

Health expectancy: an indicator for change?

Jan J Barendregt, Luc Bonneux, Paul J Van der Maas for the
Technology Assessment Methods Project Team

Abstract

Study objective – Health expectancy is an increasingly used indicator of population health status. It collapses both mortality and morbidity into a single indicator, and is therefore preferred to the total life expectancy index for populations with low mortality but high morbidity rates. Three methods of calculation exist: the Sullivan, double decrement, and multi-state methods. This report aims to describe their relative advantages and limitations when used to monitor changes in population health status over time.

Design – The differences between the three methods are explained. Using a dynamic model of heart disease, the effect of the introduction of thrombolytic treatment on the survival of patients with acute myocardial infarction is calculated. The resulting changes in health expectancy are calculated according to the Sullivan and multi-state methods.

Main results – As opposed to the double decrement and the multi-state methods, the Sullivan method produces spurious trends in health expectancy in response to the change in survival.

Conclusions – Estimates of health expectancy in a dynamic situation can be very misleading when based on the Sullivan method, with its attractively moderate data requirements. The multi-state method, which requires longitudinal studies of population health status, is often indispensable.

(*J Epidemiol Community Health* 1994;48:482-487)

In an article published in the December 1992 issue of this Journal Margaret Bone evaluated the international efforts to measure the health expectancy as an index of a population's state of health, and its application as an indicator of changes in population health.¹ For several years now an international network of researchers has been working to clear up methodological issues, to standardise methods, to encourage the collection of appropriate data, and to gain acceptance for the health expectancy indicator in health policy.²

The health expectancy (or healthy life expectancy) is derived from both mortality and morbidity, and indicates which part of the total life expectancy is spent in good health. Time trends in health expectancy help to determine whether we are improving the nation's health, or are just being more successful in preventing severely ill people from dying. This property makes health expectancy an important index in the ongoing debate, sparked off by Fries, on the compression or expansion of morbidity.³

There is no doubt that the success of the health expectancy indicator is largely due to its intuitive appeal. It seems a straightforward extension of the notion of life expectancy and is generally interpreted as the average number of years a newborn will live without (serious) disease. It also has attractive flexibility, depending on the definition of "healthy": there have been disease free, disability free, and quality adjusted life expectancy indicators.² As with life expectancy itself, however, there is more to the health expectancy indicator than meets the eye. For one thing, there are three different methods of calculating health expectancy: the Sullivan, the double decrement, and multi-state methods.⁴ Each has different data requirements and produces different results. Most of the studies cited above use Sullivan's method, named after the researcher that pioneered its use.⁵ The popularity of this method has a good reason: it is the least demanding in terms of data requirements. There is a price, however: the Sullivan method gives reliable results only in a static environment.

This article considers the conditions under which simple health expectancy indicators based on Sullivan's method can be used, and those which require the more demanding multi-state method. The introduction of thrombolytic treatment in hospitals and its effect on the health expectancy is used by way of illustration.

Methods

The health expectancy indicator is known in various guises, which differ in their definition of health and how it is measured, but otherwise are very similar. We will confine ourselves to the disease free life expectancy definition, more particularly to life expectancy free from heart disease, but the argument can be generalised to apply to the whole range of health expectancy indicators.

The health expectancy indicator is clearly an offspring of the standard life expectancy indicator, and employs the same method: a life table. In a standard life table, a birth cohort of usually 100 000 people is subjected to mortality probabilities in relation to age. For each age, the total number of years the shrinking cohort has yet to live is calculated, and dividing this total by the number of people still alive gives the life expectancy for that particular age. The Appendix (A) contains a more formal description of this calculation.

THREE WAYS OF DETERMINING THE HEALTH EXPECTANCY

Three variations of this procedure may be used to calculate the health expectancy:

(1) *Sullivan's method* uses disease prevalence data in relation to age to subtract the number of

Department of
Public Health,
Erasmus University,
Rotterdam,
PO Box 1728,
3000 DR Rotterdam,
The Netherlands
J J Barendregt
L Bonneux
P J Van der Maas

Correspondence to:
Mr J Barendregt.

Accepted for publication
January 1994

years the cohort still has to live with disease from the total number of years. The health expectancy is then calculated by dividing this number of healthy years by the number of people alive. As a rule, the disease prevalence data are from a cross sectional survey;

(2) *The double decrement method* uses disease incidence data: the birth cohort is subjected to both mortality and incidence probabilities, the former corrected for disease specific mortality. From the cohort of people who are neither dead nor ill the life expectancy is calculated according to the standard procedure;

(3) *The multi-state method* also uses incidence probabilities to calculate disease prevalence, but in addition allows for one or more disease states including, when applicable, a 'cured' state where the cohort may be subject to recurrent disease. It depends on the definition of "healthy" which of these states will count towards the health expectancy, but once this is decided the calculation proceeds in the standard way. In Appendix B an equation is given for health expectancy according to the Sullivan method and the multi-state method with only one disease state, and no cure or recurrent disease.

It is evident that the multi-state method is by far the most demanding in terms of data requirements. In fact only longitudinal studies with a long follow up period and a sufficient number of rounds can provide the necessary detail. Still, it is very attractive because it can capture the natural course of a disease, and can encompass patients who are cured or have intermittent disease free periods.

The prevalence data used in Sullivan's method reflect implicitly this natural course too, but in a very complex way. Prevalence is a stock variable: current prevalence of spinal injuries among 40 year old (former) car drivers covers an accumulated 22 years of car accidents that happened to drivers of as many different ages. As is shown in Appendix C current cross sectional prevalence is a function of a long series of past incidence and mortality rates (and cure rates, when applicable).

Incidence and mortality are flow variables: current acute mortality from spinal injuries among 40 year olds reflects this year's car accidents to drivers of that age only. While the multi-state method uses flow variables only, the Sullivan method employs the flow variable "mortality" to calculate the total number of years lived, and the stock variable "prevalence" for the diseased number of years. The inconsistent mix of stock and flow variables may lead to odd results.

The double decrement life table, like the multi-state one, uses flow variables only, is therefore consistent, and has much less demanding data requirements. The drawback is that, in effect, it treats disease incidence the same as mortality.

COMPARING SULLIVAN AND MULTISTATE METHODS
To illustrate our argument we shall compare results from the Sullivan method with those from the multi-state method. In order to ensure comparability with the Sullivan method we use

the same incidence data and combine these with survival data to calculate a prevalence. In both cases, the possibility of cure is ignored – that is, all patients who have had a myocardial infarction will contribute to the prevalence of heart disease for the rest of their lives.

The model

We use a dynamic population model for ischaemic heart disease to compare the two methods. It is a state-transition model, that owes much to the coronary heart disease policy model of Weinstein *et al.*⁶ After patients enter the disease model with a first manifestation of heart disease – angina pectoris, an acute coronary event, or heart failure – they are subject to risks of (possibly repeated) events such as an operation or another acute coronary event. Depending on their state when an event occurs, they may be referred to a new state, they may die, or they may remain where they are.

Input includes the incidence of myocardial infarctions in relation to age, based on the nationwide Dutch hospital register.⁷ Survival after admission to hospital with myocardial infarction is based on the same register and, for the long term, on the results of the Framingham Study. Combined with an "all other causes" mortality we were able to reproduce very well the observed mortality from myocardial infarction in The Netherlands.⁸

The model is dynamic in the sense that the prevalence in each state depends on the prevalence in the previous time period, and on inflow and outflow variables. These inflow and outflow variables can be manipulated to simulate changes in incidence and survival over time. Output options include disease free life expectancies according to the Sullivan method (using the dynamically calculated prevalence) and the multi-state method (using a synthetic prevalence based only on current incidence and survival).

Effect of thrombolysis

For our calculations we use the hypothetical example of the simultaneous introduction of thrombolytic treatment in all Dutch hospitals. The assumption is that all hospitals introduced thrombolysis in 1991, and that none used it before then. We assumed a conservative, 25% reduction in acute, in-hospital deaths from myocardial infarction after thrombolysis, and calculated long term effects on (healthy) life expectancy keeping incidence and all other survival parameters constant.⁹

Results

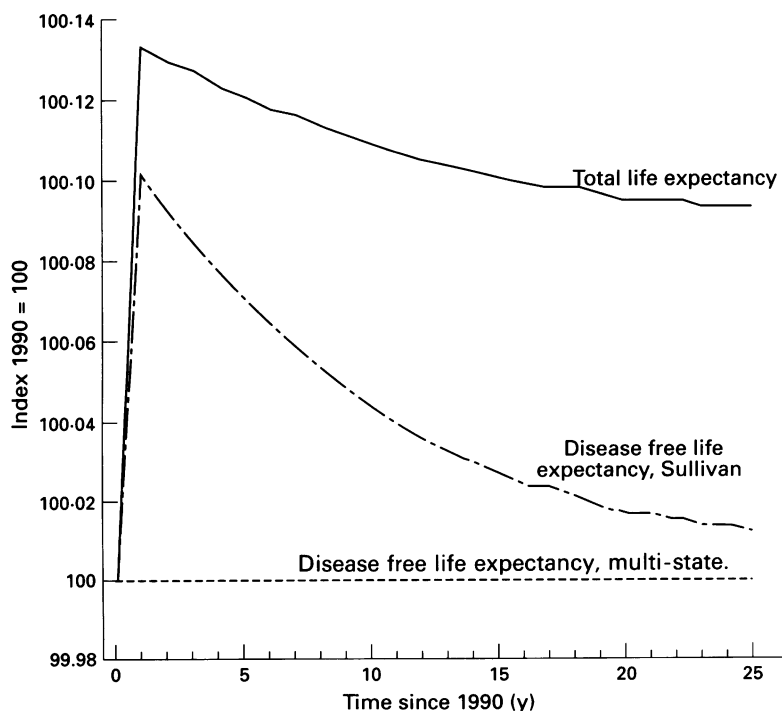
What can we expect from this intervention? Since thrombolysis has an effect on acute deaths only, we expect the total life expectancy to increase suddenly between 1990 and 1991. Since the incidence of myocardial infarction is kept constant, and because all people are considered "heart patients" after a myocardial infarct, the life expectancy free from heart disease should remain unchanged.

From equations (3) and (5) we can see how a change in disease specific mortality will, through the concomitant change in total mortality, instantaneously affect the total number of years lived by the synthetic cohort, but will only gradually be reflected in population prevalence. This results in an abrupt change in the Sullivan health expectancy (up when mortality goes down, and vice versa), followed by a slow approach to the correct value. The synthetic prevalence of the multi-state method (equation (7)) is, however, like the total number of years lived, adjusted instantly.

In the figure we present the results from a 25 year simulation with the model. Total life expectancy increases suddenly, but tapers off somewhat afterwards as the now surviving patients drift into chronic heart disease and its high mortality. The disease free life expectancy according to the multi-state method remains unchanged, as expected. The health expectancy according to the Sullivan method, however, shows a quite different pattern: it rises initially almost as much as total life expectancy, then starts falling and reaches the multi-state line asymptotically after about 40 years.

Here the Sullivan method shows the disadvantage of mixing stocks and flows. The constant incidence in combination with the lower mortality will, in the end, necessarily produce a higher prevalence. But the prevalence at any age is a function of incidences and mortalities at all lower ages. After a change in incidence or mortality, the prevalence in a population will have reached its new equilibrium value only when all cohorts dating from before the change are extinct.

While the prevalence, and consequently the number of years lived with disease according to



Total and disease free life expectancy after adoption of thrombolytic treatment in men. Multi-state method versus Sullivan's method.

the Sullivan method, is still catching up, the lower mortality has already raised the total number of years lived by the life table cohort. Since the health expectancy is calculated using the difference of total and disease years lived, the fast adjusting total life expectancy initially pulls up the slowly adjusting Sullivan health expectancy.

The Sullivan indicator shows an increase followed by a decrease when it ought to remain unchanged. Only when the prevalences have settled down to the new equilibrium does the Sullivan disease free life expectancy indicator produce no spurious trends. In Appendix (D) we derive how many years it takes for the Sullivan health expectancy to reach this equilibrium: this is determined by the difference between the highest age in the life table considered (typically something like 95 years or so) and the lowest age at which the disease under consideration becomes important (for chronic diseases typically somewhere between 20 and 60). It turns out therefore that incidence and mortality (and cure, if included) must have been constant for a period from 35 to 75 years, depending on the disease.

Discussion

The results show that the health expectancy according to Sullivan's method can produce misleading results when dynamic effects are present. It is true that our example is geared to show this: a disease free life expectancy indicator of a disease that affects people rather early in life, yet has a relatively good survival, will be much more "off track" than, for instance, an indicator of life expectancy without terminal disease.

We can see from equations (4) and (5) in the appendix which disease characteristics will reduce the deviation from the correct value. The less current prevalence is influenced by past prevalence, the sooner the Sullivan health expectancy will have approached the right value. This occurs, for example, when mortality is high or incidence increases sharply with age.

Current heart disease prevalence is dependent on past prevalence, and heart disease is therefore a good example to provide the clarification the issue apparently still needs. Bone attributes the call for the use of longitudinal data to the need to take into account explicitly reversals in disability.¹ This is an important point but, as shown above, not the whole story. Robine and Ritchie are aware of the problem of interpreting a time series of health expectancies according to Sullivan's method, but they contribute this to the combination of period data (current mortality probabilities) with cohort data, which they say currently observed disability prevalence is.¹⁰ But observed disability (or disease) prevalence is just as much period data as observed mortality probabilities: both are influenced by the history of the cohorts that make up the current population, prevalence only more so. Therefore their suggested remedy, to derive disability free life expectancy from period data only, will not work. The problem originates from combining stock and flow, not period and cohort data.

In a more recent article Robine *et al* point out that health expectancies based on a combination of stock and flow variables pose problems when making comparisons between countries.¹¹ This is undoubtedly true, and the health expectancy is meant to estimate the impact of *current* health risks, not so much the legacy of the past. Although the authors discuss the development of health expectancy over time, however, no mention is made of the fact that similar stock and flow problems occur when estimating trends.

It could be argued that the example of the sudden introduction of thrombolytic treatment is a theoretical and disruptive event, while things do not normally show such dramatic dynamics. But while its introduction in clinical practice did not happen overnight, an English study shows that thrombolysis took about three years to go from practically zero to a new treatment plateau – long perhaps from a clinical point of view, but short compared with the kind of time lags associated with the Sullivan method (see Appendix (D)).¹²

In addition, although total mortality develops rather smoothly over time, disease specific mortality rates show diverse and much more dynamic patterns.¹³ For example, male mortality from ischaemic heart disease in The Netherlands, standardised for population structure and indexed at 1950 = 100, reached a high of 301 in 1972, and had fallen to 191 in 1990 (with some allowance for codification changes over the period).¹⁴ Presumably the underlying morbidity from ischaemic heart disease will not have stayed unchanged either. Given the very long time lags involved, such changes will surely bias Sullivan health expectancy trend estimates.

Most health expectancy estimates published so far are disability free life expectancies, or some comparable health measure that is the compound result of a large number of diseases.² Interpreting a time series of Sullivan health expectancies based on such a compound health measure becomes a daunting task indeed. The health expectancy thus measured is, at any point in time, a function of a large number of disease specific prevalences, each being at some unknown point on the way to a new (and probably shifting) equilibrium. To disentangle this jumble and distinguish true change from delayed adjustments seems impossible.

Recent studies report that health expectancy is increasing less than total life expectancy, or not even increasing at all.^{15,16} But the decline in mortality rates in middle and old age started somewhere in the late 60s or early 70s, and given the response of the Sullivan health expectancy trend estimates to mortality declines – an initial overestimate of health expectancy followed by a long term decline towards the correct value – these estimates will still display a downward trend as a reaction to previous declines in mortality. In combination with current declines in mortality, the observed trend in Sullivan health expectancies may go either way, but provides no useful information. The observed stagnation in health expectancy that has accompanied recent increases in total life

expectancy may, therefore, very well be an artefact of the Sullivan method.

The double decrement method provides no viable alternative. Although it is methodologically sound, it is far too crude for monitoring population health status by equating disease incidence with death. This means that it is insensitive to changes in disease prevalence or severity that originate from new or better treatments that patients may benefit from or even be cured by.

This leaves the multi-state method. It is consistent because it calculates disease prevalence using only current flow variables, which eliminates the problem of an inherited stock of patients that plagues Sullivan's method. It is potentially subtle enough to track the changes in health status after, for instance, the introduction of thrombolysis, that will not just prevent a number of deaths, but will also limit the damage to the heart for a number of patients who would have lived anyway. To use the multi-state method to its full potential, however, detailed longitudinal data are needed, also on the disease specific level. The kind of model we have employed in this study also gives a better understanding of the dynamics of population health status, and it has similar data requirements. Only when we can monitor population health status *and* explain the direction it is moving in, can we understand what we are monitoring and make informed suggestions about improvement.

Sullivan's method was intended to give an estimate of the health expectancy, using readily available or easily obtained data. In a population with an increasing total life expectancy it will be biased upward, but this may be a disadvantage well worth putting up with. The method should not be pushed beyond this goal, however, because what is a small bias in health expectancy will confound a trend analysis of a time series of health expectancies since the direction and size of the bias are time dependent. So the question "Is health expectancy a valuable indicator for changes in population health?" deserves a conditional "Yes" – the condition being longitudinal studies, that can provide the input for multi-state tables and dynamic models.

The short cut Sullivan's method provides, as compared to the multi-state method, is a dead end when it comes to the analysis of changes in health expectancy over time.

Appendix

(A) LIFE EXPECTANCY

Life expectancy is calculated by submitting a so called "synthetic" birth cohort of 100 000 people to age specific mortality probabilities. The birth cohort is called "synthetic" because, and this is important to note, the mortality probabilities used are the current probabilities. A real birth cohort has been submitted throughout life to mortality probabilities in the past that were, as a rule, higher than the current ones, and a cohort of newly borns will face future probabilities that we expect to be lower yet. Because of these changes in mortality, the current life expectancy at age 0 will not be equal

to the average number of years to be lived by a cohort of newborns.

If we denote the synthetic cohort with l_a , with a the age index and a_{\max} the highest age considered, and the mortality probabilities with q_a , then the life expectancy e_a is given by the total number of years to be lived by the cohort at age a divided by the number alive at that age. Formally:

$$e_a = \frac{\sum_a^{a_{\max}-1} 0.5(l_a + l_{a+1}) + l_{a_{\max}} e_{a_{\max}}}{l_a} \quad (1)$$

Using $l_{a+1} = l_a(1 - q_a)$ we can rewrite (1) to:

$$e_a = \frac{\sum_a^{a_{\max}-1} l_a(1 - 0.5q_a) + l_{a_{\max}} e_{a_{\max}}}{l_a} \quad (2)$$

Life expectancy is a static estimator: because $l_{a+1} = l_a(1 - q_a)$ for all a any change in the current mortality q_a is instantly reflected in the total number of years lived to its full effect. In fact, the life expectancy estimator assumes that the mortality probabilities have been unchanged for at least a_{\max} years.

(B) HEALTH EXPECTANCY

Health expectancy is calculated by not letting years defined as unhealthy count towards the total number of years to be lived. When P_a is the age specific prevalence of unhealthiness (expressed as a proportion) then the health expectancy h_a can be calculated by:

$$h_a = \frac{\sum_a^{a_{\max}-1} l_a(1 - 0.5q_a)(1 - P_a) + l_{a_{\max}} e_{a_{\max}}(1 - P_{a_{\max}})}{l_a} \quad (3)$$

Equation (3) can be used for both the Sullivan and the multi-state methods, the difference is where the prevalence P_a comes from.

(C) EMPIRICAL VERSUS SYNTHETIC PREVALENCE

The Sullivan method uses empirical prevalence data obtained through a survey or some similar method, while the multi-state method uses the prevalence calculated in the life table itself from the current incidence and mortality probabilities. The cross sectional prevalence for a very simple disease process with incidence I_a and mortality M_a and no cure is given by:

$$P_a^t = (P_{a-1}^{t-1} + I_{a-1}^{t-1})(1 - M_{a-1}^{t-1}) \quad (4)$$

Applying (4) recursively to itself, and assuming that prevalence at birth equals zero, we can rewrite (4) as follows:

$$P_a^t = \sum_{j=0}^{a-1} (I_j^{t-a+j} \prod_{k=j}^{a-1} (1 - M_k^{t-a+k})) \quad (5)$$

The prevalence as given by (5) is used by the Sullivan method, for the multi-state method the following equations apply:

$$P_a^t = (P_{a-1}^{t-1} + I_{a-1}^{t-1})(1 - M_{a-1}^{t-1}) \quad (6)$$

$$P_a^t = \sum_{j=0}^{a-1} (I_j^{t-a+j} \prod_{k=j}^{a-1} (1 - M_k^{t-a+k})) \quad (7)$$

This multi-state disease prevalence might be called, as an analogue to the life table cohort, a "synthetic prevalence". This synthetic prevalence is, unlike a cross sectional prevalence, not a stock variable, because, being a function of current incidence and mortality only, it does not depend on past values.

(D) TIME LAGS

From (5) and (7) it can be deduced under which conditions the Sullivan and multi-state disease prevalences at age a and time t will be equal:

$$I_j^{t-a+j} = I_j^t \forall j \in [0..a-1]$$

$$M_j^{t-a+j} = M_j^t \forall j \in [0..a-1]$$

To put it differently, the past age specific incidence and mortality must have been equal to the current incidence and mortality for, depending on age, up to a years. When a_{\min} stands for the lowest age at which incidence occurs this requirement relaxes to $a - a_{\min}$ years. Because the calculation of the health expectancy always uses the prevalence of the highest age a_{\max} (see equation 3), this implies that the Sullivan health expectancy equals the multi-state health expectancy when age specific incidence and mortality have been constant for up to $a_{\max} - a_{\min}$ years. The same applies for cure probabilities, if included.

This research was funded by the Ministry of Welfare, Health and Cultural Affairs, The Netherlands.

Members of the Technology Assessment Methods project team are Barendregt JJ, van Beeck E, Boer R, Bonneux L, Bonsel GJ, van den Bosch JMH, Gunning-Schepers L, Habbema JDF, van Hout BA, van Ineveld BM, Koopmanschap MA, Looman CWN, Lubbe J, van der Maas PJ, Mackenbach JP, Oortmarssen GJ, van Rooijen L, Rutten FFH.

- 1 Bone MR. International efforts to measure health expectancy. *J Epidemiol Community Health* 1992;46:555.
- 2 Mathers CD, Robine JM. *Health expectancy indicators: a review of the work of REVES to date*. 6th Work-group meeting REVES, International Research Network for Interpretation of Observed Values of Healthy Life Expectancy. Montpellier: REVES, 1992:116.
- 3 Fries JF. Aging, natural death, and the compression of morbidity. *N Engl J Med* 1980;303:130.
- 4 Robine JM, Romieu I, Mathers CD. *Multiplication of health expectancy calculations and international comparison problems*. WHO/CBS Consultation to develop common methods and instruments for health interview surveys: third consultation. Amsterdam: CBS 1992.
- 5 Sullivan D. A single index of mortality and morbidity. *Health Services and Mental Health Administration (HSMHA) Health Reports* 1971;86:347.
- 6 Weinstein MC, Coxson PG, Williams LW, Pass TM, Stason WB, Goldman L. Forecasting coronary heart disease incidence, mortality, and cost: The coronary heart disease policy model. *Am J Public Health* 1987;77:1417.
- 7 LMR (Information Center for Health Care). Files from the National Medical Registration concerning admissions and intervention in 1988. Utrecht: LMR, 1990.
- 8 Bonneux L, Barendregt JJ, Meeter K, Bonsel GL, van der Maas PJ. Estimating clinical morbidity due to ischemic heart disease and congestive heart failure. *Am J Public Health* 1994;84:20.

- 9 Gruppo Italiano per lo Studio della Streptochinasi Nell Infarto Miocardico (GISSI). Long-term effects of intravenous thrombolysis in acute myocardial infarction: final report of the GISSI Study. *Lancet* 1987;1:871.
- 10 Robine JM, Ritchie K. Healthy life expectancy: evaluation of global indicator of change in population health. *BMJ* 1991;302:457.
- 11 Robine JM, Michel JP, Branch LG. Measurement and utilization of healthy life expectancy: conceptual issues. *Bull World Health Organ* 1992;70(6):791.
- 12 Ketley D, Woods KL. Impact of clinical trials on clinical practice: example of thrombolysis for acute myocardial infarction. *Lancet* 1993;342:891.
- 13 Mackenbach JP. *Mortality and medical care*. Rotterdam: Erasmus University Rotterdam, 1988. PhD Thesis.
- 14 Centraal Bureau voor de Statistiek. Overledenen naar doods-oorzaak, leeftijd en geslacht, serie B1. Voorburg, published yearly.
- 15 Crimmins E, Sato Y, Ingegneri D. Changes in life expectancy and disability-free life expectancy in the United States. *Population Development Reviews* 1989;15:235.
- 16 Mathers CD. *Disability-free and handicap-free life expectancy in Australia*. Health differentials series no 1. Canberra: Australian Institute of Health, 1990.