threshold because HSIL+ lesions are reason for treatment, and LSIL+ smears get follow-up in most countries (the definition of ASCUS smears is too aspecific for cross-study comparison). It can then be derived that these technologies would only be cost-effective on the lower range of the observed unit costs, and when the maximum observed sensitivity is assumed (figure 7.1). This does not yet account for the fact that ThinPrep™ showed worse specificity in most studies, whereas this could not be determined for AutoPap™.

7.4 Discussion
We analyzed the relationship between test accuracy, life years gained and unit cost per test for cervical cancer screening to compare the possible cost-effectiveness of new screening tests with that of the conventional Pap test. We used a validated simulation model that describes the natural history of cervical cancer. Within a realistic range of assumptions we calculated the diagnostic accuracy and cost levels at which any cytologic screening test would be acceptable for population based screening compared with the conventional Pap test. There is weak evidence that ThinPrep™ smears and the AutoPap™ system provide higher sensitivity than conventional Pap tests, whereas the impact on specificity is uncertain. However, even with favourable assumptions, their diagnostic accuracy appears to be below acceptable cost-effectiveness thresholds.

Limitations
Apart from avoided mortality from cervical cancer, new screening tests may also influence quality of life. A possible reduction in the number of invasive cancers has to be weighed against the increased diagnosis and treatment of preinvasive lesions. (There are no indications that current new technologies will reduce the number of false positive smears and associated anxiety and discomfort.) However, the balance between positive and negative effects of screening is yet undetermined because no valid estimates are available on the quality of life implications of surveillance, referral, treatment and invasive cervical cancer.

In the present study we only considered screening tests with better characteristics than the conventional Pap test, analogous to the claims of newly developed technologies and the debate on the limitations of the conventional Pap test. However, it is possible that a screening test with worse test characteristics than the Pap test but with a much lower price, similar to e.g. visual examination, would be attractive for screening policies in low resource settings.

We calculated cases detected and life years gained assuming a screening test sensitivity up to 100%, to calculate an upper bound for acceptable costs per test. However, a test sensitivity of 100% is unrealistic because new tests may avoid that relevant cells are overlooked (screening error) or misinterpreted (interpretation error), but they cannot solve the problem of sampling error.

The model sensitivity of the Pap test for the preinvasive stages is comparable to the findings of a recent meta-analysis on Pap test accuracy, who found a median sensitivity of 83% at a cytologic LSIL+ threshold for the detection of HSIL+ lesions [197]. In the Netherlands, these diagnostic thresholds have clinical relevance for follow-up and treatment respectively. Our baseline sensitivity of 80% was derived from screening data of British Columbia, and resulted in the best model fit with the screening data. Nevertheless, we also assumed a lower Pap test sensitivity of 60% or 70%, that comes closer to the findings of another meta-analysis [82]. Pap test sensitivity might vary among countries and even laboratories, and the contribution of a more sensitive screening instrument may therefore be higher in some settings than in other.

We calculated combinations of test characteristics and cost per slide for which the incrmental C/E of a new cytologic screening test would be similar to that of current Pap test screening. The C/E of current cervical cancer screening will be an important normative threshold for any policy change, whether being a change in the screening schedule, specific policies for high-risk women, or the implementation of a new screening instrument. This approach is different from Myers [196] and Brown [40], who calculated the incremental C/E of new screening instruments given their costs and test accuracy, but without considering this normative threshold.

Comparisons with other studies
Are the outcomes comparable with other studies? Brown calculated the incremental C/E ratios for AutoPap™ (rescreening), Papnet™ (rescreening) and ThinPrep™ (primary screening), with favourable results for AutoPap™ [40]. An important reason for this favourable outcome is a high estimate of the sensitivity of AutoPap™, that was based on results from one trial [49]. Brown

multiplied the sensitivity of AutoPap™-assisted rescreening with the assumed false-negative rate of the Pap test (20%) in the model instead of the much lower false-negative rate in the trial (12.8%). When estimated correctly, the sensitivity of AutoPap™-assisted rescreening will be 1.11 (relative sensitivity in the trial, 20% rescreening cut-off) x 80% (model sensitivity) = 89% instead of the 95.4% estimated by Brown. Moreover, the trials from which the test accuracy of AutoPap™, ThinPrep™ and Papnet™ was derived, showed serious flaws in study design that could have biased the results in favour of the new screening technologies [29 49 120 136 144 237 244 247]. Of the trials with unbiased study designs, three have shown similar test sensitivities for ThinPrep™ and Papnet™ compared with the Pap test [50 68 225], and one study showed a higher test sensitivity [85].

A second reason for the favourable outcome for AutoPap™ in Brown *et al.* is that some model assumptions were favourable for a more sensitive test. He assumed an exponential distribution of the duration of preclinical stages [73]. Compared with a Weibull-distribution as in our model, this generates more rapid growing lesions. These lesions have only one opportunity to be detected by screening, and an exponential distribution is thus favourable for a more sensitive test. However, a Weibull distribution gave a better fit with screening data [287]. Other assumptions of Brown *et al.* were no previous screening, no hysterectomies, and a higher background incidence of invasive cancer, that all result in higher estimates of life years saved (table 7.3). When we applied similar assumptions, but used a Weibull-distribution of the duration of preclinical stages, we estimated an increase in life expectancy of 0.21 days by using AutoPap™ compared with 0.96 days in Brown *et al.*, for a three year screening interval (3% discounting).

Myers [196]. estimated the cost per life year gained of the Pap test and a hypothetical screening test, with 51% and 99% sensitivity, respectively, and equal unit costs, at $2,853 and $2,919, respectively, for 5-yearly screening between age 15 and 85 compared with no screening. The Pap test sensitivity of 51% is rather low compared with our own estimate and findings reported elsewhere [40 82 197]. The cost per life year gained are lower than our estimates, which can partly be explained by more favourable assumptions for screening: a longer average duration of the preinvasive stage (about 20 years compared with 11.8 years in our model), higher costs of cancer treatment, a previously unscreened population, and 100% screening attendance including women at high-risk of cervical cancer.

Smith [252] predicts that primary screening with AutoPap™ saves costs, despite $4.50 higher cost per test, assuming 9.4% additional test sensitivity for all preclinical stages and lifetime screening. Their results are

counterintuitive, because in their model health benefits and net cost savings of AutoPap™ accumulate with increasing screening intensity, whereas it is expected that at higher screening intensities a more sensitive test will prevent less additional cancers and save less costs.

Schechter [241] estimated cost per life year gained for Papnet™ at $12,194 in a situation of screening women 20-64 years at 5 year intervals, assuming 3-5% additional sensitivity for SIL, a lower test specificity, and $7 incremental cost per test. Their results compare favourably with ours. This can partly be attributed to assumptions favourable for screening, such as a previously unscreened population, and 100% screening attendance.

Other considerations

Trials and tentative calculations on the cost-effectiveness of new screening technologies [113] use intermediate end points (e.g. abnormal cytology or histology) as a proxy for outcome, because mortality as an outcome measure would need huge trials and a long follow-up period. The present study is a clear example that simulation modelling can help to translate these intermediate measures into mortality reduction, taking into account the natural history of cervical cancer, screening attendance, individual test sensitivity, and programme sensitivity. Moreover, modelling is a tool by which (cost-)effectiveness of screening can be calculated for any screening context.

It has been argued that liquid based cytology (LBC) will reduce the proportion of inadequate smears, and therefore save costs and discomfort [187]. The evidence on this matter is equivocal, with some studies reporting higher or similar proportions of inadequate smears, and some studies reporting lower proportions with LBC [124]. Moreover, the proportion of inadequate smears is generally low, and in most countries with population based screening even lower than our baseline assumption of 1.0% [117]. Therefore, any cost savings will generally be relatively modest. A decrease in the inadequate rate from 1.0% (as currently in the Netherlands) to 0% would increase the unit cost threshold of a new test with only €0.32 (1% of Pap test unit cost). A major exception to this rule is the United Kingdom, where the proportion of inadequate smears was about 10%, and cost savings were a major argument to introduce LBC in their national screening programme [199]. Considered the large international variation in Pap smear quality, it would be important to assess introduction of LBC in comparison with other means to improve Pap smear quality, such as training of smear takers.

The reported cost estimates of LBC do not account for any additional costs for postal services because of special transport requirements and that may be substantial (about €1 per slide). In addition, the costs of consumables in a

situation where LBC is applied on a large scale is uncertain, and the reported costs should therefore be regarded as indicative.

Our findings have important implications for current screening programmes, because in several countries liquid based or automated systems are considered or used already for population based screening. Because in many countries screening is offered at 3 or 2 year intervals, and considered that the incremental C/E of these systems deteriorates with screening intensity (higher programme sensitivity), this implies considerable existing or potential health system inefficiencies.

A precise estimate of the incremental cost-effectiveness of liquid based and automated screening systems has partly been inhibited because unbiased estimates of test accuracy are rare. Because unbiased estimates of test accuracy were lower than in studies with weak study designs, lessons should be learned for any future trials for the evaluation of new tests for cervical cancer screening.

Human papillomavirus testing (HPV) is increasingly being considered as an alternative to cytologic screening [55]. HPV testing has been found to be more sensitive than the Pap test, but at the expense of a decrease in test specificity. As a result, the screening interval may be extended for women who are (repetitively) HPV negative, but a much larger number of screened women will need additional surveillance and diagnosis. However, a quantitative deliberation of the pros and cons of HPV testing in primary screening, either as an adjunct to cytology or as a complete substitute, is not possible before the results of current prospective randomized trials become available [273]. These are expected to become available in the next coming years. Because a conversion to LBC or automated cytologic screening, to HPV testing, or both has severe organizational implications, countries that did not yet convert to automated or LBC screening systems may probably better await these trial results.

We conclude that current liquid based and automated systems for cervical cancer screening show unfavourable cost per life year gained compared with conventional Pap test screening, even in situations with a low screening intensity, and with realistic assumptions on Pap test accuracy. This conclusion is even stronger in countries with more intensive screening policies (e.g. 2- or 3-yearly screening). Only in situations with low Pap smear quality, low Pap test sensitivy and low incremental costs per test, new cytologic screening technologies might be a cost-effective alternative. Yet, this should be weighed against other measures to improve the efficiency of cervical cancer screening, such as alternative measures to improve Pap smear quality, HPV screening, and targeting high-risk populations that are not reached in current screening

programmes. The present use of liquid based or automated technologies for population based screening in several countries is likely to be inefficient and is not evidence-based.

Appendix

The Papnet™ system (TriPath Inc.) [162] is a neural network based technology that selects the 128 most suspicious cells or cell groups from a conventional slide to be photographed. The images can be assessed from a computer screen, and if necessary through additional conventional microscopy. The Papnet™ system has been FDA approved as rescreening instrument for smears initially assessed as normal.

The AutoPap™ system (TriPath Inc.) [215] is an automated device that, after scanning of a conventional slide, assigns a score to each slide that indicates the likelihood that abnormalities are present. This score is based on an algorithm. The purpose is to form an enriched sample of slides that are to be manually screened. The AutoPap™ system has been FDA approved both as a rescreening device and as a primary screening device. When used as a primary screening device, the FDA panel recommended a 75% threshold for selecting slides that should be manually screened.

The ThinPrep™ system (Cytyc Inc.) and SurePath™ (formerly AutoCyte™ PREP) system (Tripath Inc.) are liquid-based slide preparation systems. Noncellular material (blood, etc.) and inflammatory cells are filtered before the cells are deposited in a thin layer on the slide. This facilitates the detection of abnormal cells whereas also the area on the slide that should be screened is reduced, thereby reducing the amount of review time. The ThinPrep™ system and AutoCyte™ PREP system have been FDA approved for primary screening. The AutoCyte™ SCREEN system (Tripath Inc.) [19] combines the technologies of the AutoPap™ system and the AutoCyte™ PREP system. Monolayer AutoCyte™ PREP slides are scanned by the system that gives a score indicating the likelihood that abnormalities are present.

A fuller description of these devices has been given in Rosenthal *et al.* [236].

Table 7.5 Summary statistics and design characteristics of studies on liquid based (A. ThinPrep, B. SurePath) and automated cytology (C. AutoCyteSCREEN, D. AutoPap, E. Papnet).

Diagnostic threshold screening test / reference test	N	Relative sensitivity			Relative specificity			Verified	Population	Reference test
		LSIL+ / LSIL+	LSIL+ / HSIL+	HSIL+ / HSIL+	LSIL+ / LSIL+	LSIL+ / HSIL+	HSIL+ / HSIL+			
A. ThinPrep™										
within subjects design ('split sample')										
Sheets, 1995 [244]	445	1.09	--	--	0.992	--	--	100%	women referred for colposcopy	histology
Ferenczy, 1996 [85]	364 [a]	1.11	1.06	--	0.985	0.925	--	100%	women referred for colposcopy	histology
Roberts, 1997 [234]	35,560	--	--	1.03 [b]	--	--	--	74% / 75%	screening and clinical	histology
Sherman, 1998 [245]	7,360	1.18	--	--	--	--	--	100%	high-risk screening	cytologic diagnosis by independent pathologist's masked review
Hutchinson, 1999 [114]	8,930	1.22	1.12	1.05	1.003	0.998	1.001	100%	high-risk screening	cytology or histology
Monsonego, 2001 [184]	5,428	1.18	--	--	0.996	--	--	100%	screening	most abnormal test result after panel review
Park, 2001 [214]	483	0.95	0.97	0.98	1.011	1.016	1.010	33% / 33%	women referred for colposcopy	biopsy
Coste, 2003 [50]	2,586	0.98 [h]	0.98 [h]	0.93 [h]	0.979 [h]	0.985 [h]	0.999 [h]	100%	screening and clinical	histology
between subjects design										
Bolick, 1998 [25]	39,408 (Pap) 10,694 (TP)	1.12 2.59 [c]	--	--	0.985	--	--	15% / 14%	screening	biopsy

Diagnostic threshold screening test / reference test	N	Relative sensitivity			Relative specificity			Verified	Population	Reference test
		LSIL+ / LSIL+	LSIL+ / HSIL+	HSIL+ / HSIL+	LSIL+ / LSIL+	LSIL+ / HSIL+	HSIL+ / HSIL+			
Papillo, 1998 [213]	18,569 (Pap) 8541 (TP)	-- 1.61 [c]	-- 1.57 [c]	1.09 1.71 [c]	0.996	0.994	0.999	70% / 66%	screening and clinical	biopsy
Carpenter, 1999 [46]	4,660 (Pap) 2,727 (TP)	-- 1.33 [c]	-- 1.04 [c]	1.00 1.04 [c]	0.993	0.982	0.996	47% / 55%	clinical	biopsy
Diaz-Rosario, 1999 [64]	74,573 (Pap) 56,095 (TP)	-- 0.81 [c]	-- 0.72 [c]	0.83 0.59 [c]	0.983	0.985	0.996	95% / 47%	screening	biopsy
Guidos, 2000 [100]	5,423 (Pap) 9,583 (TP)	-- 4.14 [c]	-- 5.45 [c]	0.85 4.61 [c]	0.977	0.971	0.997	76% / 71%	screening and clinical	biopsy
Weintraub, 2000 [295]	129,619 (Pap) 39,455 (TP)	1.54 2.99 [c]	1.63 2.99 [c]	1.98 3.62 [c]	0.986	0.984	0.999	42% / 34%	screening	histology
UK Pilot, 2003 [187]	67,856 (Pap) 34,128 (TP)	--	--	1.24 [c i]	--	--	1.000	no data	screening	histology

B. SurePath™
within subjects design ('split sample')

	N									
Bishop, 1997 [19]	2032	0.93 (1.38)	1.00 (1.11)	0.38 (0.68)	0.978 (0.995)	0.979 (0.981)	1.003 (0.999)	38% / 23%	clinical	histology or repeat smear (consensus

Diagnostic threshold screening test / reference test	N	Relative sensitivity			Relative specificity			Verified	Population	Reference test
		LSIL+ / LSIL+	LSIL+ / HSIL+	HSIL+ / HSIL+	LSIL+ / LSIL+	LSIL+ / HSIL+	HSIL+ / HSIL+			
										cytologic diagnosis by external expert panel)
Bishop, 1998 [21]	8,893	1.14	--	--	0.996	--	--	26% / 24%	screening and clinical	histology or repeat smear
Minge, 2000 [181]	2,156	0.93	--	--	0.984	--	--	no data	clinical	biopsy
Hessling, 2001 [107]	2,438	0.99	1.00	1.01	--	--	--	100%	enriched sample of slides from clinical population	most abnormal cytologic result as determined by majority opinion of three pathologists
Bergeron, 2001 [16]	500	0.98 [d]	0.93 [d]	0.92 [d]	0.879 [d]	0.844 [d]	0.935 [d]	100%	women presenting for cone biopsy	biopsy
between subjects design										
Vassilakos, 2000 [288]	19,923 (Pap) 81,120 (SP)	1.03 1.37 [c]	1.00 1.41 [c]	1.21 1.71 [c]	0.980	0.979	1.000	68% / 31%	screening and clinical	histology
UK Pilot, 2003 [187]	43,280 (Pap) 21,257 (SP)	--	--	0.87 [c i]	--	--	1.000	no data	screening	histology
C. AutoCyte™ SCREEN *within subjects design ('split sample')*										
Minge, 2000 [181]	2,138	1.44	--	--	0.994	--	--	no data	clinical	histology

Diagnostic threshold screening test / reference test	N	Relative sensitivity			Relative specificity			Verified	Population	Reference test
		LSIL+ / LSIL+	LSIL+ / HSIL+	HSIL+ / HSIL+	LSIL+ / LSIL+	LSIL+ / HSIL+	HSIL+ / HSIL+			
D. AutoPap™										
Primary screening										
Wilbur, 1998/1999 [299 300]	25,124	1.08 [e]	1.05 [e]	--	1.006 [e]	1.005 [e]	--	100% of discrep ant cases	screening	judgement of discrepant cases by panel of three external pathologists
rescreening (quality control with 10% cut-off)										
Colgan, 1995 [49]	3,487	1.06	1.12	--	--	--	--	100% of discrep ant cases	smears previously assessed normal or ASCUS	judgement of discrepant cases by panel of three external pathologists
Stevens, 1997 [255]	1,840	1.00	1.00	1.00	--	--	--	100% of discrep ant cases	smears previously assessed normal	judgement of discrepant cases by panel of two internal pathologists
Patten, 1997 [216]	12,048	>1.00	>1.00	--	--	--	--	100% of discrep ant cases	smears previously assessed normal	judgement of discrepant cases by panel of three external pathologists
E. PAPNET™										
Primary screening										
Sherman, 1998 [246]	7,323	0.71	0.75	0.73	1.023	1.028	1.007		high-risk	cytology and

Diagnostic threshold screening test / reference test	N	Relative sensitivity			Relative specificity			Verified	Population	Reference test
		LSIL+ / LSIL+	LSIL+ / HSIL+	HSIL+ / HSIL+	LSIL+ / LSIL+	LSIL+ / HSIL+	HSIL+ / HSIL+			
PRISMATIC, 1999 [225]	20,008	0.98	1.00	1.01	1.095	1.088	1.010	100% of ab-normals	screening screening	histology judgement of abnormal smears and random selection of normal smears by independent panel of three pathologists
rescreening design Koss, 1994 [136]	201	only Papnet rescreen sensitivities were given.						100%	abnormal smears with biopsy-confirmed SIL+	biopsy
Boon, 1994 [29]	63	--	2.06 / 0.92 [f]	1.53 / 0.57 [f]	--	--	--	100%	false-negative smears with HSIL+ on biopsy in next screening round	biopsy
Rosenthal, 1996 [237]	62	--	1.75 [g]	--	--	--	--	100%	false-negative smears with invasive carcinoma at follow-up and normal controls	cytology or histology
Jenny, 1997 [120]	516	1.16 / 1.08	1.17 / 1.08	--	--	--	--	100%	smears of women with SIL+ on biopsy	biopsy

Diagnostic threshold screening test / reference test	N	Relative sensitivity			Relative specificity			Verified	Population	Reference test
		LSIL+ / LSIL+	LSIL+ / HSIL+	HSIL+ / HSIL+	LSIL+ / LSIL+	LSIL+ / HSIL+	HSIL+ / HSIL+			
Kaufman, 1998 [127]	160	only Papnet rescreen sensitivities were given						100%	ASCUS smears of which biopsy was available within 1 year	biopsy
Doornewaard, 1999 [68]	6,063	1.01	0.97 [f]	1.36 [f]	1.001	0.999 [f]	0.999 [f]	100%	screening and clinical	most severe diagnosis (biopsy or repeat smear) during 7 years follow-up

LSIL = CIN 1, HSIL = CIN 2/3, TP = ThinPrep, SP = SurePath
a Based on reported sensitivities and specificities, but underlying data not given.
b Includes 'inconclusive slides', i.e. high grade abnormalities cannot be excluded.
c Ratio of detection rates.
d Unsatisfactory smears have been excluded from the numerator.
e Test performance for ASCUS+ on cytology and LSIL+ or HSIL+ respectively on histology.
f Threshold of reference test is CIN III+.
g Threshold of reference test is invasive cancer.
h Results from 'optimised reading'. Relative sensitivity and specificity were higher in the screening population than in the clinical population.
i Calculated with the positive predictive value multiplied with the number of HSIL+ smears.

Abstract

Objective To evaluate the utility of high-risk human papillomavirus (HR-HPV) testing for triage of women referred for colposcopy because of abnormal smears.

Methods We considered women with persistent mild or moderate dyskaryosis and women with severe dyskaryosis who were referred for colposcopy. For both patient groups we evaluated three alternative management policies: 1. conventional management based on histological assessment, 2. HR-HPV-triage with direct treatment without prior histological assessment for HR-HPV-positive women and conventional management for HR-HPV-negative women, and 3. direct treatment without histological assessment for all referred women. For each policy the average number of medical procedures, doctor visits and the costs per referred woman were calculated. Based on a literature review, the results were tested and translated to other patient groups.

Results Per woman with persistent mild or moderate dyskaryosis and compared with conventional policy, HR-HPV-triage will avoid 0.51 colposcopically directed biopsies, but adds 0.05 local treatments of the cervix (i.c. loop excision of the transformation zone) and 0.09 outpatient visits, and will cost $134 extra. HPV triage is less efficient in women with borderline or mildly dyskaryotic cytology. In women with severe dyskaryosis, direct treatment is more efficient as conventional management or HPV triage.

Conclusion The decision to introduce HPV testing or direct treatment in women with persistent mild or moderate dyskaryosis strongly depends on the relative burden attributed to a colposcopically directed biopsy and an outpatient visit compared to LETZ treatment of the cervix. For women with severe dyskaryosis, direct treatment should be seriously considered.

Human papillomavirus testing for triage of women referred because of abnormal Pap smears

8

8.1 Introduction

Specific high-risk human papilloma virus (HR-HPV) types are associated with the occurrence and/or development of cervical neoplasia [41 165]. The value of HPV testing as a primary screening tool has been addressed elsewhere [273]. The present study asks whether HPV testing can be used as a further diagnostic tool in women with abnormal smears. This issue has been raised in several studies [52 57 58 87 119 126].

Conventionally, the histological assessment of colposcopically directed biopsy and the adequacy of colposcopy in overseeing the neosquamocolumnar junction, determine further management of women with abnormal smears. This study investigates the consequences of using HPV testing to triage women either to direct treatment (HR-HPV positive) or to conventional practice (HR-HPV negative). This policy would reduce the number of colposcopically directed biopsies, but at the expense of some over-treatment and possibly extra costs. We investigated how large these numbers would be and what would be the associated costs.

8.2 Materials and methods

Study population

In a prospective study we considered 221 consecutive women aged 30-60 years who had persistent (i.e. two consecutive smears) mild or moderate dyskaryosis or a single smear reported as severe dyskaryosis, and who consequently were referred for further assessment to the university hospital in Groningen, the Netherlands [41]. In the Netherlands organized cervical cancer screening takes place targeting women between 30-60 years.

HPV-testing and histology

HPV-testing and histological assessment by colposcopically directed biopsy was performed in all women [41]. HPV was detected using the GP5/6 general primer mediated polymerase chain reaction (PCR), which was one of the most sensitive tests for the detection of HPV at that time. Besides, this test enabled the analysis of alternative groups of HPV types. Positive samples were

subsequently analyzed by means of type-specific primers for low-risk HPV-types 6, 11 and HR-HPV types 16, 18, 31, and 33 separately [173].

Management policies
In both patient groups we compared the numbers of medical procedures and their costs for the following alternative management policies:

1. conventional management: treatment is based on colposcopic and histological assessment. This is the current management policy in the Netherlands and also in many other countries.
2. women are triaged by HPV testing: HR-HPV-positive women are treated directly with loop excision of the transformation zone (LETZ) without prior histological assessment, while HR-HPV-negative women follow the conventional management policy.
3. direct LETZ treatment without prior virologic or histological assessment. This policy is not current practice, but is added for comparative purposes, and is the aggressive counterpart of the first two management policies.

Diagnostic and treatment procedures
We collected the number of diagnostic and treatment procedures in the study population which was conventionally treated [42], and translated these figures to the situations following each of the three management policies. These policies concern a setting in which LETZ is the local treatment of choice. In a situation without prior biopsy (policy 2 and 3) this is important because LETZ provides a histological diagnosis. We assumed that women with histological low-grade squamous intraepithelial lesions (LSIL) are in principal not treated. As a result, the number of treatments in these women will be similar to what has been observed in women with normal histology. We further assumed that women who are not treated in the conventional and HPV triage policy, are treated with LETZ in the direct treatment policy. As a result, the policies differ only in the frequency of LETZ and not of conisation and hysterectomy. Treatment of women with high-grade SIL (HSIL) is according to what has been observed in the study. In our calculations, women with LSIL or HSIL who are treated are followed up with annual Pap smears during five years. Women with normal histology or with low-grade SIL (LSIL) who are not treated, are followed up with two (six-monthly) smears and colposcopies over one year. Each smear, biopsy and treatment requires an outpatient visit. In all policies an extra visit is assumed at the start to discuss the management process with the patient. In the conventional policy (policy 1), the result and consequences of the histological examination are discussed during an additional visit. In the HPV policy (policy 2) one extra visit for HPV testing is assumed.

Unit costs of diagnostic and treatment procedures

Costs of HPV testing were estimated using the PCR (Polymerase Chain Reaction) technique and Southern blotting. Two alternative testing procedures were considered. More are available but their costs are intermediate. In the baseline estimation, one PCR using a general primer (GP) for all known HPV types is performed, while subsequent PCRs are performed on GP-positive samples with type-specific primers. In the alternative situation, prior specified HPV-types are detected through one PCR using a cocktail probe. This procedure is the cheapest one possible because only one PCR is needed, while the first one has the advantage that infection with other (although unknown) HPV types can be confirmed.

In the cost calculations unit costs are determined by the number of PCRs performed per test. Economies of scale by a more efficient use of equipment, housing and standard quality control measures, are accounted for: the average costs per PCR decrease when the production scale of the laboratory (number of PCRs per year) increases. The application of the PCR-technique requires specific investments and standard control procedures in order to avoid contamination and confounding of test results.

Relevant information for the cost assessment was gathered on the use of materials, required personnel, equipment, administration and overhead costs from laboratories for several production levels. The resource needs for each production level reflect a situation in which HPV testing is routinely applied. An additional number of 2000 PCRs processed for non-HPV tests was assumed, because equipment and housing can be used for the processing of other PCR-based tests. As a consequence, the costs per HPV test depend on the type of PCR testing procedure used and on the laboratory scale. For example, of women with two LSIL smears 56% are HPV positive for all types. Then, with the baseline technique $1 + 3 \times 0.56 = 2.68$ PCRs are used per test, compared with one PCR per test when using a cocktail probe (PCRs for control samples not included). With the baseline technique, unit costs per HPV test are less than 2.68 times the cocktail probe costs (which requires only 1 PCR), especially at high levels of scale.

Cost of colposcopy was assessed by interviewing colposcopists for time investment, by reviewing financial accounts of gynecology departments, and by cost analysis of the equipment [269]. The unit costs of hospital days and outpatient visits include hotel costs, nursing and medical staff, standard medical equipment, medication and overhead costs [210]. For other procedures we assumed that current fees were representative for their costs.

LETZ is predominantly an outpatient procedure, but we assumed that 10% requires a short hospital stay of a few hours. Because a societal perspective

is taken [96], time costs are included for women for outpatient and general practitioner visits, based on the average hourly labor wage for women. Costs are expressed in 1999 US$ using an exchange rate of 2.07 Dutch guilders for one dollar.

Literature search

In order to test the external validity of the findings on HPV prevalence in our study, and to test our results of the HPV triage to alternative patient groups and HPV tests, we performed an extensive literature search. We searched Medline (key words: "cervical intraepithelial neoplasia", "vaginal smears", "squamous intraepithelial lesion", "ASCUS", "LSIL", "HSIL", "abnormal smear", in combination with "human papillomavirus") for studies on HR-HPV prevalence in referral populations, with known histology results. We only included studies in which results were presented for well-defined patient groups (referral cytology containing not more than two subsequent cytomorphologic categories), and HPV testing by second generation Hybrid Capture (HC-II) or PCR (including at least HR-HPV 16, 31 and 33), because only these tests have a relatively high sensitivity for the detection of HPV [54]. Studies with high risk populations (e.g. HIV) were excluded. Of all studies we determined population characteristics and major test characteristics of referral cytology and HPV testing.

8.3 Results

HPV positivity

47% of women with persistent cytologic mild or moderate dyskaryosis and 69% of women with severe dyskaryosis were positive for HR-HPV 16, 18, 31 or 33 (table 8.1). The presence of only HR-HPV 16, 31 and/or 33 increased the probability of high-grade SIL (HSIL), because their positive predictive value (PPV) was higher than the PPV of cytology [41]. We used HSIL as the endpoint parameter, because these women are treated, and women without SIL or with LSIL are in principal not treated. Because triage by HPV testing between direct treatment and conventional management is only more efficient when the PPV can be increased, only these HR-HPV types will be considered in the calculations.

Table 8.1 HPV status by histological diagnosis in women of 30-60 years with two smears of mild or moderate dyskaryosis or one smear of severe dyskaryosis [41].

	Histology*				HPV prevalence	Sp	PPV
	Normal †	LSIL	HSIL	Total			
two smears of mild or moderate dyskaryosis							
HPV 16, 31 or 33	6	4	36	46	59.0% (46.7-71.4%)	77.8% (65.6-89.9%)	78.3% (66.3-90.2%)
HPV 18‡	1	2	1	4	1.6% (0.0-4.8%)	93.3% (86.0-100.0%)	25.0% (0.0-67.4%)
HPV 6/11‡, no HPV	15	17	24	56	n.a.	n.a.	n.a.
All smears	22	23	61	106	n.a.	n.a.	57.5% (48.1-67.0%)
one smear of severe dyskaryosis							
HPV 16, 31 or 33	0	0	70	70	66.7% (57.6-75.7%)	100.0% (n.a.)	100.0% (n.a.)
HPV 18‡	0	2	7	9	6.7% (1.9-11.4%)	80.0% (55.2-100.0%)	77.8% (50.6-100.0%)
HPV 6/11‡, no HPV	6	2	28	36	n.a.	n.a.	n.a.
	6	4	105	115	n.a.	n.a.	91.3% (86.2-96.5%)

SIL=squamous intraepithelial lesion, Sp=specificity of HPV test for HSIL, PPV=positive predictive value for HSIL
* Histology based on colposcopically directed biopsy. CIN 1 = low-grade SIL (LSIL), CIN 2/3 = high-grade SIL (HSIL)
† Including borderline changes.
‡ Single HPV 18 or 6/11 without 16, 31 or 33.

Table 8.2 Average number of diagnostic and treatment procedures per referred woman in case of conventional management and direct treatment policies, based on observed figures in Academic Hospital of Groningen [42].

	Histology Normal*	LSIL	HSIL
Conventional management policy			
Biopsy†	1.36	1.00	1.16
LETZ	0.11	0.11‡	0.60
Conisation	0.25	0.25‡	0.41
Hysterectomy	0.00	0.00	0.02
Direct treatment policy			
Biopsy†	0.00	0.00	0.00
LETZ	0.75	0.75¶	0.60
Conisation	0.25	0.25¶	0.41
Hysterectomy	0.00	0.00	0.02

* Including borderline changes.
† Colposcopically directed biopsy.
‡ We assumed that in the conventional management policy women with LSIL are in principal not treated. These women are treated similar to what has been observed in women with normal histology.
¶ In our study population, women with LSIL received LETZ in 74% of the cases and conisation in 26%.

Women with persistent mild or moderate dyskaryosis
We observed that 36% of women with normal histology were actually treated based on the colposcopic impression. Therefore, we used this same share for women with LSIL (table 8.2). Table 8.4 presents the number of treatment procedures resulting from each management policy by combining the figures presented in tables 8.1 and 8.2 and the unit costs of medical procedures in table 8.3. Per woman referred because of two consecutive smears reported as mild or moderate dyskaryosis and compared with the conventional management policy (policy 1), HPV triage (policy 2) avoids on average 0.51 colposcopically directed biopsies, but adds 0.05 LETZ treatments and 0.09 outpatient visits, with an additional cost of $134 (table 8.4). In other words, per additional LETZ treatment and 1.7 (CI 2.4, -0.6) outpatient visits, 9.6 (CI 5.7, 23.9) colposcopically directed biopsies are avoided. If the alternative cheap HPV test were used, additional costs would be only $37.

Compared with the conventional management policy (policy 1), direct treatment without prior histological assessment (policy 3) avoids 7.3 outpatient visits of which 4.6 include colposcopically directed biopsy per additional LETZ treatment. In addition, this policy would save $114 per woman.

Table 8.3 Unit costs ($ 1999) of medical procedures, visits and hospital days.

Item	Costs
Pap smear	21
HPV testing *	202
Primary colposcopy	80
Secundary colposcopy	59
Biopsy	44
LETZ†	284
Conisation ‡	1,460
Hysterectomy ‡	3,747
Outpatient visit excluding costs of procedures	47
Hospital day	266
Time costs for the woman per outpatient visit	7
Time costs for the woman per GP visit	4

* Including the collection of sample material.
† Including hospital days. It was observed in the Groningen hospital that 10% of LETZ (unit cost $265) need day care (unit cost $181).
‡ Including hospital days and pre-operative diagnostics. Average number of hospital days for conisation (unit cost $422) and hysterectomy (unit cost $1,007) are 3.6 and 10.0 respectively [249].

Table 8.4 Results for women referred after two consecutive mild or moderate dyskaryotic smears: predicted number of medical procedures per woman by management policy (differences with respect to conventional policy).

Medical procedures	Conventional policy	HPV triage		Direct treatment	
HPV testing	0.00	1.00	(+1.00)	0.00	(0.00)
Outpatient visits	8.83	8.92	(+0.09)	7.02	(-1.81)
		(9.03, 8.82) ‡	(+0.20,-0.01)		
Colposcopy	2.14	1.42	(-0.73)	0.00	(-2.14)
		(1.51, 1.33) ‡	(-0.64,-0.82)		
Biopsy*	1.15	0.63	(-0.51)	0.00	(-1.15)
		(0.67, 0.60) ‡	(-0.47,-0.55)		
LETZ	0.41	0.46	(+0.05)	0.66	(+0.25)
		(0.49, 0.43) ‡	(+0.08,+0.02)		
All treatment†	0.77	0.82	(+0.05)	1.02	(+0.25)
		(0.85, 0.79) ‡	(+0.08,+0.02)		
Years in follow-up	4.32	4.31	(-0.02)	4.51	(+0.19)
		(4.36, 4.26) ‡	(+0.03,-0.07)		
Costs ($)			(+134)		(-114)
			(+148,+121)		

* Colposcopically-directed biopsy.
† LETZ, conisation or hysterectomy
CI   ‡ Confidence intervals (95% CI) are based on the statistical uncertainty of the HPV prevalence (table 8.1), and range from unfavourable to favourable for HPV triage

Women with severe dyskaryosis

In women referred with one smear of severe dyskaryosis, HPV triage (policy 2) does not induce additional treatment of the cervix compared with the conventional policy (policy 1), and saves on average 0.72 colposcopically directed biopsies and 0.33 outpatient visits per woman, but with additional

costs of $133 per woman. Direct treatment (policy 3) will save on average 2.12 outpatient visits of which 1.16 include colposcopically directed biopsy and $103 per woman, and adds on average 0.05 LETZ treatment per woman. In other words, 45 outpatient visits of which 25 include biopsy can be saved against one additional LETZ treatment. The explanation is that the PPV of a severely dyskaryotic smear for the presence of HSIL is already very high (91.3%), and an additional biopsy will save only few women from being treated.

Table 8.6 Comparison (number of procedures) of management policies in alternative referral populations. Populations are ranked by severity of referral cytology.

Referral cytology (HPV test)	HPV triage versus conventional management		Direct treatment versus conventional management	
	Avoided biopsies per additional LETZ	Avoided visits per additional LETZ	Avoided biopsies per additional LETZ	Avoided visits per additional LETZ
2x mild or 1x moderate/severe dysplasia (PCR) [58]	7.8	1.3	4.0	6.9
2x borderline/mild or 1x moderate/severe dysplasia (PCR) [250]	9.0	2.7	4.2	7.8
Present study: 2x mild/moderate dysplasia (PCR)[42]	*9.6*	*-1.7*	*4.6*	*7.3*
2x mild or 1x moderate dysplasia (PCR) [149]	3.7	-1.2	2.6	5.0
2x borderline/mild dysplasia (PCR) [106]	3.7	-1.2	2.6	4.8
1x mild dysplasia (HC II) [17 48 87 145]	2.1	-0.4	2.0	2.6
1x borderline (HC II) [17 48 87 145 163 248 253]	2.5	0.8	2.2	4.3

LETZ = loop excision of the transformation zone.

Sensitivity analysis

The study results are to a large extent determined by three key characteristics: 1) we used PCR-based HPV-testing with HR-HPV types 16, 31 and 33, 2) the population is relatively old compared with other studies, and 3) the referral criteria for colposcopy were relatively conservative: two smears reported as mild or moderate dyskaryosis or one smear reported as severe dyskaryosis. From the literature review appeared that the prevalence and positive predictive value (PPV) of HR-HPV for HSIL in the present study are comparable to studies with similar populations and PCR-testing (table 8.5) [58 250]. In situations with

Table 8.5 HPV status in women referred for colposcopy and with known histology, separate for HC-II testing and PCR-testing. Studies are ranked by severity of referral cytology.

Study	Referral cytology	Age range (mean)	N	PPV cytology for high grade† (%)	HR-HPV prevalence in high grade† (%)	specificity HR-HPV for high grade† (%)	PPV HR-HPV for high grade† (%)	NPV HR-HPV for normal / low grade† (%)	HR-HPV-types
HC-II testing									
Ferris, 1998 [87]	borderline	18+ (27)	143	6	89	40	9	99	16,18,31,33,35,39,45,51,52,56,58,59,68
Clavel, 1999 [48]	borderline	15-72 (37)	23	9	100	57	18	100	id.
Manos, 1999 [163]	borderline	14-92 (40)	973	7	89	64	15	99	id.
Bergeron, 2000 [17]	borderline	15-75 (35)	111	11	83	62	21	97	id.
Shlay, 2000 [248]	borderline	15-76 (34)	195	8	93	74	23	99	id.
Lin, 2000 [145]	borderline	50+ (62)	74	36	100	74	69	100	id.
Solomon, 2001 [253]	borderline	18+ (29)	1149	11	96	49	20	99	id.
Lytwyn, 2000 [152]	borderline or mild dyskaryosis	16-50 (30)	87	9	88	51	15	98	id.
Ferris, 1998 [87]	mild dyskaryosis	18+ (27)	99	12	92	13	13	92	id.
Clavel, 1999 [48]	mild dyskaryosis	15-72 (37)	56	4	100	24	5	100	id.
Bergeron, 2000 [17]	mild dyskaryosis	15-75 (35)	267	5	93	44	8	99	id.
Lin, 2000 [145]	mild dyskaryosis	50+ (62)	45	47	100	46	62	100	id.
PCR-testing									
Cuzick, 1995 [56]	borderline	20-45 (31)	58	14	63	86	42	93	16,18,31,33
Adam, 1998 [2]	borderline or mild dyskariosis (2x)	14-75 (28)	454	15	65	60	22	91	16,18,31,33,35
Herrington, 1995 [106]	borderline or mild dyskaryosis (2x)	--	165	23	82	75	49	93	16,18,31,33
Cuzick, 1995 [56]	mild or moderate dyskaryosis	20-45 (31)	52	38	35	69	41	63	16,18,31,33
Nobbenhuis, 1999* [204]	mild or moderate dyskaryosis	18-55 (32)	297	34	87	52	48	89	16,18,31,33,35,39,45,51,52,56,58,59,66,68

Table 8.5 HPV status in women referred for colposcopy and with known histology, separate for HC-II testing and PCR-testing. Studies are ranked by severity of referral cytology.

Study	Referral cytology	Age range (mean)	N	PPV cytology for high grade† (%)	HR-HPV prevalence in high grade† (%)	specificity HR-HPV for high grade† (%)	PPV HR-HPV for high grade† (%)	NPV HR-HPV for normal / low grade† (%)	HR-HPV-types
Londesborough, 1996 [149]	mild (2x) or moderate dyskaryosis	16-69 (31)	258	25	75	74	49	90	16,18,31,33,35,45,52,58
Bollen, 1997 [26]	mild or moderate dyskaryosis (1 or 2x)	16-65 (35)	190	29	68	70	49	84	16,18,31,33,35
Bollen, 1997 [26]	mild or moderate dyskaryosis (1 or 2x)	16-65 (35)	190	29	95	40	40	95	16,18,31,33,35,39,45,51,52,56,58
Burger, 1995 [42]	mild or moderate dyskaryosis (2x)	all ages (35)	157	58	66	67	73	59	16,18,31,33
Burger, 1995 [42]	mild or moderate dyskaryosis (2x)	30-60 (±40)	106	58	59	78	78	58	16,31,33
Sigurdsson, 1997 [250]	borderline or mild dyskaryosis (2x), moderate or severe dyskaryosis	18-71 (33)	358	54	79	72	78	75	16,18,31,33,35
Cuzick, 1994 [58]	mild (2x), moderate or severe dyskaryosis	-- (32)	133	55	79	75	79	75	16,18,31,33,35

Borderline = ASCUS, mild dyskaryosis = LSIL, moderate or severe dyskaryosis = HSIL, HC = hybrid capture; HR-HPV = high-risk HPV, PPV = positive predictive value, NPV = negative predictive value.
* Histology is highest grade found by biopsy, LETZ, or conisation during surveillance when women reached severe dysplasia assessed by colposcopy or at the end of the study (5 year follow-up) in a prospective study.
† We used high-grade SIL as endpoint parameter for calculating test characteristics, because these women are treated, and women without CIN or with CIN 1 are in principal not treated.

more relaxed referral criteria, such as one smear of mild or moderate dyskaryosis or two borderline smears, as presently in e.g. the Netherlands and the UK, or even one borderline smear, as in the US in some settings, HPV triage by PCR testing with selected high-risk types similarly improves the PPV for the presence of HSIL but on a lower level (41-49% compared to 78% in the present study) [26 56 106 149]. Two studies with relatively young populations showed a modest improvement of the PPV by HPV triage [2 56]. The HC-II HPV test, which contains a broad range of HR-HPV types, has mainly been applied in populations with borderline or mildly dyskaryotic smears. In these populations, the HC-II shows a high prevalence of HPV and negative predictive value, but the PPV for the presence of HSIL is only modestly increased.

In table 8.6 the calculated performance of the HPV triage (policy 2) and direct treatment management (policy 2) policies in alternative referral populations is presented compared to conventional management (policy 1). The treatment numbers of table 8.2 were combined with the observed HPV prevalences in reviewed studies (table 8.5). For this purpose, HPV findings in similar populations and similar HPV tests were aggregated. It appears that HPV triage avoids similar numbers of biopsies per added LETZ as in the present study in comparable patient groups, but appears less favourable in populations with less severe referral cytology or when the HC-II HPV test is used. A similar pattern is observed for the direct treatment policy.

8.4 Discussion

For women referred because of persistent mild or moderate dyskaryosis on cytology, the choice between the considered strategies will, among others, depend on the quality of life (QoL) effects that are attributed to outpatient visits, colposcopically directed biopsy and conservative treatment (LETZ) respectively. Such QoL measurements are however not available. Pain and discomfort associated with biopsy and LETZ should be considered, as well as uncertainty involved in awaiting histological results. Because HPV triage can save 9.6 biopsies per additional LETZ compared with conventional policy, this policy will be preferred when undergoing LETZ is considered less than 10 times the burden of undergoing colposcopically directed biopsy (neglecting the small increase in outpatient visits). In addition, if the overall QoL of the HPV triage policy would be more favourable compared to conventional management, this must be weighed against the extra financial demands of $114 per referred woman.

For women referred because of a smear reported as severe dyskaryosis, direct treatment comes out as an attractive policy, also compared with the HPV

mediated policy. It saves a lot of diagnostic procedures at the cost of only very few added treatments. This confirms findings reported elsewhere [111].

The present study concerns a situation where LETZ is the local treatment of choice for SIL. LETZ has the advantage that it can be applied without prior histological examination. When for instance cryotherapy is included as an alternative, which requires prior biopsy, direct treatment is impossible and HPV triage will be less favourable.

We considered an HPV triage by which women with only HR-HPV 16, 31 or 33 would be treated directly, because in our study HR-HPV 18 did not improve the PPV for the presence of HSIL. As table 8.5 shows, HPV tests that include a lot of HR-HPV types beyond 16, 31 and 33 (including the commercially available HC-II) are not suitable for triage between direct treatment or colposcopy because a high HPV prevalence goes at the expense of a low PPV for the presence of HSIL. We conclude that more research should be done directed at the identification of those HR-HPV types that are predictive for the presence of HSIL.

There are indications that HPV triage performs more favourable in relatively old populations, considered the high PPV found in e.g. Lin [145] and in the group of women beyond age 30 compared to those below this age in our study (data not shown). In younger populations, the PPV of HPV for the presence of HSIL is relatively low. More evidence is needed to assess age as a possible additional triage criterium especially in populations with mild cytology.

In the present study, 64% and 50% of women with normal histology or LSIL are not treated in the conventional and HPV triage policy respectively, but have been followed up during one year with two six-monthly papsmears including colposcopy. Is this without risk? HSIL has been found in LETZ or cone biopsy in 38% [47] and 41% [251], respectively, of women who initially showed negative or LSIL at colposcopically directed biopsy. Some of these women - those who are not treated and develop cancer – would be missed in the conventional and HPV triage policy. In addition, from the Dutch national pathology database we found that of women with mildly or moderately dyskaryotic smears and negative histology shortly after, 3.5% had developed carcinoma in situ or borderline invasive cancer within five years [71]. It is not clear whether these cases have been picked up by follow-up smears or by a primary (screening) smear. Lifetime progression from LSIL to HSIL or invasive cancer has been estimated at 12% [212], while statistical analyses of screening data show slightly higher progression rates from LSIL to HSIL [31 287]. Considering that these cases are mostly picked up during follow-up or next

screening rounds, and that carcinoma in situ and borderline invasive cancer are still highly treatable, these figures seem tolerable.

HPV triage as evaluated in the present study is complementary to the current debate on the utility of HPV triage in women with less severe cytology (one ASCUS or LSIL smear) than in the present study. In these women, HR-HPV positives are referred for colposcopy and HR-HPV negative women are kept under cytologic surveillance, which is a triage protocol different from the present study [51 52 87 126 248 253 304]. For this triage definition, a high negative predictive value combined with a high HPV prevalence is supportive for HPV triage. Such test performance has been shown in several studies, but in some studies histological HSIL patients show negative on HPV test results (table 8.5, first part). Therefore the increased efficiency in patient management (avoided colposcopy) should be weighed against the few patients that are missed by HPV testing, and against the burden and risks associated with cytologic follow-up: prolonged uncertainty, non-compliance and disease progression. Of women with LSIL referral smears who have been followed-up cytologically for two years, 23-33% are found to be lost to follow-up and of the remainder 55% had progressed to HSIL on histology [3 89].

The ongoing technological development of HPV tests results in better test characteristics, but possibly in lower costs as well. Besides, lower cost levels could already be attained if HPV testing is concentrated in a few laboratories, taking advantage of economies of scale. We showed that at a unit price of $68 HPV testing results in similar costs per referred woman as the conventional management policy in women with smears reported as LSIL. This break-even point might vary among countries depending on the relative costs of HPV testing, colposcopically directed biopsies, LETZ and outpatient visits. For the US, $110-200 has been reported as unit costs for colposcopy and $84-100 for biopsy [126], which will result in even larger cost savings in the HPV triage and direct treatment policies than presented here.

We conclude that the choice between conventional management, direct treatment and HPV testing as a triage instrument in women with persistent mild or moderate dyskaryosis will depend on the relative burden for the patient that is attributed to outpatient visits and colposcopically directed biopsies compared to LETZ. A quantitative assessment of this burden, expressed in the reduction in quality of life for each of these procedures, will permit a more informed decision. Direct treatment with ablative techniques such as LETZ in women referred because of severely dyskaryotic smears seems to deserve serious consideration, because this will considerably reduce the burden of diagnostic procedures (colposcopy and biopsy), and hardly increase overtreatment.

General discussion 9

9.1 Introduction

We will summarize the main findings, discuss their methodological robustness, compare the findings with other research, draw conclusions and make recommendations. We will also integrate the described research by discussing the contribution of cost of illness estimates, burden of disease estimates and economic evaluations to the effectiveness and efficiency of health care.

9.2 Cost of illness in the Netherlands

Main findings

Health care costs (synonymous for medical costs) are dominated by old age and disability (chapter 2). Per capita health care costs are strongly age-dependent: they are relatively high in the first year of life, low during childhood and adulthood, and increase exponentially beyond age 50. Mental disorders and musculoskeletal diseases, two disabling but predominantly non-fatal disease clusters, outrank major killers as cancer, coronary heart disease and stroke. Almost 60% of total health care costsis accounted for by females, reflecting their larger life expectancy and the costs of reproduction.

How valid are the results?

The study provides broad insights into the cost distributions rather than precise cost estimates. Generally, we had to use indicators of health care consumption by diagnostic and demographic variables, that have a strong but imperfect relationship with real resource use. For instance, we were unable to distinguish between high and low intensity hospital days, thereby underestimating costs of diagnoses that account for a relatively large share of intensive care, such as coronary heart disease and injury.

For some types of health care, consumption data by diagnostic and demographic variables were not available or lacked detail or quality. Examples are old people's homes and home care, that together account for 15% of health care costs. As a result costs of chronic disabling diseases with high needs for professional care, e.g. musculoskeletal diseases, will have been underestimated, and it is difficult or even impossible to indicate any confidence bounds.

We defined disease clusters to avoid that results would be distorted by misclassification of diagnoses in health care registers. An example is mental disorders that are often too complex for mapping in a unidimensional classification. However, for other health problems more detailed results would be useful. In the field of injuries, both the physical injury and the causal mechanisms are relevant aspects of the diagnosis. It is also for these reasons that multidimensional classifications have been developed for mental disorders (DSM-III) [7], and injuries (ICECI) [298], in addition to the international classification of diseases and injuries (ICD) that we used [296].

Comorbidity
An important matter in COI studies is how to attribute costs in case of comorbidity, the fact that many patients (often elderly) suffer from more than one condition. These conditions may be unrelated (e.g. arthrosis and Parkinson's disease), whereas some diseases are risk factors for other diseases (e.g. diabetes and coronary heart disease). In our analysis, costs were attributed to the primary diagnosis, which is the principal diagnosis that gave cause for health care consumption. These costs include possible extra resource use because of other health problems. For instance, a woman with a hip fracture might have a longer hospital stay and more intensive care if she also suffers from a neurologic disorder. All costs have then been attributed to the hip fracture. Whether costs of specific diseases are thus over- or underestimated, as stated elsewhere [219], depends on the aim of the analysis, such as estimating savings through prevention. In our example, had the hip fracture not occurred, than also the extra resources because of the neurologic disorder would not have been used, and it can be justified to allocate all costs to the hip fracture. Alternatively, had the neurologic disorder been absent, only the extra costs related to this condition would have been saved. Another example is decubitus, that can be prevented by direct measures or by interventions that tackle the conditions because of which patients are bedridden. Depending on the aim of the analysis, decubitus related costs should be attributed to decubitus or these underlying conditions.

The problem of comorbidity can be solved by attributing costs to more than one diagnosis or risk factor. This would give insight into how costs of diseases and risk factors are related. This approach has been adopted in a recent generic COI study for the Netherlands, but it increases the data needs considerably [219].

Comparison with other studies

Differences between our COI results and those of other studies can have many causes. Studies may differ in their level of comprehensiveness, the definition and classification of diseases, in applying a top-down or bottom-up approach (including differences in the analysis of comorbidity), and in the quality of data sources [219].

The comprehensiveness of our study with regard to the inclusion of health care compares favourably with other studies, that exclude parts of psychiatric care, nursing home care and other elderly care [146 150]. This explains our high cost estimates in old age and mental disorders, including dementia. Other studies may include indirect costs of lost production due to disease and disability, thereby inflating the total cost estimate considerably, depending on the method used to value productivity costs [14]. We disregarded productivity costs because of the lack of reliable methods to measure them, the lack of good quality work absence data in the Netherlands, and because a description confined to health care costs contributes to the interpretability of results and the utility for health care policy.

In our COI study costs reflect health care consumption in a given year (i.c. 1994) that can be attributed to population groups defined by age, sex and diagnostic group. The resulting costs per population group do not show how these costs are distributed *within* each population group. Some persons will consume much health care and others less or even no health care at all. For diseased persons health care need is much related to the disease stage. For instance, in cancer patients costs may be U-shaped, with high costs at the time of diagnosis and treatment, relatively low costs during follow-up, and increasing costs upto the end of life in case the cancer is incurable [301]. In other words, our study does not give information on the lifetime distribution of health care consumption at the individual level. The steap increase of per capita health care costs by age may be due to higher levels of disability in old age, but may also reflect high costs in the last year of life [91 160 220 254]. Whether costs are related to (the proximity of) death or to disability in the many years preceding death is an important matter for societies with ageing populations. If costs are related to disability rather than death, a decrease in mortality will lead to more years lived with disability and therefore increasing health care consumption. If costs are related to dying rather than ageing, health care consumption will be postponed over the human lifespan when mortality declines. Accounting for health care costs in the last year of life thus has a moderating effect on the projected increase in health care costs. Because about 10% of total costs are in the last year of life, the impact is very small [220]. Another study found that not accounting for costs in the last year of life leads to

a 20% overestimate of the increase in health care costs, but the analysis was limited to hospital and primary care only [160]. It can be concluded that population ageing will lead to an increase in health care costs, because the majority of costs are related to chronic, degenerative diseases that start many years before the end of life (chapter 2).

Recommendations and future research
The comparibility of COI data may be increased by the development of guidelines for conducting and reporting COI studies. In this regard, a facilitative effort is the EUCOMP project of national statistical offices in EU member states, that aims to describe the content of health care providers in national statistics of health care costs and production. This helps to define packages of health services that are internationally comparable [285]. When these data could be integrated with COI data, this would greatly enhance the investigation of international differences in health care expenditures, and the underlying supply and demand factors.

Another challenge will be to attribute costs to specific health risks [233]. This is essential for targeting prevention. Because most diseases have multiple risk factors, and single risk factors often regard more than one disease, a comprehensive approach that accounts for multicausality would enable to quantify the combined (economic) effect of single or multiple risk factors beyond specific diseases [81].

9.3 Medical costs of injury
Main findings
We linked a national injury surveillance system (LIS) with a bottom-up costing model by which health care consumption and costs were estimated per individual patient (chapter 3). It thus became possible to estimate total health care costs of injury on a continuous basis, and for any subgroup of injury patients. Health care costs of injury were 1.1 billion euro or €1,019 per patient in 1998 (more recent years are also available). Peaks in total costs were observed in males between age 15 and 44, primarily due to high numbers of injury in this age group, and in females beyond age 65, primarily due to high costs per patient such as in hip fractures. Minor injuries without need for hospitalization, predominantly superficial injury and open wounds, together accounted for more than a third of health care costs of injury. Independent determinants of individual health care consumption were age, sex, hospitalization, injury diagnosis, motor vehicle crash, and number of injuries.

Medical costs of injuries could be used as an indicator for the relative importance of specific injuries. This is particularly useful for injuries, that

include high frequency minor injuries and low frequency severe injuries. Although in many cases health care consumption is strongly related to a patient's health status, in some cases it does not reflect health care needs. Medical costs should therefore be interpreted with caution when used as an indicator of population health.

Because of the linkage with the national injury surveillance system, the costing model can be used for continuous monitoring of injuries on aggregate and more detailed level. Because the model is incidence-based, and lifetime costs can be estimated per patient, it can provide necessary input for the economic evaluation of preventive interventions and trauma care.

How valid are the results?

Although the cost of injury study was comprehensive, considering a broad range of injuries and medical care, costs have been underestimated. Injury patients who are treated by primary care providers (e.g. general practitioners) were not considered. Their number is estimated at 1.3 million per year, and even exceed the number of patients treated in Emergency Departments (ED) [62]. Because injuries treated by primary care providers are predominantly minor, not resource intensive injuries, they would add about 10% to our total cost estimate (chapter 3). In addition, we did not include long-term costs of injury after the first year post-injury because of lack of valid data.

The statistical uncertainty of our cost estimates is primarily determined by the uncertainty of the incidence estimates. The incidence of unadmitted patients has been derived from the national injury surveillance system (LIS). LIS is based in 17 hospitals, which is the number of observations for the calculation of statistical uncertainty. The uncertainty of national incidence estimates will be higher when the variance of the incidence is large among hospitals, such as in ice-skating injuries. However, our cost estimates can still be regarded robust for the following reasons. First, the uncertainty of the incidence estimates is limited to nonadmitted patients that account for only a third of total medical costs (we used the national hospital discharge register (LMR) with national coverage for the number of admitted injury patients). Second, we reported results for broadly defined patient groups that are unlikely to have large variation in incidence among hospitals.

To calculate total ED costs, we estimated average costs per ED visit. These average ED costs should not be confused with marginal ED costs, defined as the costs of adding one extra ED visit. The marginal to average cost ratio of EDs has been found to be far below one because of the high fixed costs of running an ED, that must be staffed for 24 hours to treat real emergencies [263]. Williams estimated the marginal to average cost ratio at 0.41 for nursing

resources and infrastructure and 0.35 for physician resources [303]. A time and motion study in an academic hospital showed that 45% of nursing time was directly related to patient treatment [276]. Our average costs per ED visit can therefore not be interpreted as cost savings for instance by shifting nonurgent visits to primary care.

Comparison with other studies
Our estimates of the costs of injuries are more comprehensive and detailed compared to those of generic, top-down COI studies [221 275]. Costs of unintentional injury were slightly higher in our study (1.1 billion euro) than in Polder *et al.* (1.0 billion euro) [221]. Costs of traffic injury, upper extremity injuries and superficial injury (including contusions) were higher in our study, predominantly because we separately distinguished ED costs. Our costs of hip fracture were lower, because we excluded old people's homes and we limited the length of stay in nursing homes to the average observed in patients without comorbidity.

In the international literature, not many comparable cost studies exist (chapter 4). In Rice *et al.* [231], per capita medical costs of injury were three times our estimate, because of higher costs per patient (costs per patient were similar, but Rice *et al.* also included minor injuries treated by primary care providers), and because of a 1.5 times higher incidence. Other estimates for the United States were similar to those of Rice *et al.* [180] or slightly lower because of the exclusion of long-term costs [178]. For Australia, per capita medical costs were estimated at $145 in a bottom-up analysis [293] (about twice our estimate), compared to $105 in a generic, top-down COI study (2000 US$PPP) [164]. Interestingly, the difference in per patient costs between the US and the Netherlands is analogous to the two times higher total health care costs per capita in the US (PPP adjusted). Because per capita health care costs are comparable between the Netherlands and Australia (PPP adjusted), this does not explain the much higher costs of injury in Australia compared with the Netherlands.

Many cost of injury studies include productivity costs and 'human costs' of pain and suffering. Productivity losses are estimated at about three times the medical costs, whereas human costs are even higher [22 59]. For the Netherlands, van Beeck *et al.* estimated productivity losses due to injury in 1988 at $3,293 million with the human capital method (HCM) and at $702 million with the friction cost method (FCM), compared to $952 million for medical costs [275]. Both types of costs need a further discussion.

The inclusion of productivity losses in COI studies is theoretically justified because disease and injury negatively influences the availability of

scarce resources (labour) that have opportunity costs. An matter of concern is the validity of the cost estimates. The HCM estimates productivity losses from the occurrence of the disease until full recovery or, in case of permanent disability, the age of retirement. In contrast, the FCM accounts for the possibility that sick workers are replaced sooner or later by formerly unemployed persons [131]. The duration of the period until replacement, the friction period, primarily depends on the unemployment rate and the type of job. The HCM premises that societal welfare loss is the summation of (lifetime) individual productivity losses due to temporary or permanent disability and premature death. This implicit individual perspective is in contrast with the societal perspective of the FCM. The validity of both methods should be further tested. The HCM is appropriate in situations of full employment and scarcity of labour, and in studies that adopt an individual perspective. The FCM will by definition provide an underestimate when not accounting for future friction periods that may occur in situations of full employment. These future friction periods are however difficult to operationalize. Another concern of the FCM is that the main parameters, the duration of the friction period and the production elasticity of labour, are likely to be job specific, but we are unaware of any empirical data. In addition, more empirical data are needed on compensation of lost work hours by colleagues and others [118], and on reduced productivity during work hours ('sickness presenteeism') [39].

Estimates of 'human costs' are based on monetary valuations of lost quality adjusted life years (QALY), as empirically derived by willingness-to-pay methods (WTP). However, the valuations heavily depend on the method used to elicit preferences. In a systematic review of studies, the monetary value per QALY ranged from $25,000 (HCM) to more than $400,000 (revealed preferences for job risks), and with estimates from stated preference methods somewhere inbetween [108]. In contrast, WTP estimates appear scope insensitive: valuations have a weak relationship with the size of the benefits that are to be valued. This makes WTP particularly inappropriate for burden of disease estimates. In a systematic review of WTP studies, these findings made Olsen *et al.* to conclude that the WTP method is "sensitive to theoretically irrelevant information, and insensitive to theoretically relevant information" [209].

Recommendations and future research
The use of cost of injury studies in prioritizing health policy, and the necessary conditions, need to be further explored [188]. Another recommendation is that data on costs of injury should be complemented with cost-effectiveness information of injury control measures. Preventive interventions and trauma care, either existing or newly developed, need to be evaluated to further

develop, implement or discontinue these activities. In the Netherlands few examples exist of economic evaluations in injury control [61 211 222].

Health care costs of injury are predominantly determined by the incidence and severity of injuries. Because information on injury severity is not uniformly registered in most hospitals [274], the use of health care costs as an alternative indicator for the burden of injury needs to be tested in two ways. First, the combined analysis of trends in incidence and health care costs of specific injuries will give a first indication of trends in injury severity, and may generate hypotheses that can be tested in further analyses. Second, previous findings on the relationship between individual patient costs and injury severity as classified by validated instruments (Abbreviated Injury Scale, AIS, and Injury Severity Scale, ISS) need to be further explored and tested [155 159 180]. In addition to severity, this research should consider other determinants of individual health care consumption, such as comorbidity and socio-economic status.

Finally, as with generic COI studies, the (international) comparability of cost of injury studies could be enhanced by the development of a taxonomy and of guidelines for conducting and reporting cost of injury studies.

9.4 Injury related disability

Main findings

We found that the average health status of non-hospitalized patients 2 months post-injury, measured by the generic EuroQol instrument (EQ-5D+), was comparable to the general population (chapter 5). However, patients with injuries to the vertebral column and the extremities or with skull-brain injury reported lower than normal levels of functioning. An average of 5 work days were lost per non-hospitalized injury, and 5% had not yet returned to work after 2 months.

Hospitalized patients reported higher prevalences of disability than non-hospitalized patients in all health domains of the EQ-5D+. The mean EQ-5D summary measure increased from 0.62 after 2 months to 0.74 after 5 months and remained below the population norm at 9 months, particularly in patients below age 60. Hospitalized patients with injury to the spinal cord or vertebral column or a lower extremity fracture reported the worst health status after 2 months, also when adjusted for age, sex and educational level. Those with a paid job on average lost 72 work days, and 40%, 20% and 10% had not yet returned to work after 2, 5 and 9 months, respectively. Age, sex, educational level, injury diagnosis, and several indicators of injury severity were independent and significant predictors of functional outcome.

How valid are the results?

The main limitation of our study was the low response rate. We adjusted the results for systematic non-response with available data on background characteristics, but particularly the 9 month results cannot be considered representative for specific subgroups, such as young adults and skull-brain injuries.

We identified three main causes of the low response. First, the population was very heterogeneous, included persons of different socio-demographic and ethnic groups, and encompassed patients with various injuries: minor burns as well as hip fractures and severe brain injury. Young adults and elderly, and persons with lower education in general respond less to postal questionnaires. We could have limited our research to persons below age 64 and native Dutch speakers, but the aim of our research was to collect representative and comprehensive data on functioning after injury in the Dutch population. Second, we used postal questionnaires without reminders and other stimuli that could have increased response rates. Reminders, monetary incentives, personal follow-up, and personal interviews in stead of postal questionnaires in general lead to higher response rates [74]. However, these measures need considerable amounts of resources. Third, a number of persons refused to participate because of insufficient mental or physical fitness. For these persons the use of proxies to fill in the questionnaire would have been a good alternative.

Another limitation of our study was that we did not have adequate information of post-clinical mortality. By definition the responders all had survived their injury. The mean case fatality rate in injury victims that reach the hospital is about 2% [274].

We did not collect information on comorbidity, which is particularly relevant in elderly patients. Comorbidity is predominantly prevalent among elderly, and is an important independent determinant of mortality and disability [266]. We used norm scores of health in the general population to adjust for disability because of pre-existing conditions.

Comparison with other studies

Studies conducted so far on post-injury functioning are very heterogeneous. Differences are in general related to the patient sample, measurement instruments, follow-up intervals, and other design issues. Our study is one of the few that did not exclude patient groups a priori. This explains why hospitalized patients in our sample were on average less severely injured than in other studies. For instance, a larger proportion of hospitalized patients had returned to work after 9 months (90%) than in MacKenzie *et al.* after one year

(57%) [159]. Our observed EQ-5D summary scores after 5 and 9 months (0.74) compare favourably with the average Quality of Well-Being (QWB)-score of 0.63 and 0.67 after six and twelve months, respectively, in Holbrook *et al.*, also when accounted for population norms for these measures [109 110]. In Holbrook *et al.* 62% were traffic injuries compared to only 29% in Dutch hospitals.

 There are few studies on disability in non-hospitalized patients. In one study with non-hospitalized ED patients, 68% were at least partially restricted in work activities for one day or more (probably related to the treatment itself), and 10% reported restricted activities of daily living. After 1 month, 10% were at least partly disabled for work and 1% were restricted in their daily activities [292]. These rapid recovery rates explain why we found on average normal levels of functioning after 2 months. One should also consider that in the Netherlands more than 50% of injury patients at an ED are non-urgent, and have minor injuries such as contusions, abrasions, open wounds and small burns. The vast majority of these injuries leads to temporary disability, although a small proportion will result in long-term disability [258]. It has even been asserted that a large proportion of prevalent disability can be attributed to non-hospitalized injury patients [217]. We found that, among non-hospitalized patients, particularly injuries to the vertebral column had less than normal levels of functioning after 2 months.

Recommendations and future research
The heterogeneity of research conducted so far on functioning and disability in injury patients is likely related to the heterogeneity of injuries itself. All the more this stresses the need for uniform methodologies to generate comparative information among groups, over time, and among countries. This uniformity can be increased by defining a number of standards for study design, and by reaching agreement on these in the injury research community [72]. In short, these standards consider what, when and how should be measured:
a. a classification of injuries by which diagnoses with similar functional sequelae and speed of recovery are clustered, e.g. with use of the recently released International Classification of Functioning [297].
b. a minimum set of measurement intervals that match the several stages of the recovery process: the acute treatment phase, rehabilitation phase, adaptation phase and the stable end situation [72]. If necessary, for specific injuries additional measurements could be added to this minimum set, to match the particular speed of recovery.
c. the use of a generic instrument for measuring health that covers physical, mental and social functioning as the general constituents of health. If

necessary, this generic instrument should be supplemented with specific instruments by which injury-specific types of functional loss and restrictions in activities and social participation can be measured. Any used instruments should be reliable and valid, should closely match the functional consequences of injuries, and should be easily administrable and not too time consuming. The validity of generic instruments in injury populations should be further tested [274].

d. utility based instruments enable the calculation of a summary measure based on the scores on each domain, and facilitate a rapid comparison across different injury groups (see also paragraph 9.7).

e. in addition to health status, a standard set of personal and injury related variables should be collected that are associated with disability, including proxy measures in case specific information is not available (e.g. injury severity).

The data on injury disability presented in this thesis are an example of the systematic collection of data for calculating the burden of injury, as stimulated by the Global Burden of Disease project [194]. Future efforts should integrate these data with data on functional outcome from other studies, e.g. major trauma patients [283 291] and tibial fractures [112]. An important application will be to estimate the burden of injury in the Netherlands by combining these data with incidence and mortality data, distributed by relevant accident categories. We made preliminary estimates showing that the total number of years lived with disability due to injury in 1999 is about 122,000, and about 10,000 when adjusted for comorbidity. Of these, 38% is due to home and leisure injuries and 27% to traffic injuries [171]. Also, our estimates of disability need to be validated and enhanced by new follow-up studies using designs that guarantee a sufficient reponse, with shorter time intervals for non-hospitalized patients as applied by us, and extending the follow-up period beyond one year to capture long-term consequences of injury.

9.5 Cervical cancer screening
Main findings
In chapters 6 and 7 (part II) we investigated the test characteristics of newly developed cytologic technologies for cervical cancer screening, and their (cost-)effectiveness compared to screening with the Pap test. In a critical review of trials we found that there is weak evidence that one liquid based 'thinlayer' cytology (LBC)system (ThinPrep™) and two automated systems (AutoCyte™ SCREEN, AutoPap™) are more sensitive than the Pap test, at the loss of some specificity. We designed a decision analytic framework based on a cost-

effectiveness analysis (CEA) to indicate the minimum test performance (sensitivity, specificity, smear adequacy) of any new cytologic screening test for which it would have an acceptable C/E ratio, given te costs per test. Considering the costs of the current technologies, and that in most countries screening is more intensive than in the baseline calculations, it is unlikely that they are more cost-effective than the Pap test.

How valid are the results?
We estimated the effectiveness of screening with the MISCAN microsimulation model. The parameters in the MISCAN model that describe the natural history of cervical cancer and the Pap test sensitivity have been quantified with screening data from British Columbia [287]. These quantifications were found to be consistent with international data on interval cancers [116] and resembled the incidence estimates of Gustafsson that were based on Swedish data [101]. Also, it was possible to reproduce the epidemiology of cervical cancer in the Netherlands with the MISCAN model for the period before screening as well as for the years after the introduction of screening [267].

We used life years gained as primary outcome measure in the economic evaluation without accounting for quality of life (QoL), because of the lack of estimates of changes in QoL due to cervical cancer screening. An increased test sensitivity will reduce the number of invasive cancers and endstage disease, but will result in more primary treatment of pre-clinical stages. Although the QoL implications of these opposite effects could not be quantified, it is unlikely that the inclusion of QoL effects would considerably change our conclusions.

Comparison with other studies
Trials of new cytologic tests necessarily use intermediate outcome measures such as numbers of abnormal cytology or histology, because measuring differences in mortality would require large numbers of participants and a very long follow-up. Our modelling approach has the advantage that these intermediate outcome measures can be translated into health effects (invasive cancers prevented, life years gained). In addition, a priori evaluation of the C/E of new screening tests is possible to inform the design of possible trials of new screening tests.

Our conclusions on the accuracy of new cytologic tests are in line with other reviews [37 197]. We were able to include some more recent trials, but this did not change the outcome.

In Brown *et al.,* incremental cost-effectiveness rapidly deteriorated with more intensive screening, as in our analysis [40]. AutoPap came out relatively favourable, which was partly due to an overestimate of the test sensitivity (i.c.

95%). Moreover, the duration of pre-clinical stages was described with an exponential distribution compared to the Weibull distribution in the MISCAN model. As a result, the proportion of fast growing cancers is overestimated by Brown *et al.,* and therefore also the favourable effects of screening with a more sensitive test.

Myers estimated incremental cost per life year gained below $3,000 with a hypothetical new test for a 5 year screening interval [196]. However, he assumed a low Pap test sensitivity (51%) and a high sensitivity of the new test (99%). We reproduced the analysis with our MISCAN model as far as possible, and came up with higher incremental costs per life year gained, whereas the estimated gain in life years seemed to be fairly comparable. The differential outcome is due to the lower costs of screening and higher treatment costs of invasive cancer (and therefore larger cost savings) in Myers' model.

Human papillomavirus testing (HPV) is increasingly being considered as an adjunct or even a substitute for cytologic screening [55]. HPV testing has been found to be more sensitive but less specific than the Pap test. So far, small longitudinal studies indicate that the screening interval can be lengthened for women who are HPV and cytology negative, and this may also apply to HPV negative / cytology positive women. For the large number of HPV positive women, efficient follow-up strategies should be designed to minimize the burden to these women. The HART-study recently showed that one repeat testing after 12 months may be sufficient for women with HPV and negative or borderline cytology to decide whether or not women should be referred for colposcopy [55]. In addition, larger longitudinal trials should give more definite information for any decision to change the screening policy and follow-up regimen. Because a conversion to LBC or automated cytologic screening, to HPV testing, or both has severe organizational implications, countries that did not yet convert to automated or LBC screening systems may probably better await these trial results.

Recommendations and future research
Because an acceptable cost-effectiveness of new cytologic tests for cervical cancer screening is difficult to achieve, alternative strategies should be investigated to make screening more efficient. A major source of cervical cancer mortality is non-participation in screening. Although participation is never mandatory, efficient interventions that increase participation are urgently needed.

A major source of uncertainty in the economic evaluation of liquid based and automated screening technologies is their true test characteristics. Many trials have been performed with weak designs. The evaluation of

diagnostic tests is complex, which urges the development of practical guidelines for conducting trials in this field. Trials should also capture possible learning effects.

The use of LBC for population based screening of cervical cancer is already widespread in several countries, particularly those with more intensive screening policies than in the Netherlands. These practices are likely to be inefficient. Because the application of LBC is also considered in the Netherlands, and already practiced in some laboratories, it should be investigated whether the current policies for implementing and financing new diagnostic tests guarantee an efficient use of health care resources.

9.6   Follow-up of abnormal cervical screening smears
Main findings
We analyzed whether testing for high-risk human papillomavirus (HR- HPV) would result in a more efficient follow-up of screened women who are currently referred for colposcopy because of persistent mild or moderate dyskaryosis or a single severe dyskaryosis smear (chapter 8). Women with a positive HPV test would be treated directly with loop excision of the transformation zone (LETZ), without prior histological assessment, and women with a negative HPV test would get conventional management (colposcopically directed biopsy).

Compared to conventional management, HR-HPV triage of women with persistent mild or moderate dyskaryosis will avoid histological assessment at the expense of some overtreatment: per woman on average 0.51 colposcopically directed biopsies are avoided, but at the expense of 0.05 LETZ treatments and 0.09 outpatient visits per woman. Also costs are $134 higher per woman. These numbers imply that about 10 colposcopically directed biopsies are avoided per additional LETZ.

In women with severe dyskaryosis, direct treatment was more efficient than HPV triage. Compared to conventional management, 45 outpatient visits, of which 25 with colposcopically directed biopsies could be avoided per additional LETZ. Considered these numbers, direct treatment should seriously be considered in these women.

How valid are the results?
We considered women between age 30 and 60 because these women are screen-eligible in the Netherlands. The positive predictive value (PPV) of HPV testing seems to be lower in younger females, and so will be the benefits of HPV triage.

We did not have information about patient preferences for the considered management policies. Therefore, no definitive conclusions could be drawn on the relative effectiveness of the alternative management policies.

Comparison with other studies
The prevalence of HPV is very much determined by the type of HPV test and the age and referral criteria of the patient population [54]. HPV prevalence and PPV for the presence of CIN 2-3 (HSIL) in our study was comparable to studies with similar referral populations and HPV tests [58 250].

The management protocol considered by us is relatively unique. Other studies consider the use of HPV testing as a triage instrument for colposcopy in women with ASCUS (≈Pap 2) or LSIL (mild dyskaryosis) smears [87 125 248 253]. Women with a negative HPV test do not get further follow-up. Recently, the ALTS-trial concluded that HPV testing in women with a single LSIL smear has limited potential, because 83% of women were HPV positive (HC-II test) [259]. Therefore, only a minority of these women would benefit from HPV triage. The ALTS-trial also revealed that in women with an ASCUS smear, 96% of HSIL would be detected with HPV triage, whereas the negative predictive value (for <HSIL) of the HPV test was 99%. This implies that almost half of the women who are referred because of one ASCUS smear would avoid further diagnosis, at the expense of a very small number of missed cases. HPV testing was superior to a single repeat smear, having a lower PPV and lower sensitivity than the HPV test. A recent meta-analysis showed that HPV testing in women with one ASCUS smear showed higher sensitivity and similar specificity compared to repeat cytology [8].

Whether HPV triage of women with ASCUS is cost-effective in the Netherlands remains unclear until these results are translated to the Dutch screen-eligible age group and laboratory setting: in the US the proportion of ASCUS smears is 5-15%, in the Netherlands the proportion of Pap 2 smears has been reduced to 2%. HPV triage has been estimated at $33,600 / life year saved compared with no follow-up [128], but the assumed specificity was more favourable (75-85%) than observed in the ALTS-trial (49%).

Future research
We were not able to measure quality of life implications of further diagnosis and treatment. Future research should therefore consider the preferences of females for different follow-up policies. The recently started project on quality of life implications of cervical cancer screening and follow-up is important in this respect.

In the Netherlands the follow-up policy has recently changed towards earlier referral after abnormal repeat smears, with the aim to restrict the duration of follow-up. Future research should weigh the pros and cons of all possible follow-up strategies, including HPV testing, with use of data of past and ongoing trials (e.g. the BOB-study).

9.7 Burden of disease, cost of illness studies and economic evaluation

What are the relative merits of burden of disease (BOD) and cost of illness (COI) studies and cost-effectiveness analyses (CEA) in allocating health care resources?

The recent debate on BOD studies particularly concerned the Global Burden of Disease project of WHO [185 191 302]. It seems that experts in health economics and public health have chosen one of two camps: those who are criticasters of BOD/COI studies and those who are not. The most recent result of this debate is the publication of two voluminous books on the value of summary measures of population health and on health systems performance assessment [193 195]. We support BOD/COI studies, and will argue that BOD/COI studies and CEAs are complementary.

BOD/COI studies provide a cross-sectional description of population health and health care costs, respectively, subdivided by diseases. In a subsequent step the health burden and costs can be further attributed to (multiple) risk factors (see also paragraph 9.2). In other words, BOD and COI studies provide essential data regarding the equity of health needs and access to health care, respectively. CEAs, on their part, provide information on the efficiency of health care. If such information would be available for all health interventions, optimal efficiency in health care can be achieved by giving priority to interventions that maximize health at the least costs. Both BOD/COI data and data on the efficiency of health care are a prerequisite for societies that aim to optimize the level and distribution of health in accordance with societal values.

Critics put forward that BOD/COI estimates would set priorities based on the size of the health problem, measured in epidemiological indicators or health care consumption. This would at worst simply increase health care costs of those diseases that already account for a large share. However, none of the proponents of BOD/COI studies has ever advocated this. An additional criticism is that they do not provide information on the efficiency of health care, and therefore would not be an aid for decision making on alternative interventions [45 53 302]. We oppose, however, that BOD/COI studies never pretended to indicate the efficiency of health care, and that indeed CEAs are essential for this aim.

BOD/COI studies can be used to identify those diseases or risk factors with a high current or future burden. These health problems might not be identified without the data provided by BOD/COI studies. Subsequently it can be investigated whether the identified health problems are eligible for interventions, and whether there is any evidence on the effectiveness and efficiency of these interventions. A first application of BOD/COI studies is therefore the prioritization of research and health policy with respect to the design and implementation of (cost-)effective interventions. Comprehensive BOD/COI studies are more suitable for this purpose than disease-specific studies. Disease-specific studies can be (and often are) used for single disease advocacy, which might give unjustified priority to diseases for which data are available. Also, the sum of single disease estimates may easily exceed the total disease burden and health expenditures. In contrast, the coherent framework of comprehensive COI and BOD studies will put specific diseases or injuries into perspective and may highlight health problems that receive insufficient attention. In other words, without BOD/COI data the search for cost-effective interventions will be a blind search. In this regard BOD studies, and we add COI studies here, have been mentioned as an essential part of the 'public health accounts' similar to what the national accounts are for macro-economic policy, and nobody would question the utility of the latter [191].

Second, descriptive statistics on population health and health care costs in a coherent BOD/COI framework are needed for the comparison of health and health care between countries, over time, and between population groups. Many of the observed differences and changes may require further explanation and exploration [279]. For instance, higher age-specific per capita costs in females compared to males indicate differences in specific health care needs that could be due to underlying socio-demographic and biomedical factors, and that may warrant attention of health care planners. Also, differences in mortality or health care costs among geographic regions or among socio-economic and ethnic groups may point at specific access barriers to health care. Because these applications require valid and consistent data that may not be available on a sufficient detailed level, BOD/COI studies also play a vital role in identifying important data gaps.

Third, COI studies provide insight in the drain on health care resources subdivided by diseases, and in the relative importance of preventive, curative or caring activities. As shown in chapter 2, large fractions of health care resources are spent on caring activities for people with degenerative disease, such as dementia and mental retardation. This illustrates that COI studies aid in identifying diseases and health care activities where priority setting based on cost-effectiveness is not applicable: homes for disabled persons, nursing homes,

etcetera. At least, the concept of 'effectiveness' has a different content in prevention and cure (e.g. healthy life years) than in care (e.g. independent living, dignity), and therefore standard CEA methods do not apply to care activities.

Fourth, BOD/COI data can be used as an input into and reference framework for CEAs [191]. Summary measures of population health have not only been designed for descriptive purposes, but can also be used as outcome measures in CEAs. National estimates of costs of smoking related diseases and coronary heart disease have been used to explore the economic consequences of smoking and coronary heart disease interventions, respectively [15 27]. Reasoning that the costs and (health) benefits of an intervention should be equal to the difference in health care spending and population health with and without this intervention, regularly conducted BOD/COI studies can be used to monitor the actual impact of interventions at national level [191]. Because the medical practice is never a controlled experiment, it will not always be possible to attribute actual changes in population health and health care spending to specific interventions or even a combination of interventions. Nevertheless, specific changes in the burden and cost of illness could at least be indicative for the succes or failure of interventions in routine practice.

So far we did not distinguish between COI and BOD studies with respect to their contribution to priority setting in health care. Although patterns of resource consumption often resemble the distribution of the burden of disease (need), and high medical costs are indicative for a high burden in terms of need (chapter 2), COI data should be interpreted with caution. Some diseases may be untreatable and may for this reason be cheap or expensive, whereas considerable resources may be spent in order to make a disease a negligible health burden.

Some final remarks consider the use of CEAs to prioritize health care. First, it is a general misconception that knowledge about the incremental costs and health benefits of health care technologies is sufficient for making decisions on their implementation and financing, apart from legal, ethical and political considerations. Because the budget is limited, one needs a C/E threshold to judge whether the intervention is acceptable. The problem of finding a C/E criterium to judge whether an intervention is acceptable is often faced in practice by policy makers and health care planners, and might partly explain why the results of CEAs are so difficult to implement in practice [67]. As a result, in actual decisions other, more arbitrary, criteria are used, such as the C/E of other (similar) interventions or rules of thumb (see also chapter 7).

In case of an explicit budget constraint and an additional number of assumptions, the CEA paradigm prescribes that an intervention is acceptable if

its C/E ratio does not exceed the C/E ratio of the marginal programme, i.e. the final programme with the highest C/E that was accepted [294]. In this approach information is needed on the C/E of all possible health interventions before an implementation decision can be made [121]. However, in practice an explicit budget constraint is usually absent, interventions may need extra resources beyond the health care budget, or may need fixed assets that are already in use by other interventions. Some propose cost-benefit and WTP analysis as a solution [67 121]. An alternative approach would be to collect information on the C/E of activities that would be displaced by the intervention being evaluated (opportunity costs) [18 242 294]. The programme to be displaced should have a lower C/E ratio and free up enough resources to fund the new programme. This approach provides a second best solution because it improves but not optimizes resource allocation.

Second, cost-effectiveness rankings of interventions often have wide confidence intervals, neglect considerations of fairness and equity, and may hide underlying methodological discrepancies [70]. As a result, the ranking of interventions has some degree of arbitrariness, which has compromized previous experiences in which CEAs were used or advocated for setting priorities in health care, among which the well-known Oregon-experiment [23 24 94]. This underlines the importance of methodological development in HTA research in addition to developing guidelines for conducting HTA. This can be illustrated by innovating efforts to present uncertainty [36 84 115 284], measure productivity losses [39 131], and elicit preferences for health care [240].

Third, CEAs do not give information on the total budget impact of interventions, or on impact on total population health once the intervention would be implemented on national level. This is a typical example where BOD/COI estimates, describing the pre-intervention situation, and cost-effectiveness data can be sensibly integrated.

Fourth, decisions based on CEAs of new interventions may not guarantee an optimal resource allocation when the reference scenario (the situation by which the new intervention is compared) is not efficient. Often the reference scenario is 'usual care', usually a heterogeneous mixture of interventions. If usual care is itself not efficient, the new intervention may compare very favourable. Both usual care and the new intervention should then be compared with the null scenario, i.e. a situation without any intervention [191].

9.8 Conclusions and recommendations

1. *How are medical costs of injury at national level distributed by type of injury and health care sector, and what are their major determinants?*

About a third of health care costs of injury is due to high frequency minor injuries that do not need hospitalization, whereas hip fracture, with high costs per patient, accounts for a fifth of injury-related health care costs. Young males and elderly females contribute significantly to total costs of injury. Hospital care accounts for more than two-third of total medical costs of injury, of which a quarter is made in the Emergency Department. In addition to age and sex, several indicators of injury severity were identified as determinants of individual health care consumption: injury diagnosis, hospitalization, motor vehicle crash, and number of injuries.

We recommend the following:
- Our surveillance based model for calculating the medical costs of injury should be applied internationally and may be applicable in other realms of population health (e.g. cancer, psychiatric disorders).
- Medical costs of injury should be used by policy makers as an indicator to prioritize the development of interventions that prevent injury and improve trauma care.
- The efficiency of alternative preventive interventions should be evaluated.
- The efficiency of the current treatment of minor injuries (i.e. injuries without need for hospitalization) should be investigated.
- Costs of injury estimates should be used to develop and test indicators of injury severity that differentiate among injuries in terms of health care need.

2. *How is injury related disability at national level distributed by type of injury, and what are its major determinants?*

The majority of non-hospitalized patients has a normal level of functioning 2 months post-injury, but persons with vertebral column injury, extremity injury, and skull-brain injury were still below general population norms. Those with a paid job lost on average one week and this was longest for patients with upper extremity fractures. The health status of hospitalized injury patients increased up to 5 months post-injury, but remained below population norms even after 9 months. On average they were 14 weeks absent from work; 40%, 20% and 10% had not yet returned to work after 2, 5 and 9 months respectively. Hospitalized

patients with injury to the spinal cord or vertebral column, and those with a lower extremity fracture reported the worst health status. Both non-hospitalized and hospitalized injury patients contribute substantially to injury-related disability and particularly work absence. We identified proxy indicators of injury severity (hospital length of stay, ICU admission, motor vehicle crash, medical operation, number of injuries) that appeared to be independent and significant predictors of functioning in addition to age, sex and educational level.

We recommend the following:
- To identify the risk factors that contribute significantly to the burden of injury.
- To measure short-term disability in minor, non-hospitalized injuries.
- To measure permanent functional consequences and their predictors in major injuries, including non-hospitalized vertebral column injury, extremity injury, and skull-brain injury.
- Studies investigating quality of life in injury patients should apply uniform designs (patient selection criteria, measurement instruments, intervals).
- The relationship between poor post-injury functioning and socio-demographic and injury-related characteristics should be further investigated, and the implications for injury prevention and trauma care determined.

3. *What are the test characteristics of newly developed cytologic technologies for cervical cancer screening, and how (cost-)effective are these technologies compared to screening with Pap smears?*

There is weak evidence that some liquid based and automated screening technologies (ThinPrep™, AutoCyte™ SCREEN, AutoPap™) are more sensitive than the Pap test, at the loss of some specificity. The exact test characteristics are however not known due to flaws in trial designs. Despite this uncertainty, it is unlikely that screening with these tests is within acceptable cost-effectiveness levels in the Netherlands. In situations with more intensive screening schedules than in the Netherlands, as in many western countries, it is even less likely that these tests are cost-effective. In contrast, these tests might be cost-effective in situations with low Pap test sensitivity and low Pap test adequacy.

We recommend the following:

- The test characteristics of liquid based cytology (LBC) and automated screening devices should be determined. Any trials should have a sufficient duration to capture any learning effects.
- Practical guidelines should be developed to facilitate the complex evaluation of diagnostic tests.
- It should be investigated whether the current policies for implementing and financing new diagnostic tests guarantee an efficient use of health care resources.

4. *Can the follow-up of women with abnormal Pap smears be made more efficient by human papillomavirus testing?*

In women with persistent mild or moderate dyskaryotic smears, HPV triage (positives are treated, negatives will receive biopsy) will avoid unnecessary colposcopy and biopsies at the expense of some overtreatment. About 10 colposcopically directed biopsies can be avoided per additional local treatment of the cervix (LETZ), at the expense of some extra costs. This HPV triage protocol is less efficient in women with borderline cytology or a single mildly dyskaryotic smear because of its poor positive predictive value in these women. Also in women with severe dyskaryosis HPV triage is not efficient, but direct treatment (i.e. without histological assessment) can avoid 45 outpatient visits and 25 colposcopically directed biopsies per additional LETZ.

We recommend the following:
- Measurement of the quality of life implications of cervical cancer screening and follow-up.
- HPV triage should be considered in women with persistent mild or moderate dyskaryosis smears, taking women's preferences into account.
- Direct treatment (i.e. without histological assessment) with LETZ should be seriously considered in women with severely dyskaryotic smears.

5. *To what extent do burden of disease studies, cost of illness studies and economic evaluation studies provide helpful information for the prioritization of health care?*

Burden of disease (BOD) and cost of illness (COI) studies are complementary to economic evaluation studies (CEAs) in prioritizing health care. They a) help identify health areas where research and the design and implementation of (cost-)effective interventions is most needed, b) provide useful input for the economic evaluation of these interventions, c) generate comparative information on population health and health care costs that deserve further

exploration, d) help identify important epidemiological and health care data gaps, e) give insight in the relative importance of prevention, curative and caring activities for specific diseases, and f) act as a reference framework to trace the actual impact of interventions on population health and health care expenditures.

We recommend the following:
- Existing comprehensive BOD and COI studies in the Netherlands should be regularly updated, and should be integrated with disease-specific studies.
- Cost of illness estimates should be attributed to risk factors using a comprehensive approach, to quantify the combined (economic) effect of single or multiple risk factors beyond specific diseases.
- Research should be conducted into acceptable cost-effectiveness thresholds that are in line with society's preferences for health care.
- Cost-effectiveness analyses should evaluate the existing situation ('usual care') if insufficient knowledge exists on the individual or combined health effects of current interventions.

References

1. Aaron H, Schwartz WB. Rationing health care: the choice before us. Science 1990;247:418-22.

2. Adam E, Kaufman RH, Berkova Z, Icenogle J, Reeves WC. Is human papillomavirus testing an effective triage method for detection of high-grade (grade 2 or 3) cervical intraepithelial neoplasia? Am J Obstet Gynecol 1998;178:1235-44.

3. Alanen KW, Elit LM, Molinaro PA, McLachlin CM. Assessment of cytologic follow-up as the recommended management for patients with atypical squamous cells of undetermined significance or low grade squamous intraepithelial lesions. Cancer 1998;84:5-10.

4. Andersen RM, Newman JF. Societal and individual determinants of health care utilization. Milbank Mem Fund Q Health Soc 1973;51:95-124.

5. Anke AG, Stanghelle JK, Finset A, Roaldsen KS, Pillgram-Larsen J, et al. Long-term prevalence of impairments and disabilities after multiple trauma. J Trauma 1997;42:54-61.

6. Anonymous. The screening muddle. Lancet 1998;351:459.

7. APA. Diagnostic and Statistical Manual of Mental Disorders, third edition: American Psychiatric Association, 1980.

8. Arbyn M, Buntinx F, Van Ranst M, Paraskevaidis E, Martin-Hirsch P, et al. Virologic versus cytologic triage of women with equivocal Pap smears: a meta-analysis of the accuracy to detect high-grade intraepithelial neoplasia. J Natl Cancer Inst 2004;96:280-93.

9. Australian Institute of Health & Welfare. Australia's health 1996. http://www.aihw.gov.au/publications/h_online/ah96/index.html.

10. Autier P, Haentjens P, Bentin J, Baillon JM, Grivegnee AR, et al. Costs induced by hip fractures: a prospective controlled study in Belgium. Belgian Hip Fracture Study Group. Osteoporos Int 2000;11:373-80.

11. Badia X, Diez-Perez A, Alvarez-Sanz C, Diaz-Lopez B, Diaz-Curiel M, et al. Measuring quality of life in women with vertebral fractures due to osteoporosis: a comparison of the OQLQ and QUALEFFO. Qual Life Res 2001;10:307-17.

12. Baker SP, O'Neill B, Haddon W, Jr., Long WB. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. J Trauma 1974;14:187-96.

13. Balen v. Hip fracture in the elderly: impact, recovery and early geriatric nursing home rehabilitation (thesis). Rotterdam: Erasmus University, 2003.

14. Barendregt JJ, Bonneux L. The trouble with health economics. Eur J Public Health 1999;9:309-12.

15. Barendregt JJ, Bonneux L, van der Maas PJ. The health care costs of smoking. N Engl J Med 1997;337:1052-7.

16. Bergeron C, Bishop J, Lemarie A, Cas F, Ayivi J, et al. Accuracy of thin-layer cytology in patients undergoing cervical cone biopsy. Acta Cytol 2001;45:519-24.

17. Bergeron C, Jeannel D, Poveda J, Cassonnet P, Orth G. Human papillomavirus testing in women with mild cytologic atypia. Obstet Gynecol 2000;95:821-7.

18. Birch S, Gafni A. Cost effectiveness/utility analyses. Do current decision rules lead us to where we want to be? J Health Econ 1992;11:279-96.

19. Bishop JW. Comparison of the CytoRich system with conventional cervical cytology. Preliminary data on 2,032 cases from a clinical trial site. Acta Cytol 1997;41:15-23.

20. Bishop JW. The cost of production in cervical cytology. Comparison of conventional and automated primary screening systems. Anatomic Pathology 1997;107:445-50.

21. Bishop JW, Bigner SH, Colgan TJ, Husain M, Howell LP, et al. Multicenter masked evaluation of AutoCyte PREP thin layers with matched conventional smears. Including initial biopsy results. Acta Cytol 1998;42:189-97.

22. Blincoe LJ, Seay AG, Zaloshnja E, Miller TR, Romano EO, et al. The economic impact of motor vehicle crashes 2000. US Department of Transportation: NHTSA, 2002.

23. Bodenheimer T. The Oregon Health Plan--lessons for the nation. First of two parts. N Engl J Med 1997;337:651-5.

24. Bodenheimer T. The Oregon Health Plan--lessons for the nation. Second of two parts. N Engl J Med 1997;337:720-3.

25. Bolick DR, Hellman DJ. Laboratory implementation and efficacy assessment of the ThinPrep cervical cancer screening system. Acta Cytol 1998;42:209-13.

26. Bollen LJ, Tjong AHSP, van der Velden J, Brouwer K, Mol BW, et al. Human papillomavirus deoxyribonucleic acid detection in mildly or moderately dysplastic smears: a possible method for selecting patients for colposcopy. Am J Obstet Gynecol 1997;177:548-53.

27. Bonneux L, Barendregt JJ, Nusselder WJ, der Maas PJ. Preventing fatal diseases increases healthcare costs: cause elimination life table approach. Bmj 1998;316:26-9.

28. Boon ME, Kok LP. Neural network processing can provide means to catch errors that slip through human screening of pap smears. Diagn Cytopathol 1993;9:411-6.

29. Boon ME, Kok LP, Nygaard-Nielsen M, Holm K, Holund B. Neural network processing of cervical smears can lead to a decrease in diagnostic variability and an increase in screening efficacy: a study of 63 false-negative smears. Mod Pathol 1994;7:957-61.

30. Bos A, van Ballegooijen M, Habbema JDF. Screening history of invasive cervical cancers in the Netherlands 1994-1997. unpublished 2004.

31. Bos AB, van Ballegooijen M, van Oortmarssen GJ, van Marle ME, Habbema JD, et al. Non-progression of cervical intraepithelial neoplasia estimated from population-screening data. Br J Cancer 1997;75:124-30.

32. Bosch MM, Rietveld-Scheffers PE, Boon ME. Characteristics of false-negative smears tested in the normal screening situation. Acta Cytol 1992;36:711-6.

33. Brazier J, Jones N, Kind P. Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. Qual Life Res 1993;2:169-80.

34. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002;21:271-92.

35. Brenneman FD, Redelmeier DA, Boulanger BR, McLellan BA, Culhane JP. Long-term outcomes in blunt trauma: who goes back to work? J Trauma 1997;42:778-81.

36. Briggs AH, Goeree R, Blackhouse G, O'Brien BJ. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. Med Decis Making 2002;22:290-308.

37. Broadstock M. Effectiveness and cost-effectiveness of automated and semi-automated cervical screening devices. A systematic review of the literature. Christchurch: Christchurch School of Medicine, 2000.

38. Brouwer WB, Koopmanschap MA, Rutten FF. Patient and informal caregiver time in cost-effectiveness analysis. A response to the recommendations of the Washington Panel. Int J Technol Assess Health Care 1998;14:505-13.

39. Brouwer WB, Koopmanschap MA, Rutten FF. Productivity losses without absence: measurement validation and empirical evidence. Health Policy 1999;48:13-27.

40. Brown AD, Garber AM. Cost-effectiveness of 3 methods to enhance the sensitivity of Papanicolaou testing. Jama 1999;281:347-53.

41. Burger MP, Hollema H, Pieters WJ, Quint WG. Predictive value of human papillomavirus type for histological diagnosis of women with cervical cytological abnormalities. Bmj 1995;310:94-5.

42. Burger MPM, van Ballegooijen M. Evaluation of two-step screening for cervical intraepithelial neoplasia (in Dutch). Groningen, 1995.

43. Burstrom K, Johannesson M, Diderichsen F. Health-related quality of life by disease and socio-economic group in the general population in Sweden. Health Policy 2001;55:51-69.

44. Bush TL, Miller SR, Golden AL, Hale WE. Self-report and medical record report agreement of selected medical conditions in the elderly. Am J Public Health 1989;79:1554-6.

162

45. Byford S, Torgerson DJ, Raftery J. Economic note: cost of illness studies. Bmj 2000;320:1335.
46. Carpenter AB, Davey DD. ThinPrep Pap Test: performance and biopsy follow-up in a university hospital. Cancer 1999;87:105-12.
47. Chappatte OA, Byrne DL, Raju KS, Nayagam M, Kenney A. Histological differences between colposcopic-directed biopsy and loop excision of the transformation zone (LETZ): a cause for concern. Gynecol Oncol 1991;43:46-50.
48. Clavel C, Masure M, Bory JP, Putaud I, Mangeonjean C, et al. Hybrid Capture II-based human papillomavirus detection, a sensitive test to detect in routine high-grade cervical lesions: a preliminary study on 1518 women. Br J Cancer 1999;80:1306-11.
49. Colgan TJ, Patten SF, Jr., Lee JS. A clinical trial of the AutoPap 300 QC system for quality control of cervicovaginal cytology in the clinical laboratory. Acta Cytol 1995;39:1191-8.
50. Coste J, Cochand-Priollet B, de Cremoux P, Le Gales C, Cartier I, et al. Cross sectional study of conventional cervical smear, monolayer cytology, and human papillomavirus DNA testing for cervical cancer screening. BMJ 2003;326:733.
51. Cox JT. HPV testing: is it useful in triage of minor Pap abnormalities? J Fam Pract 1998;46:121-4.
52. Cox JT, Lorincz AT, Schiffman MH, Sherman ME, Cullen A, et al. Human papillomavirus testing by hybrid capture appears to be useful in triaging women with a cytologic diagnosis of atypical squamous cells of undetermined significance. Am J Obstet Gynecol 1995;172:946-54.
53. Currie G, Kerfoot KD, Donaldson C, Macarthur C. Are cost of injury studies useful? Inj Prev 2000;6:175-6.
54. Cuzick J, Sasieni P, Davies P, Adams J, Normand C, et al. A systematic review of the role of human papillomavirus testing within a cervical screening programme. Health Technol Assess 1999;3:1-204.
55. Cuzick J, Szarewski A, Cubie H, Hulman G, Kitchener H, et al. Management of women who test positive for high-risk types of human papillomavirus: the HART study. Lancet 2003;362:1871-6.
56. Cuzick J, Szarewski A, Terry G, Ho L, Hanby A, et al. Human papillomavirus testing in primary cervical screening. Lancet 1995;345:1533-6.
57. Cuzick J, Terry G, Ho L, Hollingworth T, Anderson M. Human papillomavirus type 16 in cervical smears as predictor of high-grade cervical intraepithelial neoplasia. Lancet 1992;339:959-60.
58. Cuzick J, Terry G, Ho L, Hollingworth T, Anderson M. Type-specific human papillomavirus DNA in abnormal smears as a predictor of high-grade cervical intraepithelial neoplasia. Br J Cancer 1994;69:167-71.
59. Danseco ER, Miller TR, Spicer RS. Incidence and costs of 1987-1994 childhood injuries: demographic breakdowns. Pediatrics 2000;105:E27.
60. de Bruin A, de Koning H, van Ballegooijen M. Pap smears and mammographies, Dutch National Health Interview Surveys 1991. Monthly Bulletin of Health Statistics 1993:12-15.
61. de Laet CE, van Hout BA, Hofman A, Pols HA. Kosten wegens osteoporotische fracturen in Nederland; mogelijkheden voor kostenbeheersing. Ned Tijdschr Geneeskd 1996;140:1684-8.
62. den Hertog PC, Geurts JJM, Hendriks HMH, Hutten JM, van Kampen LTB, et al. Ongevallen in Nederland 1997/1998. Amsterdam: Consumer Safety Institute, 2000.
63. Department of Health Housing and Community Services. Screening for the prevention of cervical cancer. Canberra: AGPS, 1991.
64. Diaz-Rosario LA, Kabawat SE. Performance of a fluid-based, thin-layer papanicolaou smear method in the clinical setting of an independent laboratory and an outpatient screening population in New England. Arch Pathol Lab Med 1999;123:817-21.
65. Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095-108.
66. Dolan P, Edlin R. Is it really possible to build a bridge between cost-benefit analysis and cost-effectiveness analysis? J Health Econ 2002;21:827-43.

67. Donaldson C, Currie G, Mitton C. Cost effectiveness analysis in health care: contraindications. Bmj 2002;325:891-4.
68. Doornewaard H, van der Schouw YT, van der Graaf Y, Bos AB, Habbema JD, et al. The diagnostic value of computer-assisted primary cervical smear screening: a longitudinal cohort study. Mod Pathol 1999;12:995-1000.
69. Doornewaard H, Woudt JM, Strubbe P, van de Seijp H, van den Tweel JG. Evaluation of PAPNET-assisted cervical rescreening. Cytopathology 1997;8:313-21.
70. Drummond M, Torrance G, Mason J. Cost-effectiveness league tables: more harm than good? Soc Sci Med 1993;37:33-40.
71. Dutch Institute for Healthcare Improvement (CBO). Application of automated screening, liquid-based cytology and HPV detection in the cervical screening programme. Utrecht: CBO, 2002.
72. ECOSA working group on post-injury levels of functioning and disability. Measuring disability of trauma patients: guidelines towards more valid estimates of the burden of injury. Rotterdam: Erasmus University, 2003.
73. Eddy DM. Screening for cervical cancer. Ann Intern Med 1990;113:214-26.
74. Edwards P, Roberts I, Clarke M, DiGuiseppi C, Pratap S, et al. Increasing response rates to postal questionnaires: systematic review. Bmj 2002;324:1183.
75. Elixhauser A, Halpern M, Schmier J, Luce BR. Health care CBA and CEA from 1991 to 1996: an updated bibliography. Med Care 1998;36:MS1-9, MS18-147.
76. Esselman PC, Ptacek JT, Kowalske K, Cromes GF, deLateur BJ, et al. Community integration after burn injuries. J Burn Care Rehabil 2001;22:221-7.
77. Essink-Bot ML. Health status as a measure of outcome of disease and treatment (thesis). Rotterdam: Erasmus University, 1995.
78. Essink-Bot ML, Bonsel GJ. How to derive disability weights. In: Murray CJM, Salomon JA, Mathers CD, Lopez AD, editors. Summary measures of population health: concepts, ethics, measurement and applications. Geneva: World Health Organization, 2002.
79. Essink-Bot ML, Krabbe PF, Bonsel GJ, Aaronson NK. An empirical comparison of four generic health status measures. The Nottingham Health Profile, the Medical Outcomes Study 36-item Short-Form Health Survey, the COOP/WONCA charts, and the EuroQol instrument. Med Care 1997;35:522-37.
80. EVAC (Evaluation Committee). Population screening for cervical cancer in the pilot regions Nijmegen, Rotterdam and Utrecht. A report by the Evaluation Committee. First and second interim report (in Dutch). Leidschendam: Ministry of Welfare, Public Health and Cultural Affairs, 1980.
81. Ezzati M, Hoorn SV, Rodgers A, Lopez AD, Mathers CD, et al. Estimates of global and regional potential health gains from reducing multiple major risk factors. Lancet 2003;362:271-80.
82. Fahey MT, Irwig L, Macaskill P. Meta-analysis of Pap test accuracy. Am J Epidemiol 1995;141:680-9.
83. Farnsworth A, Chambers FM, Goldschmidt CS. Evaluation of the PAPNET system in a general pathology service. Med J Aust 1996;165:429-31.
84. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. Health Econ 2001;10:779-87.
85. Ferenczy A, Robitaille J, Franco E, Arseneau J, Richart RM, et al. Conventional cervical cytologic smears vs. ThinPrep smears. A paired comparison study on cervical cytology. Acta Cytol 1996;40:1136-42.
86. Fern KT, Smith JT, Zee B, Lee A, Borschneck D, et al. Trauma patients with multiple extremity injuries: resource utilization and long-term outcome in relation to injury severity scores. J Trauma 1998;45:489-94.
87. Ferris DG, Wright TC, Jr., Litaker MS, Richart RM, Lorincz AT, et al. Triage of women with ASCUS and LSIL on Pap smear reports: management by repeat Pap smear, HPV DNA testing, or colposcopy? J Fam Pract 1998;46:125-34.

88. Field MJ, Gold GM, eds. Summarizing population health: directions for the development and application of population metrics. Washington DC: National Academy Press, 1998.

89. Flannelly G, Anderson D, Kitchener HC, Mann EM, Campbell M, et al. Management of women with mild and moderate cervical dyskaryosis. Bmj 1994;308:1399-403.

90. Forsen L, Sogaard AJ, Meyer HE, Edna T, Kopjar B. Survival after hip fracture: short- and long-term excess mortality according to age and gender. Osteoporos Int 1999;10:73-8.

91. Frankel S, Ebrahim S, Davey Smith G. The limits to demand for health care. BMJ 2000;321:40-5.

92. Frew EJ, Whynes DK, Wolstenholme JL. Eliciting willingness to pay: comparing closed-ended with open-ended and payment scale formats. Med Decis Making 2003;23:150-9.

93. Gerdtham UG, Jonsson B. International comparisons of health expenditure. In: Culyer AJ, Newhouse JP, editors. Handbook of health economics. Amsterdam: Elsevier Science, 2000:11-53.

94. Gezondheidsraad. Contouren van het basispakket. Den Haag: Gezondheidsraad, 2003.

95. Glancy KE, Glancy CJ, Lucke JF, Mahurin K, Rhodes M, et al. A study of recovery in trauma patients. J Trauma 1992;33:602-9.

96. Gold ME, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996.

97. Greenwood DC, Muir KR, Doherty M, Milner SA, Stevens M, et al. Conservatively managed tibial shaft fractures in Nottingham, UK: are pain, osteoarthritis, and disability long-term complications? J Epidemiol Community Health 1997;51:701-4.

98. Groenenboom GKC, Huijsman R. Elderly care in economic perspective: cost scenarios. Utrecht, De Tijdstroom: STG (Steering Committee on Future Health Scenarios, 1995.

99. Grohs DH. Impact of automated technology on the cervical cytologic smear. A comparison of cost. Acta Cytol 1998;42:165-70.

100. Guidos BJ, Selvaggi SM. Detection of endometrial adenocarcinoma with the ThinPrep Pap test. Diagn Cytopathol 2000;23:260-5.

101. Gustafsson L, Adami HO. Natural history of cervical neoplasia: consistent results obtained by an identification technique. Br J Cancer 1989;60:132-41.

102. Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed 1985;20:79-93.

103. Harlan LC, Harlan WR, Parsons PE. The economic impact of injuries: a major source of medical costs. Am J Public Health 1990;80:453-9.

104. Hartunian NS, Smart CN, Thompson MS. The incidence and economic costs of cancer, motor vehicle injuries, coronary heart disease, and stroke: a comparative analysis. Am J Public Health 1980;70:1249-60.

105. Hendrie D, Rosman DL, Harris AH. Hospital inpatient costs resulting from road crashes in Western Australia. Aust J Public Health 1994;18:380-8.

106. Herrington CS, Evans MF, Hallam NF, Charnock FM, Gray W, et al. Human papillomavirus status in the prediction of high-grade cervical intraepithelial neoplasia in patients with persistent low-grade cervical cytological abnormalities. Br J Cancer 1995;71:206-9.

107. Hessling JJ, Raso DS, Schiffer B, Callicott J, Jr., Husain M, et al. Effectiveness of thin-layer preparations vs. conventional Pap smears in a blinded, split-sample study. Extended cytologic evaluation. J Reprod Med 2001;46:880-6.

108. Hirth RA, Chernew ME, Miller E, Fendrick AM, Weissert WG. Willingness to pay for a quality-adjusted life year: in search of a standard. Med Decis Making 2000;20:332-42.

109. Holbrook TL, Anderson JP, Sieber WJ, Browner D, Hoyt DB. Outcome after major trauma: discharge and 6-month follow-up results from the Trauma Recovery Project. J Trauma 1998;45:315-23.

110. Holbrook TL, Anderson JP, Sieber WJ, Browner D, Hoyt DB. Outcome after major trauma: 12-month and 18-month follow-up results from the Trauma Recovery Project. J Trauma 1999;46:765-71.

111. Holschneider CH, Ghosh K, Montz FJ. See-and-treat in the management of high-grade squamous intraepithelial lesions of the cervix: a resource utilization analysis. Obstet Gynecol 1999;94:377-85.

112. Hoogendoorn JM, van der Werken C. Grade III open tibial fractures: functional outcome and quality of life in amputees versus patients with successful reconstruction. Injury 2001;32:329-34.

113. Hutchinson ML. Assessing the costs and benefits of alternative rescreening strategies. Acta Cytol 1996;40:4-8.

114. Hutchinson ML, Zahniser DJ, Sherman ME, Herrero R, Alfaro M, et al. Utility of liquid-based cytology for cervical carcinoma screening: results of a population-based study conducted in a region of Costa Rica with a high incidence of cervical carcinoma. Cancer 1999;87:48-55.

115. Hutubessy RC, Baltussen RM, Evans DB, Barendregt JJ, Murray CJ. Stochastic league tables: communicating cost-effectiveness results to decision-makers. Health Econ 2001;10:473-7.

116. IARC Working Group on evaluation of cervical cancer screening programmes. Screening for squamous cervical cancer: duration of low risk after negative results of cervical cytology and its implication for screening policies. Br Med J 1986;293:659-64.

117. Iftner T, Arbyn M, Ronco G, Patnick J, Anttila A. Personal communication about inadequate and mildly abnormal screen smears in Germany, Belgium, Italy, United Kingdom and Finland., 2003.

118. Jacob-Tacken KHM, Koopmanschap MA, Meerding WJ, Severens JL. Correcting for compensating mechanisms related to productivity costs in economic evaluations of health care programs. submitted for publication.

119. Jenkins D, Sherlaw-Johnson C, Gallivan S. Can papilloma virus testing be used to improve cervical cancer screening? Int J Cancer 1996;65:768-73.

120. Jenny J, Isenegger I, Boon ME, Husain OA. Consistency of a double PAPNET scan of cervical smears. Acta Cytol 1997;41:82-7.

121. Johannesson M, Meltzer D. Some reflections on cost-effectiveness analysis. Health Econ 1998;7:1-7.

122. Jorgensen HS, Nakayama H, Raaschou HO, Larsen K, Hubbe P, et al. The effect of a stroke unit: reductions in mortality, discharge rate to nursing home, length of hospital stay, and cost. A community-based study. Stroke 1995;26:1178-82.

123. Jurkovich G, Mock C, MacKenzie E, Burgess A, Cushing B, et al. The Sickness Impact Profile as a tool to evaluate functional outcome in trauma patients. J Trauma 1995;39:625-31.

124. Karnon J, Peters J, Platt J, Chilcott J, McGoogan E. Liquid-based cytology in cervical screening: an updated rapid and systematic review. Sheffield:: University of Sheffield, 2003.

125. Kaufman RH, Adam E. Is human papillomavirus testing of value in clinical practice? Am J Obstet Gynecol 1999;180:1049-53.

126. Kaufman RH, Adam E, Icenogle J, Reeves WC. Human papillomavirus testing as triage for atypical squamous cells of undetermined significance and low-grade squamous intraepithelial lesions: sensitivity, specificity, and cost-effectiveness. Am J Obstet Gynecol 1997;177:930-6.

127. Kaufman RH, Schreiber K, Carter T. Analysis of atypical squamous (glandular) cells of undetermined significance smears by neural network-directed review. Obstet Gynecol 1998;91:556-60.

128. Kim JJ, Wright TC, Goldie SJ. Cost-effectiveness of alternative triage strategies for atypical squamous cells of undetermined significance. Jama 2002;287:2382-90.

129. Kline TS. The challenge of quality improvement with the Papanicolaou smear. Arch Pathol Lab Med 1997;121:253-55.

130. Kok MR, Boon ME. Consequences of neural network technology for cervical screening: increase in diagnostic consistency and positive scores. Cancer 1996;78:112-7.

131. Koopmanschap MA, Rutten FFH, van Ineveld BM, van Roijen L. The friction cost method for measuring indirect costs of disease. J Health Econ 1995;14:171-89.

132. Koopmanschap MA, van Roijen L, Bonneux L. Costs of diseases in the Netherlands. [in Dutch]. Rotterdam: Erasmus University, Department of Public Health, Institute of Medical Technology Assessment, 1991.

133. Koopmanschap MA, van Roijen L, Bonneux L, Bonsel GJ, Rutten FFH, et al. Costs of diseases in an international perspective. Eur J Public Health 1994;4:258-64.

134. Kopjar B. Costs of health care for unintentional injury in Stavanger, Norway. Eur J Public Health 1997;7:321-27.

135. Koss LG. The Papanicolaou test for cervical cancer detection. A triumph and a tragedy. Jama 1989;261:737-43.

136. Koss LG, Lin E, Schreiber K, Elgert P, Mango L. Evaluation of the PAPNET cytologic screening system for quality control of cervical smears. Am J Clin Pathol 1994;101:220-9.

137. Koutsky LA, Holmes KK, Critchlow CW, Stevens CE, Paavonen J, et al. A cohort study of the risk of cervical intraepithelial neoplasia grade 2 or 3 in relation to papillomavirus infection. N Engl J Med 1992;327:1272-8.

138. Krabbe PF, Stouthard ME, Essink-Bot ML, Bonsel GJ. The effect of adding a cognitive dimension to the EuroQol multiattribute health-status classification system. J Clin Epidemiol 1999;52:293-301.

139. Kruijshaar ME. Data consistency in summary measures of population health (thesis). Rotterdam: Erasmus University, 2004.

140. Laara E, Day NE, Hakama M. Trends in mortality from cervical cancer in the Nordic countries: association with organised screening programmes. Lancet 1987;1:1247-9.

141. Langley JD, Phillips D, Marshall SW. Inpatient costs of injury due to motor vehicle traffic crashes in New Zealand. Accid Anal Prev 1993;25:585-92.

142. Laverty CR, Thurloe JK, Redman NL, Farnsworth A. An Australian trial of ThinPrep: a new cytopreparatory technique. Cytopathology 1995;6:140-8.

143. Leary TJO, Tellado M, Buckner SB, Ali IS, Stevens A, et al. PAPNET-assisted rescreening of cervical smears: cost and accuracy compared with a 100% manual rescreening strategy [see comments]. Jama 1998;279:235-7.

144. Lee KR, Madge R, Sheets EE. Colposcopically directed biopsy as a basis for comparing the diagnostic accuracy of the ThinPrep and Papanicolaou smear methods. Acta Cytol 1996;40:1047-48.

145. Lin CT, Tseng CJ, Lai CH, Hsueh S, Huang HJ, et al. High-risk HPV DNA detection by Hybrid Capture II. An adjunctive test for mildly abnormal cytologic smears in women > or = 50 years of age. J Reprod Med 2000;45:345-50.

146. Lindgren B. The economic impact of illness. In: Abshagen U, Munnich FE, editors. Cost of illness and benefits of drug treatment. Munich: W. Zuckschwerdt Verlag, 1990:12-20.

147. Lindqvist KS, Brodin H. One-year economic consequences of accidents in a Swedish municipality. Accid Anal Prev 1996;28:209-19.

148. Lips P, Cooper C, Agnusdei D, Caulin F, Egger P, et al. Quality of life in patients with vertebral fractures: validation of the Quality of Life Questionnaire of the European Foundation for Osteoporosis (QUALEFFO). Working Party for Quality of Life of the European Foundation for Osteoporosis. Osteoporos Int 1999;10:150-60.

149. Londesborough P, Ho L, Terry G, Cuzick J, Wheeler C, et al. Human papillomavirus genotype as a predictor of persistence and development of high-grade lesions in women with minor cervical abnormalities. Int J Cancer 1996;69:364-8.

150. Lubitz J, Beebe J, Baker C. Longevity and Medicare expenditures. N Engl J Med 1995;332:999-1003.

151. Luchter S, MacKenzie EJ, editors. Measuring the burden of injury. The 3rd International Conference; 2000 May 15-16; Baltimore, Maryland. NHTSA.

152. Lytwyn A, Sellors JW, Mahony JB, Daya D, Chapman W, et al. Comparison of human papillomavirus DNA testing and repeat Papanicolaou test in women with low-grade cervical cytologic abnormalities: a randomized trial. HPV Effectiveness in Lowgrade Paps (HELP) Study No. 1 Group. Cmaj 2000;163:701-7.

153. Mackenbach JP. Mortality and medical care (thesis). Rotterdam: Erasmus University, 1988.

154. MacKenzie EJ, Cushing BM, Jurkovich GJ, Morris JA, Jr., Burgess AR, et al. Physical impairment and functional outcomes six months after severe lower extremity fractures. J Trauma 1993;34:528-38.

155. MacKenzie EJ, Morris J, Jr., Smith GS, Fahey M. Acute hospital costs of trauma in the United States: implications for regionalized systems of care. J Trauma 1990;30:1096-101.

156. MacKenzie EJ, Morris JA, Jr., Jurkovich GJ, Yasui Y, Cushing BM, et al. Return to work following injury: the role of economic, social, and job-related factors. Am J Public Health 1998;88:1630-7.

157. MacKenzie EJ, Shapiro S, Moody M, Siegel JH, Smith RT. Predicting posttrauma functional disability for individuals without severe brain injury. Med Care 1986;24:377-87.

158. MacKenzie EJ, Shapiro S, Siegel JH. The economic impact of traumatic injuries. One-year treatment-related expenditures. Jama 1988;260:3290-6.

159. MacKenzie EJ, Siegel JH, Shapiro S, Moody M, Smith RT. Functional recovery and medical costs of trauma: an analysis by type and severity of injury. J Trauma 1988;28:281-97.

160. Madsen J, Serup-Hansen N, Kristiansen IS. Future health care costs--do health care costs during the last year of life matter? Health Policy 2002;62:161-72.

161. Malek M, Chang BH, Gallagher SS, Guyer B. The cost of medical care for injuries to children. Ann Emerg Med 1991;20:997-1005.

162. Mango LJ. Computer-assisted cervical cancer screening using neural networks. Cancer Lett 1994;77:155-62.

163. Manos MM, Kinney WK, Hurley LB, Sherman ME, Shieh-Ngai J, et al. Identifying women with cervical neoplasia: using human papillomavirus DNA testing for equivocal Papanicolaou results [see comments]. Jama 1999;281:1605-10.

164. Mathers CD, Penm R, Stevenson C, Carter R. Health system costs of diseases and injury in Australia, 1993-4. Health and Welfare Expenditure Series No 2. Canberra: AIHW, 1998.

165. Matsukura T, Sugase M. Identification of genital human papillomaviruses in cervical biopsy specimens: segregation of specific virus types in specific clinicopathologic lesions. Int J Cancer 1995;61:13-22.

166. Maurette P, Masson F, Nicaud V, Cazaugade M, Garros B, et al. Posttraumatic disablement: a prospective study of impairment, disability, and handicap. J Trauma 1992;33:728-36.

167. Max W, MacKenzie EJ, Rice DP. Head injuries : costs and consequences. J Head Trauma Rehabil 1991;6:76-91.

168. McCarthy ML, MacKenzie EJ, Bosse MJ, Copeland CE, Hash CS, et al. Functional status following orthopedic trauma in young women. J Trauma 1995;39:828-36; discussion 36-7.

169. McClure RJ, Douglas RM. The public health impact of minor injury. Accid Anal Prev 1996;28:443-51.

170. McKeown T. The role of medicine: dream, mirage of nemesis. London: Nuffield Provincial Hospitals Trust, 1976.

171. Meerding WJ. Ziektelast en kosten van ongevallen. Seminar ongevalsslachtoffers in beeld. Amstelveen: Consument en Veiligheid, 2003.

172. Meerding WJ, Birnie E, Mulder S, den Hertog PCd, Toet H, et al. Costs of injury in the Netherlands. Amsterdam: Consumer Safety Institute, 2000:79.

173. Melchers W, van den Brule A, Walboomers J, de Bruin M, Burger M, et al. Increased detection rate of human papillomavirus in cervical scrapes by the polymerase chain reaction as compared to modified FISH and southern-blot analysis. J Med Virol 1989;27:329-35.

174. Merea E, Le Gales C, Cochand-Priollet B, Cartier I, De Cremoux P, et al. Cost of screening for cancerous and precancerous lesions of the cervix. Diagn Cytopathol 2002;27:251-7.

175. Michaels AJ, Michaels CE, Smith JS, Moon CH, Peterson C, et al. Outcome from injury: general health, work status, and satisfaction 12 months after trauma. J Trauma 2000;48:841-8; discussion 48-50.

176. Miller JE, Russell LB, Davis DM, Milan E, Carson JL, et al. Biomedical risk factors for hospital admission in older adults. Med Care 1998;36:411-21.

177. Miller TR. Costs and functional consequences of U.S. roadway crashes. Accid Anal Prev 1993;25:593-607.

178. Miller TR, Lestina DC. Patterns in US medical expenditures and utilization for injury, 1987. Am J Public Health 1996;86:89-93.

179. Miller TR, Pindus NM, Douglass JB. Medically related motor vehicle injury costs by body region and severity. J Trauma 1993;34:270-5.

180. Miller TR, Pindus NM, Douglass JB, et. al. Databook on nonfatal injury : incidence, costs, and consequences. Washington, D.C.: The Urban Institute Press, 1995.

181. Minge L, Fleming M, VanGeem T, Bishop JW. AutoCyte Prep system vs. conventional cervical cytology. Comparison based on 2,156 cases. J Reprod Med 2000;45:179-84.

182. Ministry of Health. Annual overview of health care, 1997 [in Dutch]. The Hague: SDU, 1996.

183. Mock C, MacKenzie E, Jurkovich G, Burgess A, Cushing B, et al. Determinants of disability after lower extremity fracture. J Trauma 2000;49:1002-11.

184. Monsonego J, Autillo-Touati A, Bergeron C, Dachez R, Liaras J, et al. Liquid-based cytology for primary cervical cancer screening: a multi-centre study. Br J Cancer 2001;84:360-6.

185. Mooney G, Wiseman V. Burden of disease and priority setting. Health Econ 2000;9:369-72.

186. Moore R, Mao Y, Zhang J, Clarke K. Economic burden of illness in Canada, 1993: Minister of Public Works and Government Services, 1997.

187. Moss SM, Gray A, Legood R, Henstock E. Evaluation of HPV/LBC: cervical screening pilot studies. First report to the DOH on evaluation of LBC (revised January 2003). Surrey: Institute of Cancer Research, 2003.

188. Mulder S, Blankendaal F, Vriend I, Schoots W, Bouter L. Epidemiological data and ranking home and leisure accidents for priority-setting. Accid Anal Prev 2002;34:695-702.

189. Mulder S, Meerding WJ, Van Beeck EF. Setting priorities in injury prevention: the application of an incidence based cost model. Inj Prev 2002;8:74-8.

190. Munoz N, Bosch FX. HPV and cervical neoplasia: review of case-control and cohort studies. IARC Sci Publ 1992:251-61.

191. Murray CJ, Lopez AD. Progress and directions in refining the global burden of disease approach: a response to Williams. Health Econ 2000;9:69-82.

192. Murray CJL, Lopez AD. Global mortality, disability and the contribution of risk factors. The global burden of disease study. Lancet 1997:1436-42.

193. Murray CJM, Evans DB, eds. Health systems performance assessment: debates, methods and empiricism. Geneva: World Health Organization, 2003.

194. Murray CJM, Lopez AD, eds. The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Cambridge, MA: Harvard School of Public Health, 1996.

195. Murray CJM, Salomon JA, Mathers CD, Lopez AD, eds. Summary measures of population health: concepts, ethics, measurement and applications. Geneva: World Health Organization, 2002.

196. Myers ER, McCrory DC, Subramanian S, McCall N, Nanda K, et al. Setting the target for a better cervical screening test: characteristics of a cost-effective test for cervical neoplasia screening. Obstet Gynecol 2000;96:645-52.

197. Nanda K, McCrory DC, Myers ER, Bastian LA, Hasselblad V, et al. Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. Ann Intern Med 2000;132:810-9.

198. National hospital register (LMR). Data on hospitalizations and medical procedures 1998. Utrecht: Prismant, 1999.

199. National Institute of Clinical Excellence (NICE). Guidance on the use of liquid-based cytology for cervical screening. Technology appraisal 69. London, 2003.

200. NHS cervical screening programme. Cervical screening: a pocket guide. Sheffield: National Health Service, 1996.

201. NHS Executive. Burdens of disease: a discussion document. London: Department of Health, 1996.

202. Nobbenhuis MA, Helmerhorst TJ, van den Brule AJ, Rozendaal L, Voorhorst FJ, et al. Cytological regression and clearance of high-risk human papillomavirus in women with an abnormal cervical smear. Lancet 2001;358:1782-3.

203. Nobbenhuis MA, Meijer CJ, van den Brule AJ, Rozendaal L, Voorhorst FJ, et al. Addition of high-risk HPV testing improves the current guidelines on follow-up after treatment for cervical intraepithelial neoplasia. Br J Cancer 2001;84:796-801.

204. Nobbenhuis MA, Walboomers JM, Helmerhorst TJ, Rozendaal L, Remmink AJ, et al. Relation of human papillomavirus status to cervical lesions and consequences for cervical-cancer screening: a prospective study. Lancet 1999;354:20-5.

205. Noro AM, Hakkinen UT, Laitinen OJ. Determinants of health service use and expenditure among the elderly Finnish population. Eur J Public Health 1999;9:174-80.

206. Nursing homes information system (SIVIS). Patient records 1996. Utrecht: Prismant, 1999.

207. OECD. OECD Health Data: OECD, 2002.

208. Oliver CW, Twaddle B, Agel J, Routt ML, Jr. Outcome after pelvic ring fractures: evaluation using the medical outcomes short form SF-36. Injury 1996;27:635-41.

209. Olsen JA, Smith RD. Theory versus practice: a review of 'willingness-to-pay' in health and health care. Health Econ 2001;10:39-52.

210. Oostenbrink JB, Koopmanschap MA, Rutten FFH. Costing manual for economic evaluations. Amstelveen: CVZ, 2000.

211. Oppe S, De Charro FT. The effect of medical care by a helicopter trauma team on the probability of survival and the quality of life of hospitalised victims. Accid Anal Prev 2001;33:129-38.

212. Ostor AG. Natural history of cervical intraepithelial neoplasia: a critical review. Int J Gynecol Pathol 1993;12:186-92.

213. Papillo JL, Zarka MA, St John TL. Evaluation of the ThinPrep Pap test in clinical practice. A seven-month, 16,314-case experience in northern Vermont. Acta Cytol 1998;42:203-8.

214. Park IA, Lee SN, Chae SW, Park KH, Kim JW, et al. Comparing the accuracy of ThinPrep Pap tests and conventional Papanicolaou smears on the basis of the histologic diagnosis: a clinical study of women with cervical abnormalities. Acta Cytol 2001;45:525-31.

215. Patten SF, Jr., Lee JS, Nelson AC. NeoPath, Inc. NeoPath AutoPap 300 Automatic Pap Screener System. Acta Cytol 1996;40:45-52.

216. Patten SF, Jr., Lee JS, Wilbur DC, Bonfiglio TA, Colgan TJ, et al. The AutoPap 300 QC System multicenter clinical trials for use in quality control rescreening of cervical smears: I. A prospective intended use study. Cancer 1997;81:337-42.

217. Payne SR, Waller JA. Trauma registry and trauma center biases in injury research. J Trauma 1989;29:424-9.

218. Phillips DE, Langley JD, Marshall SW. Injury: the medical and related costs in New Zealand 1990. N Z Med J 1993;106:215-7.

219. Polder JJ. Cost of illness in the Netherlands (thesis). Rotterdam: Erasmus University, 2001.

220. Polder JJ, Jacobs OM, Barendregt JJ. De prijs van grijs. Medisch Contact 2003;58:2034-8.

221. Polder JJ, Takken J, Meerding WJ, Kommer GJ, Stokx LJ. Kosten van ziekten in Nederland: de zorgeuro ontrafeld. Houten: Bohn Stafleu Van Loghum, 2002.

222. Polder JJ, van Balen R, Steyerberg EW, Cools HJ, Habbema JD. A cost-minimisation study of alternative discharge policies after hip fracture repair. Health Econ 2003;12:87-100.

223. Ponzer S, Bergman B, Brismar B, Johansson LM. A study of patient-related characteristics and outcome after moderate injury. Injury 1996;27:549-55.

224. Ponzer S, Nasell H, Bergman B, Tornkvist H. Functional outcome and quality of life in patients with Type B ankle fractures: a two-year follow-up study. J Orthop Trauma 1999;13:363-8.

225. PRISMATIC Project Management Team. Assessment of automated primary screening on PAPNET of cervical smears in the PRISMATIC trial. Lancet 1999;353:1381-5.

226. Quinlan KP, Thompson MP, Annest JL, Peddicord J, Ryan G, et al. Expanding the National Electronic Injury Surveillance System to monitor all nonfatal injuries treated in US hospital emergency departments. Ann Emerg Med 1999;34:637-45.

227. Raab SS. The cost-effectiveness of cervical-vaginal rescreening [see comments]. Am J Clin Pathol 1997;108:525-36.

228. Radensky PW, Mango LJ. Interactive neural-network-assisted screening. An economic assessment. Acta Cytol 1998;42:246-52.

229. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978;299:926-30.

230. Rehabilitation data system. Data on inpatients and outpatients in 1996. Utrecht: VRIN, 1997.

231. Rice DP, MacKenzie EJ. Cost of injury in the United States: a report to Congress. San Francisco: Institute for Health and Ageing, University of California, 1989.

232. RIVM. Volksgezondheid Toekomst Verkenning 1997 deel III: Gezondheid en levensverwachting gewogen. Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu, 1997.

233. RIVM. Gezondheid op koers? Volksgezondheid Toekomst Verkenning 2002. Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu, 2002.

234. Roberts JM, Gurley AM, Thurloe JK, Bowditch R, Laverty CR. Evaluation of the ThinPrep Pap test as an adjunct to the conventional Pap smear. Med J Aust 1997;167:466-9.

235. Roberts RO, Bergstralh EJ, Schmidt L, Jacobsen SJ. Comparison of self-reported and medical record health care utilization measures. J Clin Epidemiol 1996;49:989-95.

236. Rosenthal DL. Automation and the endangered future of the Pap test. J Natl Cancer Inst 1998;90:738-49.

237. Rosenthal DL, Acosta D, Peters RK. Computer-assisted rescreening of clinically important false negative cervical smears using the PAPNET Testing System. Acta Cytol 1996;40:120-6.

238. Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. Stat Med 1991;10:585-98.

239. Rutten FF, Brouwer WB. Meer zorg bij beperkt budget; een pleidooi voor een betere inzet van het doelmatigheidscriterium. Ned Tijdschr Geneeskd 2002;146:2254-8.

240. Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, et al. Eliciting public preferences for healthcare: a systematic review of techniques. Health Technol Assess 2001;5:1-186.

241. Schechter CB. Cost-effectiveness of rescreening conventionally prepared cervical smears by PAPNET testing. Acta Cytol 1996;40:1272-82.

242. Sendi P, Gafni A, Birch S. Opportunity costs and uncertainty in the economic evaluation of health care interventions. Health Econ 2002;11:23-31.

243. Severens JL, Laheij RJF, Jansen JBM, van der Lisdonk EH, Verbeek ALM. Estimating the cost of lost productivity in dyspepsia. Aliment Pharm Therap 1998;12:919-23.

244. Sheets EE, Constantine NM, Dinisco S, Dean B, Cibas ES. Colposcopically directed biopsies provide a basis for comparing the accuracy of ThinPrep and Papanicolaou smears. Journal of Gynecologic Techniques 1995;1:27-33.

245. Sherman ME, Mendoza M, Lee KR, Ashfaq R, Birdsong GG, et al. Performance of liquid-based, thin-layer cervical cytology: correlation with reference diagnoses and human papillomavirus testing. Mod Pathol 1998;11:837-43.

246. Sherman ME, Schiffman M, Herrero R, Kelly D, Bratti C, et al. Performance of a semiautomated Papanicolaou smear screening system: results of a population-based study conducted in Guanacaste, Costa Rica. Cancer 1998;84:273-80.

247. Sherman ME, Schiffman MD, Herrero R, Bratti C, Hildesheim A, et al. Evaluation of conventional and novel cervical cancer screening methods in a population-based study of 10,000 Costa Rican women. Acta Cytol 1995;39:983.

248. Shlay JC, Dunn T, Byers T, Baron AE, Douglas JM, Jr. Prediction of cervical intraepithelial neoplasia grade 2-3 using risk assessment and human papillomavirus testing in women with atypia on papanicolaou smears. Obstet Gynecol 2000;96:410-6.

249. SIG (Information Centre for Health Care). Data files on cervical cancer, 1989-1992. Utrecht, 1994.

250. Sigurdsson K, Arnadottir T, Snorradottir M, Benediktsdottir K, Saemundsson H. Human papillomavirus (HPV) in an Icelandic population: the role of HPV DNA testing based on hybrid capture and PCR assays among women with screen-detected abnormal Pap smears. Int J Cancer 1997;72:446-52.

251. Skehan M, Soutter WP, Lim K, Krausz T, Pryse-Davies J. Reliability of colposcopy and directed punch biopsy [see comments]. Br J Obstet Gynaecol 1990;97:811-6.

252. Smith BL, Lee M, Leader S, Wertlake P. Economic impact of automated primary screening for cervical cancer. J Reprod Med 1999;44:518-28.

253. Solomon D, Schiffman M, Tarone R. Comparison of three management strategies for patients with atypical squamous cells of undetermined significance: baseline results from a randomized trial. J Natl Cancer Inst 2001;93:293-9.

254. Spillman BC, Lubitz J. The effect of longevity on spending for acute and long-term care. N Engl J Med 2000;342:1409-15.

255. Stevens MW, Milne AJ, James KA, Brancheau D, Ellison D, et al. Effectiveness of automated cervical cytology rescreening using the AutoPap 300 QC System. Diagn Cytopathol 1997;16:505-12.

256. Stone DH, Morrison A, Smith GS. Emergency department injury surveillance systems: the best use of limited resources? Inj Prev 1999;5:166-7.

257. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. 2nd ed. Oxford: Oxford University Press, 1995.

258. Sturms LM, van der Sluis CK, Snippe H, Groothoff JW, ten Duis HJ, et al. Spaakverwondingen bij kinderen: toedracht en gevolgen. Ned Tijdschr Geneeskd 2002;146:1691-6.

259. The Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesions Triage Study (ALTS) Group. Human papillomavirus testing for triage of women with cytologic evidence of low-grade squamous intraepithelial lesions: baseline data from a randomized trial. J Natl Cancer Inst 2000;92:397-402.

260. The EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990;16:199-208.

261. Torrance GW. Measurement of health state utilities for economic appraisal. J Health Econ 1986;5:1-30.

262. Tylko S, editor. Measuring the burden of injury. 4th International Conference; 2002 May 16-17; Montreal, Quebec. Transport Canada.

263. Tyrance PH, Jr., Himmelstein DU, Woolhandler S. US emergency department costs: no emergency. Am J Public Health 1996;86:1527-31.

264. US Preventive Task Force. Guide to Clinical Preventive Services. 2nd ed. Washington DC: U.S. Department of Health and Human Services, 1996.

265. van Agt HM, Essink-Bot ML, Krabbe PF, Bonsel GJ. Test-retest reliability of health state valuations collected with the EuroQol questionnaire. Soc Sci Med 1994;39:1537-44.

266. van Balen R. Hip fracture in the elderly (thesis). Rotterdam: Erasmus University, 2003.

267. van Ballegooijen M. Effects and costs of cervical cancer screening (thesis). Rotterdam: Erasmus University, 1998.

268. van Ballegooijen M, Beck S, Boon ME, Boer R, Habbema JD. Rescreen effect in conventional and PAPNET screening: observed in a study using material enriched with positive smears. Acta Cytol 1998;42:1133-8.

269. van Ballegooijen M, Boer R, van Oortmarssen GJ, Koopmanschap MA, Lubbe JTN, et al. Cervical screening: age ranges and intervals. (in Dutch). Rotterdam: Department of Public Health, Erasmus University Rotterdam, 1993.

270. van Ballegooijen M, Habbema JDF, van Oortmarssen GJ, Koopmanschap MA, Lubbe JT, et al. Preventive Pap-smears: balancing costs, risks and benefits. Br J Cancer 1992;65:930-3.

271. van Ballegooijen M, Koopmanschap MA, Habbema JD. The management of cervical intra-epithelial neoplasia (CIN): extensiveness and costs in The Netherlands. Eur J Cancer 1995;31A:1672-6.

272. van Ballegooijen M, Koopmanschap MA, Tjokrowardojo AJ, van Oortmarssen GJ. Care and costs for advanced cervical cancer. Eur J Cancer 1992;28A:1703-8.

273. van Ballegooijen M, van den Akker-van Marle ME, Warmerdam PG, Meijer CJ, Walboomers JM, et al. Present evidence on the value of HPV testing for cervical cancer screening: a model-based exploration of the (cost-)effectiveness. Br J Cancer 1997;76:651-7.

274. van Beeck EF. Injuries: a continuous challenge for public health (thesis). Rotterdam: Erasmus University, 1998.

275. van Beeck EF, van Roijen L, Mackenbach JP. Medical costs and economic production losses due to injuries in the Netherlands. J Trauma 1997;42:1116-23.

276. van Beek J, Janssens L. Werklastmeting op de Eerste Hulp. Amsterdam: Academisch Medisch Centrum, 1995.

277. van den Akker-van Marle ME, van Ballegooijen M, van Oortmarssen GJ, Boer R, Habbema JD. Cost-effectiveness of cervical cancer screening: comparison of screening policies. J Natl Cancer Inst 2002;94:193-204.

278. Van den Bosch EW, Van der Kleyn R, Hogervorst M, Van Vugt AB. Functional outcome of internal fixation for pelvic ring fractures. J Trauma 1999;47:365-71.

279. van der Maas PJ. Applications of summary measures of population health. In: Murray CJM, Salomon JA, Mathers CD, Lopez AD, editors. Summary measures of population health: concepts, ethics, measurement and applications. Geneva: World Health Organization, 2002.

280. van der Maas PJ, de Koning HJ, van Ineveld BM, van Oortmarssen GJ, Habbema JDF, et al. The cost-effectiveness of breast cancer screening. Int J Cancer 1989;43:1055-60.

281. van der Meer JBW. Equal care, equal cure? (thesis). Rotterdam: Erasmus University, 1998.

282. van der Sande R, Lamberts SW, Rooijmans HG. Kennis op de plank? Het nuttig effect van onderzoeken uit het programma van het fonds Ontwikkelingsgeneeskunde. Ned Tijdschr Geneeskd 2003;147:2390-3.

283. van der Sluis CK. Outcomes of major trauma (thesis). Groningen: Rijksuniversiteit Groningen, 1998.

284. van Hout BA, Al MJ, Gordon GS, Rutten FF. Costs, effects and C/E-ratios alongside a clinical trial. Health Econ 1994;3:309-19.

285. van Mosseveld CJPM. International comparison of health care expenditure (thesis). Rotterdam: Erasmus University, 2003.

286. van Oortmarssen GJ, Habbema JD. Duration of preclinical cervical cancer and reduction in incidence of invasive cancer following negative pap smears. Int J Epidemiol 1995;24:300-7.

287. van Oortmarssen GJ, Habbema JDF. Epidemiological evidence for age-dependent regression of pre-invasive cervical cancer. Br J Cancer 1991;64:559-65.

288. Vassilakos P, Schwartz D, de Marval F, Yousfi L, Broquet G, et al. Biopsy-based comparison of liquid-based, thin-layer preparations to conventional Pap smears. J Reprod Med 2000;45:11-6.

289. Vazquez Mata G, Rivera Fernandez R, Perez Aragon A, Gonzalez Carmona A, Fernandez Mondejar E, et al. Analysis of quality of life in polytraumatized patients two years after discharge from an intensive care unit. J Trauma 1996;41:326-32.

290. Visser O, Coebergh JWW, Schouten LJ, van Dijck JAAM. Incidence of cancer in the Netherlands 1997: ninth report of the netherlands cancer registry. Utrecht: Vereniging van Integrale Kankercentra, 2001.

291. Vles W. Trauma registration (thesis). Utrecht: Utrecht University, 2003.

292. Waller JA, Skelly JM, Davis JH. Emergency department care and hospitalization as predictors of disability. J Trauma 1995;39:632-4.

293. Watson WL, Ozanne-Smith J. The cost of injury to Victoria. Clayton: Monash University, 1997.

294. Weinstein MC. From cost-effectiveness ratios to resource allocation: where to draw the line? In: Sloan F, editor. Valuing health care. Cambridge: Cambridge University Press, 1996.

295. Weintraub J, Morabia A. Efficacy of a liquid-based thin layer method for cervical cancer screening in a population with a low incidence of cervical cancer. Diagn Cytopathol 2000;22:52-9.

296. WHO. International Classification of Diseases. 9th ed. Geneva: World Health Organization, 1977.

297. WHO. International Classification of Functioning, Disability and Health. Geneva: World Health Organization, 2001.

298. WHO Working Group on Injury Surveillance Methods. International Classification of External Causes of Injuries (ICECI). Amsterdam: Consumer Safety Institute, 2002.

299. Wilbur DC, Prey MU, Miller WM, Pawlick GF, Colgan TJ. The AutoPap system for primary screening in cervical cytology. Comparing the results of a prospective, intended-use study with routine manual practice. Acta Cytol 1998;42:214-20.

300. Wilbur DC, Prey MU, Miller WM, Pawlick GF, Colgan TJ, et al. Detection of high grade squamous intraepithelial lesions and tumors using the AutoPap System: results of a primary screening clinical trial. Cancer 1999;87:354-8.

301. Will BP, Berthelot JM, Le Petit C, Tomiak EM, Verma S, et al. Estimates of the lifetime costs of breast cancer treatment in Canada. Eur J Cancer 2000;36:724-35.

302. Williams A. Calculating the global burden of disease: time for a strategic reappraisal? Health Econ 1999;8:1-8.

303. Williams RM. The costs of visits to emergency departments. N Engl J Med 1996;334:642-6.

304. Wright TC, Sun XW, Koulos J. Comparison of management algorithms for the evaluation of women with low-grade cytologic abnormalities. Obstet Gynecol 1995;85:202-10.

305. Zethraeus N, Gerdtham UG. Estimating the costs of hip fracture and potential savings. Int J Technol Assess Health Care 1998;14:255-67.

Summary

Population health has improved considerably in the Netherlands, as can be illustrated by the increase in life expectancy since the mid 19th century. Part of this improvement can be attributed to an increased access to health care and the rapid development of medical technology, in particular since 1950. This also led to an increase in health care costs. Currently about 10% of gross national product is spent on health care. To contain these costs and at the same time to further improve the population's health, we need to give priority to health problems where interventions are most needed and rewarding. To inform these choices, data are needed on how health care costs and population health are distributed by diseases, risk factors, and population groups. In other words, we need to know where the money goes and where the need for improving health is highest. Also, information is needed on the cost-effectiveness (efficiency) of health interventions to select interventions that provide the most 'value for money'. In this thesis we present descriptive data on health care costs and population health, and information on the cost-effectiveness of interventions, with applications in injuries (part 1) and cervical cancer (part 2). In addition, the relative importance is discussed of descriptive data on population health and health care versus economic evaluations.

Cost of illness and injury in the Netherlands
The thesis starts with a generic cost of illness (COI) study, showing that the largest proportions of health care resources are spent on chronic, disabling diseases (chapter 2). Mental diseases such as mental retardation (Down's syndrome) and dementia, and musculoskeletal disorders are among the top 5 of diagnostic groups with the highest costs. The main causes of death, i.e. stroke, all cancers combined, and coronary heart disease, rank among the top 10 with shares of 2.5-3% of health care costs, but costs of dental diseases (4%) are higher. The average health care costs per capita are relatively high in the first year of life, low during childhood and adulthood, and increase exponentially after age 50. Women contribute most to health care costs (almost 60%) which is explained by their longer life expectancy and the costs of reproductive care.

The implications are far-reaching. The skewed cost distribution by age has important consequences for societies with an ageing population. Health care needs can only increase if life expectancy increases and if costly disabling diseases (e.g. dementia, osteoarthritis, hip fracture) remain resistant to control.

In this study we broke down health care costs in 1994 using sector-specific administrative data on health care use (cross-sectional design). For specific diseases more detailed cost estimates would be useful for policy

making. For instance, injuries have heterogeneous causes, and differ in terms of severity and health care need. In this area, costs can be a useful summary measure to quantify the relative importance of specific injuries. We estimated medical costs of injury (excluding adverse medical events) at national level by type of injury and health care sector, and identified the major determinants (chapter 3). We developed an incidence-based costing model linked to the continuous national Injury Surveillance System (LIS). In LIS a representative sample of Emergency Department (ED) visits is recorded, and the system includes extensive information on the cause of injury. We collected health care consumption data from administrative systems (e.g. the hospital discharge register) and from a follow-up study among 5,755 injury patients.

Total health care costs of injury in 1998 amounted to 1.1 billion euro, or more than 3% of total health care costs. Injuries with the highest costs are hip fractures (21%), superficial injuries (14%), open wounds, skull-brain injury, and knee/lower leg fractures (each 6%). These high costs are due to a high frequency (e.g. superficial injury) or due to high costs per patient (e.g. hip fracture, skull-brain injury). The same accounts for the cost distribution by age and sex: costs are relatively high in young adult males (high frequency) and in elderly females (high costs per patient). Two-third of costs are due to hospitalized patients, that account for 9% of all injuries.

Whereas it is widely known that hip fractures cause high costs, this study also identified minor injuries as a major source of health care costs: superficial injury and open wounds together account for one fifth of injury costs. A greater efficiency might be achieved when the treatment of these minor injuries can be shifted towards primary care.

A comparison with studies from other highly developed countries revealed that differences in the reported health care costs of injury are large (chapter 4). Per capita health care costs ranged from $35-275 (year 2000 international dollars). We first analysed whether observed differences could be explained by differences in methodologies. Within-country differences were up to 40% and are by definition caused by methodological differences, particularly differences in case definition, cost items included, and approach (bottom-up versus top-down). However, real differences in injury epidemiology and costs per patient seemed to explain a larger part of between-country differences in costs of injury. Our estimates from the Netherlands occupy an intermediate position. Per capita costs were highest in the US, because of a higher incidence and higher costs per patients. Also in Australia costs were higher than in the Netherlands, despite a lower incidence, due to three-fold higher costs per patient. Per capita costs were lowest in Sweden, Norway and New Zealand. For the subcategory traffic injuries, the within-country differences in costs per

capita were higher (up to 60%) as were the between-country differences ($2-$116 per capita). Also a larger proportion of the international differences could be attributed to differences in methodology than for all injuries together. Studies that included productivity losses due to injury showed that these were consistently almost three-fold higher than the medical costs.

Functional outcome in injury patients
Apart from health care costs, quantitative data on the functional outcome of injury, and of its major determinants, is important to direct the development of preventive interventions and trauma care. With the decline in overall injury mortality, the importance of injury-related disability has increased. Because of the many functional sequelae and recovery patterns of injuries, the measurement of disability is a necessary but also challenging task. Generic (not disease-specific) instruments thereby enable a uniform comparison of injuries among each other and with other health problems. We measured disability in the first year post-injury in a comprehensive population of surviving injury patients presenting at EDs, both non-hospitalized and hospitalized, with the EuroQol generic instrument (chapter 5). This instrument measures mobility, the ability to perform self care and usual activities, pain/discomfort, and anxiety/depression. We added a question on cognitive disability. The resulting score profiles can be converted to a summary measure (utility) between 0 and 1 representing the overall level of health.

After two months the average health status of non-hospitalized patients was comparable to the general population's health, and 95% of workers had returned to work. However, non-hospitalized patients with vertebral column or extremity injury had less than normal levels of health, and patients with upper extremity fractures reported the highest work absence. The mean health status of hospitalized patients was far below general population norms at 2 months, improved up to 5 months, but stabilized thereafter at a suboptimal level, predominantly in patients with a long hospital stay. Among workers, 40%, 20% and 10% had not (yet) returned to work after 2, 5 and 9 months, respectively. Hospitalized patients with hip fractures, injuries to the vertebral column and spinal cord, and other lower extremity fractures reported the worst health status, also adjusted for age and sex. High levels of cognitive limitations were measured in patients with skull-brain injury (e.g up to 40% in those with skull fracture or intracranial injury at 2 months). These patients could insufficiently be discriminated by the EuroQol. Injury diagnosis, hospitalization, hospital duration, IC use, motor vehicle involvement, and number of injuries all were independent predictors of disability. A lower socio-economic status was associated with a worse health status and longer work absence.

Evaluation of new cytological tests for cervical cancer screening

Part 1 (chapters 2-5) of this thesis was devoted to describing the medical costs and burden of disease (specifically injuries). In part 2 (chapters 6-8) we changed the focus to health interventions. We investigated the efficiency of two measures to improve population based cervical cancer screening: the introduction of new cytological screening tests (chapters 6 and 7) and the role of human papillomavirus (HPV) testing in women with abnormal smears (chapter 8). Cervical cancer screening has occurred for decades in the Netherlands, and organized screening is offered since the late 1980s. Screening has always been done by the Pap smear. This test is often criticized for being too insensitive. Estimates of its sensitivity vary between 60-90%. Recently, automated technologies and liquid based ('thin layer') cytology (LBC) have been developed to improve test performance. We investigated the test characteristics of these new technologies, and their (cost-)effectiveness compared to screening with Pap smears. In this evaluation it is essential to distinguish between single test sensitivity and programme sensitivity. Because the natural course of cervical cancer takes on average many years – the duration of pre-invasive stages is estimated at 12 years on average – women who attend screening at regular intervals have generally more than one chance to be detected timely. The programme sensitivity therefore depends on both single test sensitivity and screening intensity.

In a systematic review of published trials, that were assessed with a standard list of quality criteria, we found that there is weak evidence that one liquid based system (ThinPrepTM) and one automated system (AutoPapTM) are more sensitive than the Pap test, at the loss of some specificity (maximum reported increase in sensitivity of 12%). Using the MISCAN microsimulation model for the evaluation of cervical cancer screening, we designed a decision analytic framework, and quantified for which combinations of test sensitivity, test specificity, and incremental unit costs per test a new screening test would be as cost-effective as the Pap test. We used the MISCAN model, which has been extensively validated and is a flexible tool for the prospective evaluation of changes in the screening programme. In the Dutch situation (screening between age 30 and 60 at 5-year intervals) a hypothetical test with optimal (100%) sensitivity and specificity may cost additionally €9 per test to be as cost-effective as the Pap test. This unit cost threshold is lower in countries with more intensive screening (because incremental health gains of a more sensitive test are lower), but is higher in countries where the Pap test sensitivity is lower than assumed at baseline. The baseline Pap test sensitivity (80%) has been derived from observed screening data. Comparing the findings in our decision analytic framework with the observed test characteristics and unit costs of current new

technologies, we concluded that it is unlikely that these technologies are as cost-effective as the Pap test. This is even less likely in countries with more intensive screening, such as the US and the UK where LBC is already widely used.

HPV triage of women with abnormal smears

The number of women with abnormal Pap smears is a continuous concern in screening programmes. Because the majority of women with a positive screening test would never develop cervical cancer in the absence of screening, due to spontaneous regression, they run the risk of overtreatment. Currently, women with (persistent) abnormal smears are referred for colposcopy, treated if necessary, and kept under surveillance. We investigated whether hr-HPV (high risk human papillomavirus) testing of women with at least persistent mild or moderate dyskaryosis increases the efficiency of further management of these women (chapter 8).

We compared three policies: a) conventional management based on colposcopy and biopsy, with only histological positive women getting treatment (LETZ treatment) b) testing women on hr-HPV, with hr-HPV positive women being treated directly without prior histological assessment, and conventional management for hr-HPV negative women, and c) a policy by which all women receive LETZ treatment directly without prior histological assessment. For each policy we calculated the mean number of clinical procedures and costs per woman. We used data on HPV-status and medical procedures from a follow-up study among 221 women who were referred for colposcopy because of persistent mild or moderate dyskaryosis or a single smear reported as severe dyskaryosis.

We demonstrated that hr-HPV triage (b), compared to conventional management (a), would avoid histological assessment at the expense of some overtreatment in women with persistent mild or moderate dyskaryosis: per woman 0.51 colposcopically directed biopsies would be avoided, but this should be weighed against 0.05 extra LETZ and $134 additional costs per woman. In other words, the benefit of 10 avoided biopsies should be weighed against the burden of one LETZ treatment. In women with one severely dyskaryotic smear, direct treatment with LETZ (c) would be more efficient than triage with HPV testing (b). Compared with conventional management (a), a direct treatment policy (c) would avoid 25 colposcopically directed biopsies for each additional LETZ treatment.

Our findings were robust because they did not change much when we used data from foreign studies on HPV prevalence in similar patient groups. However, our HPV triage policy appeared to be much less efficient in women with only mildly abnormal smears (ASCUS or mild dysplasia). Because data on

the quality of life implications of the different management policies were unavailable, we could not conclude which policy would be preferred by the women.

The merits of burden of disease and cost of illness estimates versus economic evaluation
In the General discussion (chapter 9) we integrated the findings in this thesis by discussing the relative contribution of burden of disease (BOD) and cost of illness (COI) studies (part 1) on the one hand, and economic evaluations (part 2) on the other hand. We argue that both types of analyses are necessary and complementary for resource allocation in health care. BOD and COI studies provide essential information to identify diseases, risk factors and population groups with the highest need for intervention. Comprehensive studies have the advantage of providing comparative and internal consistent data, whereas disease-specific studies may raise unjustified attention for individual health problems. In addition, comparative information on population health and costs may generate hypotheses about the explanation of observed differences. Because BOD and COI studies provide national estimates in disease burden and costs, they may be an input into and reference framework for economic evaluations of (combinations of) interventions. As BOD and COI studies provide essential data regarding equity of health needs and access to health care, respectively, information on the efficiency of health care (costs per incremental health gains) is provided by economic evaluations. Although we recognize the importance of economic evaluations, we have pointed to a number of shortcomings and additional conditions for an effective use. We mention here the need for acceptable (ranges of) cost-effectiveness thresholds, knowledge of the impact of single interventions on health budgets and on the size and distribution of health gains on population level. In addition to new interventions, also the existing situation ('usual care') should be evaluated, that is usually a heterogeneous mixture of interventions with insufficient evidence about their individual or combined effectiveness and efficiency.

Our conclusions are as follows:
1. Cost of injury estimates are a useful indicator for the importance of specific injuries. Together with epidemiological indicators they should be used to prioritize the development of interventions that prevent injuries and improve trauma care. Among others the efficiency of the current treatment of minor injuries should be investigated.
2. The lack of methodological consistency of cost of injury studies compromises their policy relevance. Guidelines should be developed to improve the comparability and transparency of these studies.

3. It is well possible to validly measure and compare the functional outcomes of heterogeneous injuries. Our empirical data should be complemented with data on short-term disability in minor injuries, and with data on permanent disability. Future research should also identify the major risk factors contributing to the burden of injury.

4. It is unlikely that liquid based cytology and automated screening technologies are currently as cost-effective as the Pap test for cervical cancer screening. We recommend that the test characteristics of these technologies should be estimated more accurately. Also, it should be investigated whether the current decision processes for implementing and financing new diagnostic tests guarantee an efficient use of health care resources.

5. Triage of women with persistent mild or moderate dyskaryosis, based on high-risk human papillomavirus testing, could lead to a less burdensome treatment of these women. For a definitive answer women's preferences for alternative follow-up policies should be investigated.

6. Comprehensive burden of disease and cost of illness estimates, and economic evaluations are complementary for the improvement of health systems efficiency. We recommend that burden of disease and cost of illness studies are regularly updated. Research should be conducted into acceptable cost-effectiveness thresholds for health care interventions, and cost-effectiveness analyses should evaluate the existing situation ('usual care') in addition to new interventions.

Samenvatting

Afgaande op de toegenomen levensverwachting sinds het midden van de 19e eeuw, is de volksgezondheid in Nederland sterk verbeterd. Deze verbetering is deels te danken aan een verbeterde toegang tot gezondheidszorg en aan de snelle ontwikkeling van de medische technologie, vooral vanaf het midden van de vorige eeuw. Dit heeft ook tot een stijging in de zorguitgaven geleid. Op dit moment wordt circa 10% van het bruto nationaal product uitgegeven aan gezondheidszorg. Om de kosten in de hand te houden en tegelijk de volksgezondheid verder te verbeteren, is het nodig om voorrang te geven aan gezondheidsproblemen waar interventies zowel het meest noodzakelijk als het meest efficiënt zijn. Om deze keuzen te kunnen maken zijn gegevens nodig over hoe de zorguitgaven en de volksgezondheid zijn verdeeld over ziekten, risicofactoren, en bevolkingsgroepen. Met andere woorden, waar wordt het geld aan besteed en waar is de behoefte aan een betere gezondheid het grootst. Daarnaast is informatie nodig over de kosten-effectiviteit (efficiëntie) van zorg, om zodoende de toegang te beperken tot die interventies die de meeste gezondheidswinst opleveren tegen de minste kosten. Beide typen gegevens komen in dit proefschrift aan bod, toegepast op twee volksgezondheids-terreinen: ongevallen (deel 1) en baarmoederhalskanker (deel 2). Ook wordt ingegaan op het relatieve belang van beschrijvende gegevens van de volksgezondheid en gezondheidszorg enerzijds, en van economische evaluaties anderzijds.

Kosten van ziekten en ongevalsletsels in Nederland
Het proefschrift begint met een generieke kosten van ziekten (KVZ) studie, waaruit blijkt dat het grootste deel van de gezondheidszorg wordt besteed aan chronische, invaliderende ziekten (hoofdstuk 2). Psychische aandoeningen, zoals zwakzinnigheid (Down's syndroom) en dementie, en ziekten van het bewegingsapparaat behoren tot de vijf aandoeningen met de hoogste kosten. De belangrijkste doodsoorzaken – beroerten, kanker, en coronaire hartziekten – staan weliswaar in de top 10 van ziekten met de hoogste kosten, maar hun aandeel van elk 2,5-3% is lager dan het aandeel van gebitsziekten (4% van de totale kosten). De gemiddelde kosten per hoofd van de bevolking zijn relatief hoog voor nuljarigen, laag bij kinderen en jongvolwassenen, en stijgen exponentieel na de leeftijd van 50 jaar. De meeste kosten (bijna 60%) komen voor rekening van vrouwen, wat komt door hun hogere levensverwachting en door zwangerschapsgerelateerde kosten.

Dit alles heeft verreikende implicaties. De scheve kostenverdeling met hoge kosten op oudere leeftijd heeft voorspelbare gevolgen voor vergrijzende

populaties. Als de levensverwachting stijgt en tegelijkertijd dure invaliderende ziekten (bijv. dementie, osteoarthritis, heupfracturen) nauwelijks kunnen worden tegengegaan, zullen de zorguitgaven slechts stijgen.

De kostenschattingen in deze studie zijn het resultaat van een opsplitsing van de zorgkosten in 1994 aan de hand van administratieve gegevens van zorggebruik per zorgsector (cross-sectioneel). Bij sommige gezondheidsproblemen zijn echter meer gedetailleerde gegevens noodzakelijk voor beleidsontwikkeling. Dit geldt bijvoorbeeld voor ongevallen, waarvan de oorzaken zeer heterogeen zijn, terwijl ook de gevolgen voor de slachtoffers erg kunnen verschillen qua ernst en zorgbehoefte. Gezien deze heterogeniteit kunnen de kosten van ongevallen een nuttige samengestelde maat zijn om het relatieve belang van specifieke ongevalscategorieën aan te geven. Wij hebben daarom onderzocht hoe de totale medische kosten van ongevallen (uitgezonderd medische oorzaken) verdeeld zijn over letseltypen en zorgsectoren, en wat de belangrijkste determinanten van kosten zijn (hoofdstuk 3). Hiervoor werd een incidentie-gebaseerd kostenmodel ontwikkeld die gekoppeld werd aan het Letsel Informatie Systeem (LIS). In het LIS wordt een representatieve steekproef van ongevalspatiënten op de Spoedeisende Hulp (SEH) geregistreerd, met uitgebreide toedrachtsinformatie. Gegevens over zorggebruik werden verzameld uit administratieve systemen (bijv. de ziekenhuisopnameregistratie) en middels een enquête onder 5.755 ongevalspatiënten.

De zorgkosten van ongevallen in 1998 bedroegen 1,1 miljard euro, meer dan 3% van de totale zorguitgaven. De meeste kosten waren voor heupfracturen (21%), oppervlakkige letsels (14%), open wonden, schedel-hersenletsel, en knie- en onderbeenfracturen (elk 6%). Hoge kosten zijn het gevolg van hoge aantallen letsels (zoals bij oppervlakkige letsels) of van dure letsels (zoals heupfracturen, schedel-hersenletsel). Ook de verdeling van de kosten naar leeftijd en geslacht kan zo worden verklaard: hoge kosten bij jongvolwassen mannen (hoge aantallen) en bij oudere vrouwen (dure letsels). Van de patiënten wordt 9% opgenomen, maar deze zijn samen verantwoordelijk voor tweederde van de zorgkosten.

Hoewel het algemeen bekend is dat heupfracturen veel kosten veroorzaken, geeft dit onderzoek aan dat ook milde letsels een belangrijke oorzaak van zorgkosten zijn: oppervlakkige letsels en open wonden zijn samen goed voor eenvijfde van de kosten. Mogelijk kan de efficiency worden verbeterd door de behandeling van deze letsels meer te verplaatsen naar de eerstelijnszorg.

We hebben onze resultaten ook vergeleken met onderzoeken uit andere hoogontwikkelde landen (hoofdstuk 4). Hieruit bleek dat de internationale

verschillen in gerapporteerde zorgkosten van ongevallen erg groot zijn. De gemiddelde kosten per hoofd van de bevolking bedroegen $35-$275 (internationale dollars peiljaar 2000). Deze verschillen kunnen deels verklaard worden door methodologische verschillen. Kostenschattingen binnen eenzelfde land verschilden ten hoogste 40% van elkaar. Dit is per definitie verklaarbaar door verschillen in methodologie, zoals selectie van patiëntgroepen, kostendefinities, en een bottom-up of top-down aanpak. Internationale verschillen in kosten werden echter grotendeels veroorzaakt door werkelijke verschillen in het voorkomen van ongevallen en in behandelkosten. Nederland neemt qua kosten een tussenpositie in. De gemiddelde kosten per hoofd van de bevolking waren het hoogst in de VS, zowel door een hogere incidentie als door hogere behandelkosten per patiënt. Ook in Australia waren de kosten hoger dan in Nederland, ondanks een lagere incidentie, door driemaal hogere behandelkosten per patiënt. De kosten waren het laagst in Zweden, Noorwegen, en Nieuw Zeeland. Voor verkeersongevallen bleken de gemiddelde kosten per hoofd van de bevolking internationaal nog meer uiteen te lopen: $2-$116. Methodologische verschillen speelden hier een nog grotere rol. Verschillende onderzoeken keken behalve naar medische kosten ook naar andere maatschappelijke kosten. Zo werden de productiviteitskosten (ziekteverzuim e.d.) consistent bijna drie keer hoger geschat dan de medische kosten van ongevallen.

Functionele gevolgen van ongevalsletsels
Naast zorgkosten zijn ook gegevens over de functionele gevolgen van ongevallen, en de determinanten hiervan, van belang voor het doelgericht ontwikkelen van preventieve interventies en van traumazorg. Doordat de sterfte door ongevallen is gedaald, is het relatieve belang van functionele beperkingen bij de overlevers gestegen. Vanwege de heterogeniteit in functionele gevolgen en herstelpatronen bij ongevalsletsels, is het meten van beperkingen niet alleen noodzakelijk maar ook uitdagend. Met generieke (niet-ziektespecifieke) instrumenten is het mogelijk om ongevalsletsels op een eenduidige wijze met elkaar en met andere gezondheidsproblemen te vergelijken. Wij hebben functionele gevolgen gemeten in het eerste jaar na het ongeval bij een brede populatie ongevalspatiënten die zijn behandeld op de SEH, zowel niet-opgenomen als opgenomen patiënten (hoofdstuk 5). We gebruikten de EuroQol, een generiek instrument waarmee beperkingen worden gemeten ten aanzien van mobiliteit, zelfverzorging, dagelijkse activiteiten, pijn of ongemak, en angst of somberheid. Een vraag over cognitieve beperkingen werd toegevoegd. De EuroQol scoreprofielen kunnen worden vertaald in een

somscore (utiliteit) tussen 0 en 1 waarmee de algehele kwaliteit van de gezondheidstoestand wordt aangegeven.

Na 2 maanden was de gemiddelde gezondheidstoestand van *niet-opgenomen* letselpatiënten vergelijkbaar met de algemene bevolking, en van de mensen met een betaalde baan was 95% weer aan het werk. Patiënten met letsel aan de wervelkolom of met extremiteitenletsel hadden echter een minder goede gezondheid. Patiënten met een armfractuur verzuimden het langst. Bij *opgenomen* patiënten was de gezondheid na 2 maanden gemiddeld ver beneden het normale niveau. De gezondheidstoestand verbeterde tot 5 maanden, en bleef daarna stabiel onder het normale niveau, vooral bij patiënten met een lange opnameduur. Van de mensen met een betaalde baan was 40%, 20% en 10% nog niet aan het werk na, respectievelijk, 2, 5 en 9 maanden. Patiënten met een heupfractuur, met letsel aan wervelkolom of ruggenmerg, of met een beenfractuur waren het minst gezond, ook na correctie voor leeftijd en geslacht. Veel patiënten met schedel-hersenletsel hadden cognitieve beperkingen (na 2 maanden bijvoorbeeld 40% van de patiënten met een schedelfractuur of hersenletsel). Deze patiënten worden onvoldoende onderscheiden door de EuroQol. Het letseltype, opnamestatus, opnameduur, IC opname, al of geen motorvoertuigongeval, en het aantal letsels bleken onafhankelijke voorspellers van beperkingen. Een lagere sociaal-economische status hing samen met een slechtere gezondheid en een langere verzuimduur.

Evaluatie van nieuwe cytologische technieken voor screening op baarmoederhalskanker
Deel 1 van het proefschrift gaat vooral over beschrijving van de medische kosten en van de ziektelast. In deel 2 richten we onze blik op gezondheidsinterventies. Wij evalueerden twee maatregelen waarmee de efficiëntie van screening op baarmoederhalskanker mogelijk verbeterd kan worden: de introductie van nieuwe cytologische screeningstesten (hoofdstukken 6 en 7) en het testen op humaan papillomavirus (HPV) bij vrouwen met een afwijkende uitstrijk (hoofdstuk 8). Reeds tientallen jaren wordt in Nederland gescreend op baarmoederhalskanker. Sinds eind jaren '80 bestaat er een landelijk bevolkingsonderzoek. De screening wordt uitgevoerd middels een uitstrijk (Pap test). De betrouwbaarheid van deze test wordt vaak betwijfeld. Schattingen van de sensitiviteit lopen uiteen van 60-90%. Recentelijk zijn geautomatiseerde technologieën en dunnelaag-cytologie systemen (liquid based cytologie) ontwikkeld, die betere testeigenschappen pretenderen. Wij onderzochten de testeigenschappen van deze technologieën en hun (kosten-)effectiviteit vergeleken met de uitstrijk. Het is van belang om in deze evaluatie onderscheid te maken tussen de sensitiviteit van een individuele test en de programmasensitiviteit. Omdat het natuurlijk beloop van

baarmoederhalskanker gemiddeld vele jaren in beslag neemt – de duur van pre-invasieve stadia wordt geschat op gemiddeld 12 jaar – hebben vrouwen die regelmatig worden gescreend in het algemeen meer dan één kans om op tijd te worden ontdekt. De programmasensitiviteit hangt daarom af van de sensitiviteit van een individuele test en van de intensiteit van screening.

In een systematische review van gepubliceerde trials, die uniform beoordeeld werden aan de hand van een lijst met kwaliteitscriteria, vonden we slechts zwak aangetoond dat één systeem voor dunnelaag-cytologie (ThinPrep™) en één geautomatiseerd systeem (AutoPap™) een hogere sensitiviteit (maximaal 12% hoger) hadden dan de Pap test, ten koste van een enigszins lagere specificiteit. Met het MISCAN microsimulatie model voor de evaluatie van screening op baarmoederhalskanker, ontwikkelden we een besliskundig raamwerk waarin we berekenden voor welke combinaties van sensitiviteit, specificiteit, en testkosten een nieuwe screeningstest even kosten-effectief zou zijn als de Pap test. Met het uitgebreid gevalideerde MISCAN model kunnen op flexibele wijze potentiële veranderingen in het screeningsprogramma worden geëvalueerd. Het bleek dat in Nederland, waar vrouwen tussen de 30 en 60 jaar om de vijf jaar onderzocht worden, een hypothetische test met een optimale (dus 100%) sensitiviteit en specificiteit maximaal €9 per test meer zou mogen kosten om even kosten-effectief te zijn als de Pap test. Dit drempelbedrag is echter lager in landen met intensievere screening (omdat de potentiële gezondheidswinst minder is), maar is hoger als de sensitiviteit van de Pap test lager is dan verondersteld (80%). De sensitiviteit van de Pap test is afgeleid van empirische screeningsdata. Wanneer deze bevindingen worden vergeleken met de geobserveerde testeigenschappen en kosten van nieuwe technologieën, is het onwaarschijnlijk dat de huidige technologieën net zo kosten-effectief zijn als de Pap test. Dit geldt des te sterker in landen waar intensievere screening plaatsvindt, zoals de VS en Engeland, waar dunnelaag-cytologie overigens reeds breed geïmplementeerd is.

HPV triage van vrouwen met een afwijkende uitstrijk
Het is belangrijk hoeveel gescreende vrouwen test-positief blijken te zijn. Omdat deze vrouwen in meerderheid nooit baarmoederhalskanker zouden krijgen door spontane regressie, lopen zij risico op overbehandeling. Op dit moment worden vrouwen met een sterk afwijkende uitstrijk of met een positieve herhalingsuitstrijk verwezen voor colposcopie, en indien nodig behandeld en regelmatig gecontroleerd. Wij onderzochten of de behandeling van verwezen vrouwen efficiënter kan door ze te testen op hr-HPV (hoog risico humaan papillomavirus) (hoofdstuk 8).

We vergeleken drie soorten beleid: a) conventionele behandeling waarbij vrouwen colposcopisch gebiopteerd worden, en bij histologische afwijkingen behandeld worden (LETZ behandeling), b) hr-HPV triage, waarbij hr-HPV positieve vrouwen direct worden behandeld (zonder eerst colposcopie of biopsie), en hr-HPV negatieve vrouwen conventioneel behandeld worden, en c) een beleid waarbij alle vrouwen direct behandeld worden met LETZ zonder voorafgaand histologisch onderzoek. Voor elk beleid werd het gemiddeld aantal medische procedures per vrouw berekend alsook de kosten. We gebruikten hiervoor gegevens over de HPV-prevalentie en de behandeling van 221 vrouwen die verwezen waren voor colposcopie vanwege twee keer een uitstrijk met milde of matige dysplasie, of vanwege één uitstrijk met ernstige dysplasie.

Vergeleken met conventionele behandeling (a) kan door HPV-triage (b) van vrouwen met herhaalde milde of matige dysplasie histologisch onderzoek worden voorkomen ten koste van enige overbehandeling: per vrouw zouden gemiddeld 0.51 minder colposcopische biopsieën plaatsvinden, tegen gemiddeld 0.05 extra LETZ-behandelingen en $134 extra kosten per vrouw. Met andere woorden, het voordeel van 10 voorkomen biopsieën moet worden afgewogen tegen de belasting van 1 LETZ behandeling. Bij vrouwen met één uitstrijk met ernstige dysplasie zou directe behandeling (c) efficiënter zijn dan HPV-triage. Vergeleken met conventionele behandeling (a) zou directe behandeling 25 colposcopische biopsieën voorkomen per extra LETZ-behandeling.

Onze bevindingen zijn redelijk zeker omdat de resulaten nauwelijks veranderden wanneer deze werden doorgerekend met gegevens over de hr-HPV prevalentie uit andere studies. De HPV-triage zoals hier beschreven bleek niet efficiënt voor vrouwen met licht afwijkende ('Pap 2') uitstrijken. Omdat we niet de beschikking hadden over kwaliteit van leven data, konden we niet aangeven aan welk beleid vrouwen de voorkeur zouden geven.

Ziektelast en kosten van ziekten versus kosten-effectiviteit van zorg
In de Discussie (hoofdstuk 9) zijn de bevindingen in dit proefschrift geïntegreerd met een discussie over het relatieve belang van gegevens over de ziektelast en kosten van ziekten (KVZ) enerzijds (deel 1), en van economische evaluaties anderzijds (deel 2). Wij beargumenteren dat beide noodzakelijk en complementair zijn voor beslissingen aangaande de verdeling van zorgmiddelen. Ziektelast- en KVZ-studies geven essentiële informatie over de (on)gelijkheid in gezondheid respectievelijk de toegang tot gezondheidszorg, terwijl economische evaluaties informatie geven over de efficiency van gezondheidszorg (kosten per eenheid gezondheidswinst). Met gegevens over

ziektelast en kosten van ziekten kunnen ziekten, risicofactoren, en bevolkingsgroepen geïdentificeerd worden met de grootste behoefte aan zorginterventies. De voorkeur gaat hierbij uit naar brede, veelomvattende ziektelast- en KVZ-studies die onderling vergelijkbare en intern consistente gegevens opleveren. Ziekte-specifieke studies kunnen onterecht de aandacht vestigen op individuele gezondheidsproblemen. Ook kunnen gegevens over de verdeling van volksgezondheid en van kosten hypothesen genereren over mogelijke verklaringen van eventuele verschillen tussen groepen. Schattingen van ziektelast en kosten van ziekten op nationaal niveau kunnen ook gebruikt worden als input en als referentiekader voor economische evaluaties van (combinaties van) interventies. Hoewel gegevens over de efficiëntie van zorg via economische evaluaties esentieel zijn voor het maken van beleidskeuzes, hebben we ook gewezen op een aantal tekortkomingen en op additionele voorwaarden voor het gebruik ervan. Te noemen zijn onder andere: wat is een maatschappelijk acceptabele grens (of bandbreedte) van kosten-effectiviteit, en wat zijn de consequenties van invoering van een interventie voor het zorgbudget en voor de volksgezondheid. Verder kan het bij de evaluatie van een nieuwe interventie noodzakelijk zijn om ook het totale pakket aan huidige zorg ('usual care') te evalueren op het niveau van ziekten. De huidige behandeling van specifieke ziekten betreft veelal een heterogene mix van interventies, waarbij onvoldoende bewijs bestaat over de (kosten-)effectiviteit van elk van deze interventies individueel, laat staan van de gecombineerde kosten-effectiviteit.

Onze conclusies zijn de volgende:
1. Schattingen van de kosten van ongevalsletsels zijn een nuttige indicator voor het belang van specifieke ongevallen. Samen met epidemiologische indicatoren zouden kosten gebruikt moeten worden voor het prioriteren van de ontwikkeling van interventies ter preventie van ongevallen en voor de verbetering van de traumazorg. De efficiëntie van de huidige zorg voor patiënten met lichte letsels zou bijvoorbeeld moeten worden onderzocht.
2. Het gebrek aan methodologische consistentie bij ramingen van de kosten van ongevallen ondergraaft de beleidsrelevantie ervan. Richtlijnen zijn nodig om deze schattingen meer vergelijkbaar en transparant te maken.
3. Het is goed mogelijk om de functionele gevolgen van verschillende ongevalsletsels op een valide wijze te meten en te vergelijken. Onze empirische gegevens zouden moeten worden aangevuld met gegevens over korte termijn beperkingen bij lichte letsels, en met gegevens over permanente gevolgen van ongevalsletsels. Ook zou moeten worden onderzocht welke risicofactoren in belangrijke mate bijdragen aan de ziektelast van ongevallen.

4. Het is niet waarschijnlijk dat dunnelaag-cytologie en geautomatiseerde screeningstechnieken op dit moment kosten-effectieve alternatieven zijn voor de Pap test bij screening op baarmoederhalskanker. De testeigenschappen van deze nieuwe technologieën zouden op een betrouwbare wijze moeten worden vastgesteld. Ook zou moeten worden bekeken of huidige beleidsprocessen bij de toelating en financiering van nieuwe diagnostische tests een efficiënt gebruik van zorgmiddelen garanderen.

5. De triage van vrouwen met persistente lichte of matige dysplasie, door middel van een test op hoog-risico typen van het humaan papillomavirus, zou kunnen leiden tot een minder belastende behandeling van deze vrouwen. Voor een definitief antwoord moeten echter de voorkeuren van vrouwen voor alternatieve vormen van behandelbeleid worden onderzocht.

6. Vergeleken met economische evaluatiestudies leveren schattingen van de ziektelast en van kosten van ziekten op nationaal niveau een complementaire bijdrage aan een efficiëntere gezondheidszorg. Wij bevelen aan dat schattingen van de ziektelast en van kosten van ziekten op nationaal niveau regelmatig geactualiseerd worden. Verder zou moeten worden onderzocht wat maatschappelijk acceptabele grenswaarden zijn voor kosten-effectiviteit in de gezondheidszorg. In kosten-effectiviteitsanalyses van nieuwe zorginterventies zou ook het huidige pakket aan zorg bij specifieke ziekten moeten worden geëvalueerd.

Dankwoord

Lieve Margreet, ik hoop dat het het waard is geweest. Zonder jouw mentale en praktische steun was het nooit gelukt. Lieve Rinko en Loek, het 'Engelse boekje' is af, maar begin er voorlopig nog maar niet in te lezen. Samen vormen jullie een liefdevol nest waar ik dagelijks naar mag terugverlangen.

Ik dank God, mijn Schepper en bron van alle wijsheid.

Curriculum vitae

Willem Jan Meerding werd geboren op 17 mei 1969 in Gouda. Hij doorliep het VWO-Atheneum van 1981 tot 1987 aan het Christelijk Lyceum en aan Scholengemeenschap De Driestar in Gouda. Hij studeerde vervolgens economie (doctoraalexamen in 1993) en rechten (propedeuse in 1989) aan de Erasmus Universiteit in Rotterdam. Na een kortstondige loopbaan als docent economie is hij vanaf 1994 verbonden aan de afdeling Maatschappelijke Gezondheidszorg van het Erasmus MC. Hij was als gezondheidseconoom betrokken bij vele nationale en internationale onderzoeksprojecten op het terrein van kosten van ziekten, kosten-effectiviteit in de gezondheidszorg, kwaliteit van leven, ongevallen, screening op baarmoederhalskanker, en tuberculose. Hij was in 1998 expert voor het WHO tuberculose-programma in Armenië en was lid van de NVVP-werkgroep cervix uteri (Richtlijn Cervixcytologie, 2002). Hij is getrouwd met Margreet Keus. Samen hebben zij twee zoons, Rinko (9 jaar) en Loek (8 jaar).