Disability weights for diseases

A modified protocol and results for a Western European region

MARLIES E.A. STOUTHARD, MARIE-LOUISE ESSINK-BOT, GOUKE J. BONSEL,
on behalf of the DUTCH DISABILITY WEIGHTS (DDW) GROUP *

Background: The objective of the study was to establish a comprehensive and consistent set of disability weights for a number of important diseases in a Western European (Dutch) context, to be applied in composite health outcome measures to quantify the burden of disease and in economic evaluation of health care services. The context of the study was the Dutch Public Health Status and Forecast study. Replication and refinement of the valuation protocol used in the Global Burden of Disease (GBD) Study was a secondary aim. Methods: The disease stages were valued in a panel study in two steps, enabling the evaluation of a large number of disease stages. The first step was a carefully designed group process, using person trade-off as the valuation method to establish disability weights for 16 selected disease stages. The second step consisted of interpolation of the remaining disease stages on a disability scale by the individual panel members. Panel members were Dutch health care professionals with sufficient knowledge of the consequences of a broad variety of diseases. Results: A comprehensive set of disease-specific disability weights for 175 disease stages associated with 52 disease categories (cf. ICD-9) was obtained. The internal consistency and validity of the set of Dutch disability weights were satisfactory. Considerable agreement existed within panels, between the panel members and panels. Conclusions: Establishing a comprehensive and coherent set of reliable disability weights, using a modified valuation protocol from the GBD Study appeared to be feasible. The results can be used in composite health outcome measures applied in public health research and in economic evaluations.

Keywords: burden of disease, composite health outcome measures, DALY, person trade-off, public health, valuation of health status

Whether it is to quantify the burden of disease or to estimate the potential health benefits of alternative investments in health care services, mortality-based measures of a population's health are no longer considered adequate. Cost-effectiveness analyses have introduced the concept of QALYs (quality-adjusted life years) instead. The Global Burden of Disease (GBD) Study has shown the potential value of this type of composite health outcome measures in the area of public health research by employing the DALY (disability-adjusted life years) measure.[1–3] The DALY measure combines the (estimated) number of life years lost due to premature death and the number of years lived with disability using a set of disease-specific empirical weights to value the level of disability, following standardized methods.[4] Disability weights represent the consequences of relative severity of each disease. The GBD Study provided quantitative, internally consistent estimates of the burden of disease, including non-fatal health outcomes, attributable to 107 causes, per sex, for different age groups and per region in the world, for 1990.

Although hotly debated,[5] the DALYs have been a step forward in quantifying health changes in the public health field. However, certain questions remained to be answered. In the Dutch disability weights study, we addressed two of the remaining issues. The first issue is the application of the methods and the comparability of the results in a national rather than a global context. The disability weights used in the GBD project were global weights and apparently too universal at the disease level. The primary aim of the current study was to establish a comprehensive and consistent set of disability weights for a number of important diseases in a Western European (Dutch) context, to be applied in composite health outcome measures to quantify the burden of disease and in economic evaluation of health care services. We incorp-

* M.E.A. Stouthard[1], M-L. Essink-Bot[2], G.J. Bonsel[3]
1 Institute of Social Medicine, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands
2 Department of Public Health, Erasmus University Rotterdam, Rotterdam, The Netherlands
3 Department of Clinical Epidemiology and Biostatistics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands
Correspondence: Dr. M.E.A. Stouthard, Institute of Social Medicine, Academic Medical Center, University of Amsterdam, P.O. Box 22700, 1100 DD Amsterdam, The Netherlands, tel. +31 20 5664618, fax +31 20 6972316, e-mail: m.e.stouthard@amc.uva.nl

DDW-Group: Jan J. Barendregt, Gouke J. Bonsel, Marie-Louise Essink-Bot, Louise J. Gunning-Schepers, Paul F.M. Krabbe, Pieter G.N. Kramers, Paul J. van der Maas, Marlies E.A. Stouthard, Harry P.A. van de Water

orated Murray's[4] approach in the recent public health status and forecast document in The Netherlands.[6] The secondary aim deals with more methodological questions concerning the feasibility of the valuation protocol, and the improvement and refinement of the GBD valuation protocol with respect to the disease description.

METHODS

The list of diseases

The GBD Study had a global scope and, hence, contained a number of diseases bearing little relevance to the population's health in developed countries. For the Dutch study a list of 52 disease categories (at the three-digit level ICD-9 classification) was made, representing the major disease categories and health-related problems in The Netherlands.[6]

The disease categories were further subdivided into assumed homogeneous diseases stages with respect to functional status, treatment and prognosis, following an iterative process using empirical data of health status measurement, clinical expert opinions and the researchers' own judgement. For instance, the disease category 'dementia' was subdivided into the disease stages 'mild dementia', 'moderate dementia' and 'severe dementia'. In all, 175 disease stages were identified for the 52 disease categories.

The duration of a disease stage to be valued was universally defined as 1 year. For some disease stages characterized by a short episode of 1 or a few weeks of illness followed in almost all instances by complete recovery (e.g. influenza), the states were described and valued as 'a short episode in an otherwise healthy year'. In analogue, episodic diseases (e.g. asthma) were described as chronic, i.e. the general consequences for every day life were valued. Hence, for all disease stages the 1 year time frame calculus was achieved.

Standardized description of health status

A standardized description of the associated functional health status was added to each disease stage description, to harmonize the mental image of that state across individuals. An extended, six-dimensional version of the original EuroQol-system,[7] labelled EQ5D+C, was used. The original five dimensions of EuroQol ('mobility', 'self-care', 'usual activities', 'pain/discomfort', and 'mood') were, for descriptive purposes in the present study, extended with a sixth dimension on 'cognitive functioning'. This dimension was experienced to be indispensable for the description of particularly neuropsychiatric conditions. Each dimension had three levels of the general form 1 = no problems, 2 = some problems and 3 = severe problems.

Modifications from the Global Burden of Disease Study

The valuation protocol from the GBD Study[4] has been modified in two ways with respect to the disease description. In the GBD Study, the presumed sequelae or end stages of disease categories were primarily valued. In the current study, diseases categories were subdivided into homogeneous disease stages and all stages were valued separately. Furthermore, whereas in GBD Study a naturalistic description of the disease (sequelae) was presented in the form of a short vignette, in the current study we presented a disease category label with a description of all assumed stages of the disease (one of them as a short vignette) and the associated generic health status description of the disease stage to be valued in terms of EQ5D+C (*figure 1*). More examples of disease stage descriptions are given in *table 1*.

VISION DISORDER

Vision disorder is subdivided in the following three stages:
(1) mild vision disorder,
(2) moderate vision disorder,
(3) severe vision disorder.

We now ask you to value the following health state:
Patients with severe vision disorder, i.e. unable to read small newspaper print, great difficulty or unable to recognize faces at 4m distance

Description
- No problems in walking about
- Some problems in washing or dressing self
- Some problems with performing usual activities
  (eg. work, study, housework, family or leisure activities)
- No pain or discomfort
- Moderately anxious or depressed
- No cognitive impairment
  (eg memory, concentration, desorganisation, IQ-level)

Figure 1 Example of disease stage description

The valuation procedure

The valuation procedure took place in two steps. In the first step, a selection of 16 'indicator disease stages' was valued by three panels, in a day-long workshop. The resulting values were used to calibrate the disability scale between 0 (no disability) and 1 (extreme disability). In the second step, all other remaining disease stages were interpolated on the disability scale by the individual panel members. The panel members valued a random set of 30 disease stages by placing each disease stage separately in between two indicator disease stages on the disability scale, in a paper and pencil procedure performed individually.

The 16 indicator disease stages were deliberately selected from the list of 175 disease stages, applying three criteria. Firstly, their expected valuations (as estimated by a statistical model) should evenly cover the total range. Secondly, they should have a sizeable public health impact. Thirdly, they should be relatively easy to recognize and interpret.

The person trade-off (PTO) method, as operationalized by Murray,[4] was applied as the valuation method for the indicator disease stages in the panel sessions. PTO is essentially a trade-off method which requires the subjects to trade-off person years lived healthy or with some de-

fined disability.[1,8] In the GBD Study and in the current study, PTO was applied in two forms, so as to stimulate the subjects to consider their valuation of each state from different viewpoints. The whole PTO procedure was designed to enhance deliberation, so that each panel member would arrive at a well-considered valuation. The first form, labelled PTO1, refers to the trade-off between extending life of 1,000 healthy persons by 1 year versus extending life of *N* persons by 1 year in a specified disabled health state (N≥1,000). The options are mutually exclusive. The second form, labelled PTO2, refers to a trade-off between extending life of 1,000 healthy persons by 1 year versus complete and instantaneous cure of *N* persons in a disabled health state and extending their life in perfect health by 1 year. Again, the options are mutually exclusive. (The exact wordings of PTO1 and PTO2 are available on request from the authors.)

PTO has been advocated as the appropriate method of deriving disability weights for the estimation of burden of disease, primarily because it forces the subjects to judge health states from a public viewpoint, e.g. for groups of individuals, not including him- or herself.[8,9] In the current study, the choice for PTO and the adherence to the GBD operationalization of the PTO tasks was further motivated by the argument of comparability with the severity weights of the GBD Study.

Subjects

A deliberate choice was made to employ health care professionals with sufficient knowledge of the consequences of a broad variety of diseases, in accordance with the GBD Study.[4] Three panels of 15 medical experts were recruited, all of them MDs, predominantly GPs with ample general medical experience and extensive research training. Hence, the subjects in the current study were the individual panel members.

Analysis

The results of the PTO valuations were converted by a linear transformation into a 0–1 scale, ranging from no disability (0) to extreme disability (1).

Generalizability theory was applied as a general approach (G study) to estimate the proportion of the total variance uniquely attributable to the object of measurement (disease stages) and the other sources of variance (e.g. measurement error and bias).[10,11] G study is a specific application of analysis of variance (ANOVA). It allows for estimation of the relative contribution (variance components) of the disease stages, the panels and the panel members nested within panels to the total variance of the model. Furthermore, G study was used to estimate the reliability of the valuations as agreement among panel members. To this end, a G coefficient can be computed as a measure of internal consistency, comparable to the intraclass correlation coefficient (ICC) for two variables, defined as the proportion of total variation in scores accounted for by the disease stages.[10,12]

The design of the Dutch Disability Weights Study has been described extensively elsewhere.[13]

RESULTS

In total 38 medical experts (28 men and ten women of mean age 47.7 years with SD=9.2 years) participated. The mean number of years of medical experience was 15.5; 21 panel members were still involved in direct patient care. No differences between panels existed regarding sex, age and medical experience.

All panel members performed the PTO task for the 16 indicators although some initially experienced some difficulty. After inspection of the results, the valuations of four panel members were discarded from the analysis because they showed unacceptable lack of variance. The interpolation task was performed consistently by all 38 medical experts.

Descriptive statistics

The mean disability weights for the indicator disease stages were used to construct a disability scale *(figure 2)*, ranging from 0 indicating the best imaginable health state at the bottom (no disability) to 1 indicating the worst imaginable health state at the top (extreme disability). To show the range of the valuations for each indicator disease stage, the standard deviation was also given. The mean disability weights for most diseases did not vary across panels. The Spearman rank order correlations between panels were high ($r_S \geq 0.94$), indicating a similar rank ordering.

To illustrate the results of the interpolation, *table 1* shows the disability weights (in 11 disability classes) for 28 disease stages from ten ICD-9 chapters. The 11 disability classes were arbitrarily based on the disability weights for the indicator disease.

The mean correlation between the interpolations of individual panel members with the rest of the group (Pearson correlation = 0.95) indicated high interrater reliability. The test-retest reliability after 2 months was also very satisfactory (Spearman rank correlation = 0.94) at the aggregated group level. (Complete data are available on request.)

Variance components analysis and reliability

G study results show that the major part of the variance can be attributed to variation in the disease stages offered for valuation *(table 2)*. The effects of panel members and panels are very small, implying that the panel members valued the disease stages according to the stimulus only and that no systematic difference existed between panels. Overall, 18.3% of the variance was attributable to the interaction terms and measurement error. The interpretation of the interaction between disease stages and panel members is that some panel members valued some disease stages systematically different.

The G-coefficient can be computed from the variance component of the disease stages as a percentage of the total variance equals 0.80, indicating acceptable internal consistency.

Validity

Table 3 compares some Dutch disability weights with the GBD disability weights for comparable indicator condi-

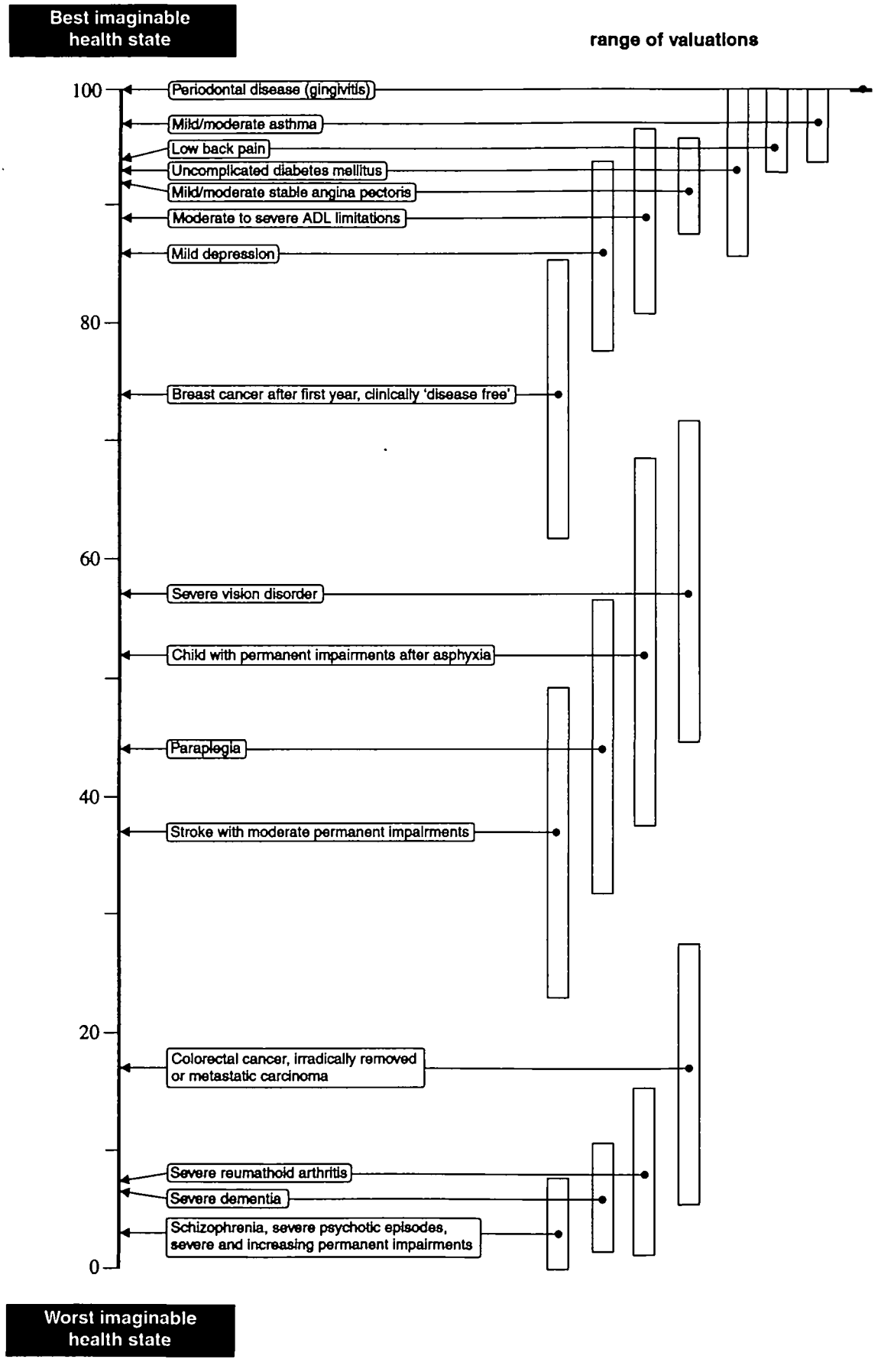

Figure 2 Disability scale, based on PTO means and 67% CI (standard deviations)

Table 1 Disability weights for 28 disease stages from ten disease categories, arranged in 11 disability classes

Disability class	Disability weights	Disease stage
1	0.00–0.01	Acute bronchitis (one episode of 2 weeks in an otherwise healthy year)
2	0.01–0.05	Mild vision disorder (i.e. some difficulty reading small newspaper print and no difficulty recognizing faces at 4 m distance)
		Acute bronchitis (more episodes of 2 weeks in an otherwise healthy year)
3	0.05–0.10	Mild/Moderate stable angina pectoris (NYHA 1–2)
		Pneumonia (one episode of 2 weeks in an otherwise healthy year)
4	0.10–0.15	Mild depression
5	0.15–0.20	Low level spina bifida aperta (sacral), permanent stage
		Moderate vision disorder (i.e. great difficulty reading small newspaper print and some difficulty recognizing faces at 4 m distance)
		Colorectal cancer, state after intentionally curative primary therapy
6	0.20–0.30	Mild rheumatoid arthritis
		Mild dementia (only daily activities significantly impaired)
7	0.30–0.40	Multiple sclerosis, 'relapsing-remitting' phase
		Moderate depression
		Stroke with mild permanent impairments
8	0.40–0.50	Colorectal cancer (stage of diagnosis and primary therapy)
		Severe vision disorder (i.e. unable to read small newspaper print and great difficulty or unable to recognize faces at 4 m distance)
		Medium level spina bifida aperta (L3–L5), permanent stage
9	0.50–0.65	Severe stable angina pectoris (NYHA 3–4)
		Stroke with moderate permanent impairments
		Moderate dementia (living independently impossible without limited supervision)
10	0.65–0.80	High level spina bifida aperta (L2 or higher), permanent stage
		Multiple sclerosis, progressive phase
		Severe depression, without psychosis and/or hallucinations
11	0.80–1.00	Colorectal cancer, irradically removed or metastatic carcinoma
		Severe depression, with psychosis and/or hallucinations
		Stroke with severe permanent impairments
		Severe rheumatoid arthritis
		Severe dementia (permanent supervision required)

tions.[1,6] Five out of 12 disease stages were classified in the same disability class. Two were classified almost exactly on the boundary and the remaining five in an adjacent class. These discrepancies can be explained from the Western European instead of global context of the valuations and from refinements in the valuation protocol. 'Infertility', 'severe vision disorder', and 'mild mental retardation' probably have less consequences in a Western European situation than in, for example, sub-Saharan Africa or Asia. The addition of a health status description, specifying the functional health status associated with the disease stage and, thus, adding information and leaving less room for variation may also account for the discrepancies. Finally, 'mild/moderate stable angina pectoris' and 'severe depression' in the Dutch study were presented as being one stage in disease processes that each consisted of several stages, whereas in the GBD Study they were offered as a single stage of each disease.

DISCUSSION

The primary outcome of the Dutch project is a comprehensive and coherent set of disease-specific disability weights for 175 disease stages associated with 52 disease categories (cf. ICD-9). This comprehensive set of disease-specific disability weights allows public health researchers to quantify the relative contributions of different disease categories to the total burden of disease in The Netherlands. The Dutch disability weights were applied in the Dutch Public Health Status and Forecast – 1997 study.[6] Similar to the GBD Study, the results indicate the importance of including both mortality and morbidity in the total burden of disease. Examples of the application of the Dutch disability weights (*table 4*) clearly show that the relative contribution of disability-weighted disease years and of years of life lost due to premature death differ for different groups of diseases. Such combined data, however, strongly rely on valid epidemiological data on frequency and duration of disease episodes. More work remains to be done in that area.

Secondary results of the Dutch disability weight study include confirmation of the feasibility of the valuation protocol and several improvements to the original GBD protocol. The differences in disability weights between the Dutch disability weight and GBD studies may reflect the more refined considerations of the different disease

Table 2 Estimated variance components (percent) of disease stages (16) × panels (3) × panel members (34)

Sources of variance	Variance	Contribution %
Disease stages (D)	0.130	79.9
Panels (P)	0.001	0.6
Panel members (subjects) nested within panels (S(P))	0.002	1.2
Disease stage × panel members nested within panels (D × S(P))	0.004	2.4
All other interactions plus error	0.026	15.9

stages. A recent study on the stability of disability weights in different countries and informant groups showed that the ranking of the disabling effect of different health conditions is fairly similar across the world, with a few exceptions such as HIV-infections and physical conditions e.g. blindness.[14]

The feasibility of establishing reliable disability weights associated with the various diseases that constitute the major part of the burden of disease in a Western European country has been demonstrated. And the addition of a standardized description of functional health status to the disease label proved indispensable to the valuation process.

Apart from estimating burden of disease, the disability weights allow us to quantify the potential benefits of any intervention that can help reduce the total burden of disease. Such a link to QALY-type measures of cost-effectiveness data from health technology assessment will contribute to the ongoing political discussion on priority setting in health policy based on the perspective of the population's health. We feel that the precision of our valuations is adequate to allow conclusions at the public health level if disease-specific trends or changes are evaluated.

The authors thank the Ministry of Public Health, Welfare and Sports for the financial support for this research project.
The authors wish to thank the GPs, some members of the clinical staff of the Academic Medical Center Amsterdam and the other panel members for their willingness to participate. We would like to thank Professor C.J.L. Murray for his interest and encouragement.

REFERENCES

1 Murray CJL, Lopez AD, editors. 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: Harvard University Press, 1996.

2 Murray CJL, Lopez AD. Mortality by cause for eight regions in the world: Global Burden of Disease Study. Lancet 1997;349:1269-76.

3 Murray CJL, Lopez AD. Regional patterns of disability-free life expectancy and disability-adjusted life expectancy: Global Burden of Disease Study. Lancet 1997;349:1347-52.

4 Murray CJL. Rethinking DALYs. In: Murray CJL, Lopez AD, editors. 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: Harvard University Press 1996:1-98.

5 Anand S, Hanson K. Disability-adjusted life years: a critical review. J Hlth Econ 1997;16:685-702.

6 Ruwaard D, Kramers PGN, editors. Public health status and forecast - 1997: health, prevention and health care in The Netherlands until 2015. Utrecht: Elsevier/de Tijdstroom, 1998.

Table 3 Comparison of Global Burden of Disease Study – Dutch disability weights

GBD Study (WHO/Worldbank)			Dutch disability weights study		
Indicator condition	Disability class	Severity weight	Disease stage	Disability class	Disability weight
Infertility	3	0.12–0.24	Infertility from STD	2	0.11
Angina	3	0.12–0.24	Mild/moderate stable angina pectoris	2	0.08
Rheumatoid arthritis	3	0.12–0.24	Mild rheumatoid arthritis	3	0.21
Deafness	4	0.24–0.36	Severe hearing disorder in elderly	5	0.37
Blindness	6	0.50–0.70	Severe vision disorder	5	0.43
Mild mental retardation	5	0.36–0.50	Mild mental handicap	4	0.29
Down's syndrome	5	0.36–0.50	Child with Down's syndrome without other congenital anomalies	6	0.51
Paraplegia	6	0.50–0.70	Paraplegia	6	0.57
Unipolar major depression	6	0.50–0.70	Severe depression	7	0.76
Active psychosis	7	0.70–1.00	Schizophrenia, several psychotic episodes, severe and increasing permanent impairments	7	0.98
Dementia	7	0.70–1.00	Severe dementia	7	0.94
Quadriplegia	7	0.70–1.00	Quadriplegia	7	0.86

Table 4 Application of disability weights in The Netherlands: some examples of the calculation of DALYs

Disease	Year prevalence (LY)	Disability weight (Q)	Q × LY	Years of life lost due to premature death	DALYs
Lung cancer	18,500	0.42	7,800	115,300	123 300
Stroke	97,200	0.61	59,300	110,400	169 700
Rheumatoid arthritis	80,700	0.53	42,700	2,700	45 500
Depression	484,200	0.23	111,700	–[a]	111 700

a: Mortality and years of life lost are very low and without virtual implication; suicides were not included.

Source: Ruwaard, Kramers[6] (p.50)

7 Brooks R. EuroQol: the current state of play. Hlth Policy 1996;37:53-72.
8 Nord E. The person trade-off approach to valuing health care programs. Med Decis Making 1995;15:201-8.
9 Pinto Prades JL. Is the person trade-off a valid method for allocating health care resources? Hlth Econ 1997;6:71-81.
10 Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press, 1995.
11 Krabbe PFM, Essink-Bot ML, Bonsel GJ. On the equivalence of collectively and individually collected responses: standard gamble and time trade-off judgements of health states. Med Decis Making 1996;16:120-32.
12 Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Control Clin Trials 1991;12:142S-58S.
13 Stouthard MEA, Essink-Bot ML, Bonsel GJ, et al. Disability weights for diseases in The Netherlands. Rotterdam: Department of Public Health, Erasmus University Rotterdam, 1997.
14 Üstün TB, Rehm J, Chatterji S, et al. Multiple-informant ranking of the disabling effects of different health conditions in 14 countries. Lancet 1999;354:111-5.

Received 23 October 1998, accepted 16 September 1999