

TOWARDS A NEW APPROACH FOR ESTIMATING INDIRECT COSTS OF DISEASE

MARC A. KOOPMANSCHAP¹ and B. MARTIN VAN INEVELD²

¹Department of Public Health and Social Medicine, Erasmus University Rotterdam,
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands and

²Institute for Medical Technology Assessment, Erasmus University Rotterdam, The Netherlands

Abstract—Many researchers in the field of evaluation of health care doubt the usefulness of estimates of indirect costs of disease in setting priorities in health care. This paper attempts to meet part of the criticism on the concept of indirect costs, which are defined as the value of production lost to society due to disease. Thus far in cost of illness studies and cost-effectiveness analyses the potential indirect costs of disease were calculated. In the following a first step will be taken towards a new method for estimating indirect costs which are expected to be effectuated in reality: the friction cost method. This method explicitly takes into account short and long run processes in the economy which reduce the production losses substantially as compared with the potential losses.

According to this method production losses will be confined to the period needed to replace a sick worker: the so called friction period. The length of this period and the resulting indirect costs depend on the situation on the labour market.

Some preliminary results are presented for the indirect costs of the incidence of cardiovascular disease in the Netherlands for 1988, both for the friction costs and the potential costs. The proposed methodology for estimating indirect costs is promising, but needs further development. The consequences of illness in people without a paid job need to be incorporated in the analysis. Also the relation between internal labour reserves and costs of disease should be further investigated. Next to this, more refined labour market assumptions, allowing for diverging situations on different segments of the labour market are necessary.

Key words—economic evaluation, indirect costs, cost of illness

INTRODUCTION

In most cost of illness-studies and in some cost-effectiveness analyses the indirect costs of disease are considered and estimated. Not all researchers are convinced that indirect cost estimates contribute to the economic evaluation of health care programmes. In this paper a first step will be taken in developing an alternative method of estimating indirect costs, which meets an important part of the criticism. The conceptual elements of this method will be described in detail. Some preliminary results will be presented for the indirect costs of the incidence of cardiovascular disease in the Netherlands for 1988. A comparison will be made with the traditional method of estimating indirect costs based on the human capital approach.

THE CONCEPT OF INDIRECT COSTS

Indirect costs of disease can be defined in several ways, depending on the perspective and objective of analysis. In this paper the indirect costs of disease will be defined as the value of production lost to society due to illness, with respect to paid labour as well as unpaid labour. While defined this way, this paper deals only with paid labour.

Production losses can be the consequence of:

- temporary absence from work;
- disablement;
- premature mortality.

Basically, production losses could even be relevant without absence from work if sick people continue working at a suboptimal level. Information on this topic is unfortunately not available, so it won't be analyzed in this paper. The specification excludes the 'costs' of pain, suffering, discomfort etc. Next to the formidable problem of valuing these items in monetary terms, it seems more appropriate to consider these psychological factors as effect variables and to quantify the consequences in terms of reduced quality of life. This procedure will also avoid possible double counting.

As a consequence we will focus on the 'pure' economic consequences of disease. Furthermore it should be stressed that we don't seek to assign a value to human life *per se*, as in some cost-benefit analyses [1]. This definition of indirect costs is therefore compatible with social cost-effectiveness analyses and cost-utility analyses.

Almost all studies estimating indirect costs use the human capital approach. The way in which the human capital approach is usually being applied leads to an estimate of the value of *potentially* lost production (or the potentially lost earnings) as a consequence of disease. We use the word *potentially*

because in case of permanent disablement or premature mortality at a specific age the total productive value (or earnings) from that age until the age of retirement is counted as indirect costs.

This concept of indirect costs in the economic appraisal in health care is not uncontroversial [2-5]. The critical comments focus on four topics:

(1) Earnings may be a poor estimate for the amount of production forgone to society. A better, though not perfect, estimate used in the rest of this paper is labour costs.

(2) Health care programmes directed to working people will always generate more gains in production than programmes for mentally handicapped or elderly. This could lead to priority setting enlarging existing inequalities in health, which may conflict with policy objectives.

(3) Sometimes it is not clear if costs are direct or indirect and this gives rise to possible double counting of costs.

(4) The real production losses can be much smaller than the potential losses because sick people can often be replaced at little cost.

Notwithstanding this critique the fact that production losses due to illness (and production gains due to health care) influence the scarcity of resources and therefore the wealth of society is not disputed. In this respect there exists no difference in assessing direct and indirect costs; both should be incorporated in economic evaluations of health care programmes.

The first and fourth criticism don't overthrow the concept of indirect costs but merely stress the importance of using estimates of the production losses which are expected to be effectuated in reality.

Addressing the first point of critique, an example in which labour costs are a poor estimate for the value of production is unpaid labour in the household. Reliable estimates of production losses should therefore include the production involved in this type of labour.

With respect to the second critical comment, the gain in society's wealth will indeed be larger for programmes directed to economically active people. Explicit estimates of these gains will be more helpful in decision making than using implicit notions. Next to this, estimates of indirect costs should never be weighed in isolation, but always as a part of the political decision making process, in conjunction with other criteria like direct costs and equity. In addition, including indirect costs has implications for equity. If a health care service is actually implemented estimates of production gains can be useful in the discussion about how the program needs to be financed.

An example of the third comment: if a sick person is a day absent from work and needs to visit a physician one could approximate the value of the time spent on the visit by the value of production lost in this timespan. If one also calculates the production lost due to a day of work absence, the sum of direct

and indirect costs is clearly an overstatement. We think this problem can be avoided by continuous awareness of the possibility of simultaneous sickness absence and medical consumption.

The last item can be met when taking into account the economic processes which make that sick people can be replaced, mitigating the amount of production losses. Several examples can be given in which the expected production loss due to illness may be much smaller than the potential loss.

—If unemployment, registered as well as hidden unemployment, is well beyond the level of inevitable frictional unemployment, sick persons can be replaced by unemployed.

—If labour reserves exist within firms with sick employees the loss in production may be negligible.

—Non-urgent work may be cancelled or postponed without consequences for the level of production.

—If replacement is only feasible by persons already working, a reallocation of employees over jobs can induce a smaller production loss, resulting from a vacant job at a lower productivity.

The underlying notion is that processes within firms and on the labour market will lead to replacement of sick persons, after a period of adaptation, reducing the production loss substantially. Unemployment is important in this respect, but it is not the only factor causing the difference between potential and real production losses. The above mentioned processes are effective in the short run when the organisation of the production and the labour market conditions are merely fixed constraints.

In the longer run unemployment may fall substantially or enhanced efficiency within firms may reduce internal labour reserves, both inducing higher production losses. However, in the longer run labour shortages and production losses will be mitigated by increasing possibilities for restructuring the organization of the production, by reallocation of employees and as a result of laboursaving investments. To illustrate this process the analysis should preferably be linked to a macro-economic model incorporating vintages of capital goods in relation to the relative costs of capital and labour.

In the following the concept of indirect costs will be modified, setting a first step towards deriving more realistic estimates of production losses due to disease.

OPERATIONALIZATION

Disease may cause losses in production, but in general this loss will be confined to a period needed to adapt to the changed situation. This period will be called the friction period. To clarify this notion further, it may be helpful to consider the possible economic consequences of disease on the micro-economic level. Due to illness, there are three possible

outcomes for production and costs during the friction period.

- Production falls;
- Production remains equal at extra labour input and extra costs;
- Production falls in spite of extra labour input and costs.

The sum of production loss and possible extra costs can vary widely, depending on the way extra labour input is realized. Acquisition of labour from within the firm will often be cheaper compared with hiring external personnel. Exact data about the magnitude of losses and extra costs as well as the frequency of these possible outcomes are lacking. Therefore a macro-economic approximation of the indirect costs should tackle this problem. The labour costs of the sick employee during the friction period may be the best possible estimate of the *average* indirect costs. Additional costs for selecting and training personnel may be relevant if the sick person must be replaced permanently.

In the Netherlands the current level of registered and hidden unemployment is substantial. Recent labour market studies indicate that with growing employment many yet non-participating women acquire new jobs. Still the participation rate for women is low, so an important part of non-participating women can be considered as involuntarily unemployed [6,7]. Furthermore, Dutch studies indicate that 15–20% of the prevalence of persons receiving disablement benefits should actually be considered as unemployed [8,9]. Therefore we assume that in general replacement of sick persons by unemployed is possible after a friction period. The production loss during the friction period is assumed to be equal to the labour costs of the sick person. For an estimate of the total indirect costs of a disease two items are to be analyzed: the costs per friction period and the frequency of friction periods.

The costs per friction period are calculated for the Netherlands 1988. Because of insufficient data on the quantity of unpaid labour and its consequences for

work absence and disablement the analysis yet had to be restricted to paid labour.

The *average* value of production per labour-year has been approximated by combining the net domestic product, measuring the total net value added by all employed labour, with the number of labour years in 1988 [10]. This average production value was made age- and sex-specific using data about the relative gross labour incomes per labour-year by age and gender [11]. We assumed a stable and linear relationship between gross income and the value added per labour year.

Data on the relative amount of full time and part time labour by age and sex were extrapolated on the basis of a detailed study [7] and used to calculate the average production value *per working person* per year. Correction for the age- and sex-specific share of employed persons in the population yields the average production value *per person* per year by age and sex.

The value per working person is applied to data on work absence and disablement. The production value per person will be used to calculate the indirect costs of mortality. We assumed equal mortality rates for employed and not-employed.

The length of the friction period was derived from data on the average period necessary to fill vacancies: 2.5 months in 1988 (not age- and sex-specific) [12]. The production values per year are applied in calculating the potential indirect costs. The values per friction period lead up to the friction costs after combining them with the frequency of friction periods. We would prefer using estimates of the marginal productivity instead of average productivity. However, reliable data on marginal productivity are difficult to obtain, especially for the services sector.

Data on work absence, disablement and mortality provide information about the frequency of friction periods [13–15]. These databanks are not linked at the level of individuals which makes some assumptions necessary. For the Netherlands a distinction must be made between employees and self-employed, as a consequence of differences in registration. For

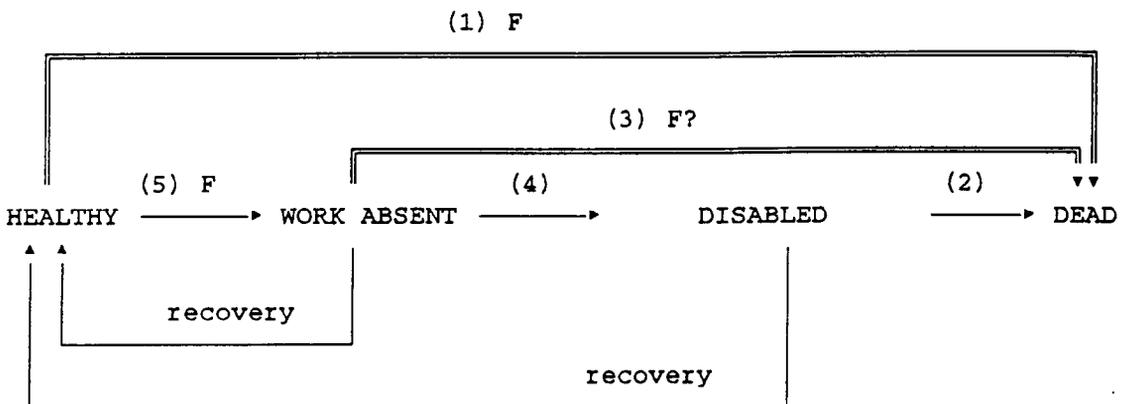


Fig. 1. Flows between healthy, absent of work, disabled or dead for employees, including the relevance of friction periods (F).

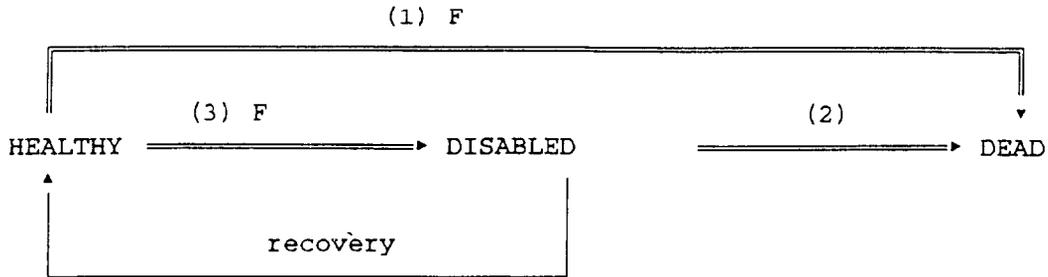


Fig. 2. Flows between healthy, disabled and dead for self-employed, including the relevance of friction periods (F).

employees the possible flows between healthy, absent from work, disabled and dead are depicted in Fig. 1.

In case of 'immediate death' (flow 1) and work absence (flow 5) a friction period is relevant. Premature death while one is absent from work (flow 3) can induce double counting of friction periods if the work absence is longer than the friction period. We can't correct for this, because the number of people dying during work absence and the relation with the length of absence are unknown. However the results in Table 1 show that the maximum level of double counting is still small.

Flows 2 and 4 don't cause friction costs because in the Dutch system one acquires the status of disabled only after a stay of one year in the work absence insurance. In this case the friction period took place in a previous year.

For self-employed people no work absence register exists, so the frequency of friction periods must be derived from data on immediate death (flow 1) and incidence of disablement (flow 3), see Fig. 2. Flow 1 was calculated by correcting total mortality for the number of disabled persons dying.

If internal labour reserves prevail, they can further reduce the frequency of friction periods and the level of production losses. A study of the Dutch industry for the period 1975-1982 estimated the level of general internal labour reserves as between 5 and 15% of the existing labour volume, concluding that 5% seems a minimum level of general internal labour reserve of firms [16]. Given the heterogeneity of labour the reserve will be lower for each individual employee, but how much lower is unknown. Because of this uncertainty no further corrections are made for internal labour reserves.

As far as the internal labour reserves exist for purposes of covering for sickness absence, these costs are clearly related to sickness. Internal labour reserves can also be useful in coping with fluctuations in the volume of production. In that case the costs are not caused by sickness.

These remarks illustrate the need for further investigation of the role of internal labour reserves, which has not been possible in this paper.

The assumptions for calculating the friction costs and the potential indirect costs can be summarized

as follows. In both cases the incidence of absence, disablement or mortality in 1988 is the viewpoint of analysis. In case of the friction cost method one friction period is counted per death until age 65, correcting for disabled persons dying. Incidence of disablement is only relevant for self-employed persons. Days of work absence are counted, but only within the friction period of two and a half months.

In estimating potential indirect costs production losses caused by mortality are counted until age 65. The average period of disablement has been estimated for each 10-year age group and sex and was applied to the incidence of disablement in 1988. All registered days of work absence are counted. Costs beyond 1988 were discounted by 5% yearly.

RESULTS

Table 1 shows the estimated indirect costs of the incidence of cardiovascular disease. The estimates are based on the production losses of paid labour only, due to insufficient data on unpaid labour. Therefore the indirect costs are underestimated, but this holds for both methods.

The results differ substantially depending on the method used, as could be expected. Total costs according to the human capital method are over 5 billion guilders, the friction cost method results in somewhat more than 400 million. The major part of this difference is due to the cost of disablement and mortality because of the great difference in the length of the period in which production losses are assumed

Table 1. Indirect costs of incidence of cardiovascular disease according to the friction cost method and the human capital approach (gross variant*) for the Netherlands in 1988 in millions of Dutch guilders. Costs beyond 1988 were discounted by 5% per year. (% of total costs in brackets)

	Friction (mln dfl)	Human capital (min dfl)
Mortality	60 (14%)	2072 (39%)
Disablement	19 (4%)	2615 (49%)
Work absence	353 (82%)	624 (12%)
Total costs	431 (100%)	5311 (100%)

*In the gross variant of the human capital approach the total average value of production is used. In the net variant the value of personal consumption of the sick person is deducted from the production value.

to prevail. The results for the costs of work absence differ less dramatically. The relative difference in costs of both methods corresponds with a Dutch study of the costs of traffic-accidents which used a variant of the friction cost-method [17].

Considering the types of cost per method, the importance of the costs of work absence diverges completely. More than 80% of the friction costs are caused by work absence compared with only 12% for the human capital method. The opposite holds for the shares of costs induced by mortality and disablement. The above mentioned possibility of double counting in case of people dying while absent from work is by definition lower than 60 mln Dfl, the friction costs caused by mortality, see Table 1.

DISCUSSION

Comparing both methods of estimating indirect costs the following remarks can be made. Both methods can provide only global results as they use aggregate data. In some studies estimates of production losses according to the human capital theory are corrected for the chance for an individual to become unemployed [18,19]. This indeed influences the individual income, but does not necessarily have an impact on society's level of production. So this won't necessarily lead to indirect costs according to our definition.

A critical remark on the friction cost method may be that its credibility depends on the existence of substantial unemployment. However, the friction cost method is sensitive to the level of unemployment. If unemployment declines, the average period to fill vacancies will become longer, inducing higher costs per friction period. The situation on the Dutch labor market has actually developed in this direction since 1986. Nevertheless in parts of the labour market with serious shortages, production losses will certainly be higher than estimated by the friction cost method.

The most important shortcoming of the results presented here is the lack of a reliable estimate of the production loss of unpaid labour, understating the production losses of women and elderly people to a large extent. This could lead to undesirable policies in favour of intensifying efforts to prevent and fight diseases from which predominantly men are suffering. This again shows the implications of calculating indirect costs for equity. Data on unpaid labour are very scarce. This also holds for the amount of work absence and disablement with respect to unpaid labour. Both are indispensable elements in estimating the indirect costs.

With respect to the internal reserves of labour within firms three qualitative comments can be made. First, it is plausible that their size will be influenced by the expected sickness absence, especially if the work can't be postponed and the sick employee must be replaced immediately (e.g. in case of bus drivers). This will diminish the amount of production lost by

disease, but at extra costs for maintaining this labour reserve. The costs of this reserve are not necessarily high, because firms may be able to hire employees for part-time jobs, who can be employed more hours if necessary [20].

Secondly, the size of internal reserves and the level of unemployment may be inversely related, because in times of high unemployment firms can more easily acquire new personnel to replace sick employees. On the other hand, in the Netherlands temporary employment agencies play an important role in providing a flexible supply of labour which can act as a buffer, reducing fluctuations in the size of internal reserves. Thirdly, internal labour reserves can be held partly in order to cope with fluctuations in the volume of production. In this case not all costs are directly related to sickness.

It can be concluded that the numerous aspects of the relation between costs of disease and internal labour reserves should be studied in more detail.

In estimating production losses one should be careful when considering only the registered sickness absence, because sick people can continue working at a suboptimal level. The extent of this phenomenon is very difficult to quantify, but especially in times of high unemployment, it can be more than negligible.

Further analysis will be necessary on the costs of selection and training if vacancies have to be filled permanently and the level of internal labour reserves for specific jobs.

Next a stratification of the labour market assumptions seems useful to incorporate differences between parts of the labour market which are relevant for the level of production losses. The labour market can be split into educational levels, each with different production values and length of the friction period. This will refine the estimates of indirect costs and takes labour shortages in some market segments into account. For such refinement data on incidence and mortality by education level should be available.

CONCLUSION

In setting priorities in health care reliable estimates of direct and indirect costs of disease are indispensable. Thus far studies only provide estimates of the potential indirect costs of disease.

The friction cost method can be considered as a promising method of estimating the indirect costs which may actually result from the incidence of disease. This approach explicitly takes into account short run and long run adaptation processes in the economy. The friction cost method is therefore highly compatible with the methodology of cost-effectiveness analysis in which the real resource costs (of a disease or technology) are subject of analysis.

Because the costs per friction period depend on labour market conditions the estimates of the indirect costs can vary widely over time and across countries. In situations with a tight labour market the differences

between the potential costs and the friction costs will become much smaller than shown in Table 1. The differences also depend on the type of disease. In case of diseases causing predominantly disablement and mortality the gap will be much wider than for diseases merely inducing work absence.

The proposed methodology needs further development. Especially the consequences of illness in the longer run have to be considered in more detail, preferably by linking the analysis to a macro-economic model. Presumably, these consequences are minor, especially when illness occurs in more flexible and adaptive sectors of the economy. Furthermore, our empirical analysis is global and preliminary. Further refinement of the underlying labour market assumptions is necessary; both with respect to distinguishing different segments on the labour market as to a better analysis of the amount of sickness-related costs caused by internal labour reserves. Extension of the analysis to the consequences of illness in people without a paid job is needed.

Acknowledgements—This study was financed by the Department of Welfare, Health and Cultural Affairs, the Sickness Fund Council and Merck & Co Inc. We thank Frans Rutten, Gouke Bonsel and Leona van Roijen for their comments on previous versions of this article.

REFERENCES

- Schoenbaum S. C., Hyde J. N. Jr, Bartoshesky L. and Crampton K. Benefit-cost analysis of rubella vaccination policy. *N. Engl. J. Med.* **294**, 306–310, 1976.
- Drummond M. F., Ludbrook A., Lowson K. and Steele A. *Studies in Economic Appraisal in Health Care*, Vol. 2. Oxford University Press, New York, 1986.
- Drummond M. F., Stoddard G. L. and Torrance G. W. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press, Oxford, 1987.
- Hartunian N., Smart N. and Thompson M. *The Incidence and Economic Costs of Major Health Impairments*. Lexington Books, Lexington, 1981.
- McGuire A., Henderson J. and Mooney G. *The Economics of Health Care, an Introductory Text*. Routledge & Kegan Paul, London, 1988.
- Labour market report 1989 (in Dutch). Department of Social Affairs and Employment. 's Gravenhage, 1989.
- Teulings C. N., Koopmans C. C., Koopmanschap M. A. and Vedder L. *Redistribution of Employment and Social Insurance*. Foundation for Economic Research, University of Amsterdam, in charge of the Department of Social Affairs and Employment (in Dutch). 's-Gravenhage, 1987.
- Roodenburg H. J. and Wong W. *The Labour Market Component in the Disablement Benefit* (in Dutch). Central Office for Planning. 's Gravenhage 1985.
- Vrooman J. C. and Kemp A. A. de. Disablement and labour market (in Dutch). *Econ. Stat. News* **3778**, 942–943, 1990.
- CPB (Central Office for Planning). *Central Economic Plan 1989* (in Dutch). 's-Gravenhage, 1989.
- CBS (Central Office for Statistics). *Social Economic Monthstatistics 1988* (in Dutch). 's-Gravenhage, 1988.
- CBS (Central Office for Statistics). *Social Economic Monthstatistics 1989* (in Dutch). 's-Gravenhage, 1989.
- CBS (Central Office for Statistics). *Diagnose Statistics (on work absence) 1987* (in Dutch). 's Gravenhage 1988.
- CBS (Central Office for Statistics). *Deaths by Cause of Death by Age and Sex 1988*. Health Monthstatistics. 's Gravenhage 1989.
- GMD, Mutual medical service. *Statistical Information 1988* (in Dutch), Amsterdam 1988.
- Koning J. de Volume and causes of labour hoarding (in Dutch). Academic Dissertation. Erasmus University Rotterdam 1987.
- Haselen H. W. J. van Production losses due to traffic accidents (in Dutch). Netherlands Economic Institute Rotterdam, 1987.
- Muller C. and Caton C. Economic cost of schizophrenia: a postdischarge study. *Med. Care* **21**, 92–104, 1983.
- Berk A., Paringer L. and Mushkin S. The economic cost of illness, Fiscal 1975. *Med. Care* **16**, 785–790, 1978.
- Reich M., Gordon D. and Edwards R. A theory of labor market segmentation. *Am. Econ. Rev.* **63**, 359–365, 1973.