Research

A generic model for the assessment of disease epidemiology: the computational basis of DisMod II

Jan J Barendregt*[1], Gerrit J van Oortmarssen[1], Theo Vos[2] and Christopher JL Murray[3]

Address: [1]Department of Public Health, Erasmus MC, Rotterdam, Netherlands, [2]Victorian Government Department of Human Services, and Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia and [3]Global Programme on Evidence for Health Policy, World Health Organization, Geneva, Switzerland

Email: Jan J Barendregt* - j.barendregt@erasmusc.nl; Gerrit J van Oortmarssen - g.vanoortmarssen@erasmusc.nl; Theo Vos - theo.vos@dhs.vic.gov.au; Christopher JL Murray - murrayc@who.int

* Corresponding author

Abstract

Epidemiology as an empirical science has developed sophisticated methods to measure the causes and patterns of disease in populations. Nevertheless, for many diseases in many countries only partial data are available. When the partial data are insufficient, but data collection is not an option, it is possible to supplement the data by exploiting the causal relations between the various variables that describe a disease process. We present a simple generic disease model with incidence, one prevalent state, and case fatality and remission. We derive a set of equations that describes this disease process and allows calculation of the complete epidemiology of a disease given a minimum of three input variables. We give the example of asthma with age-specific prevalence, remission, and mortality as inputs. Outputs are incidence and case fatality, among others. The set of equations is embedded in a software package called 'DisMod II', which is made available to the public domain by the World Health Organization.

Background

Assessment of the epidemiology of a disease is often very hard. Data on incidence, prevalence and disease specific mortality are frequently incomplete, not very reliable, or altogether lacking. The solution of choice is gathering good data, but this is time-consuming, often difficult, and always costly. When primary data collection is no real option, as in a burden of disease study where the goal is a comprehensive overview of the epidemiology of a large number of diseases, additional methods of assessing disease epidemiology are needed.

Additional information can be derived from the logical relations between the variables that describe a disease. By definition, a prevalent case must have been incident at some earlier time and age. Also, it is impossible to die or recover from a disease without having had the disease, however brief. These logical relations can be expressed as a formal model of a generic disease process. Such a formal disease model allows calculation of a complete and internally consistent description of disease epidemiology from partial data.

For the Global Burden of Disease 1990 study a generic formal disease model was implemented as a computer model called 'DisMod' [1,2]. In that study and in subsequent country studies, DisMod has been used extensively to supplement missing data and force consistency on data that

were available. DisMod is based on a set of differential equations that describe age specific incidence, remission, case fatality, and 'all other causes' mortality. With total mortality and three transition hazards – incidence, remission, and case fatality – as inputs, the equations were solved numerically using an iterative approximation method, the finite differences method. It was also possible to enter a relative risk on total mortality as an alternative input to case fatality, but, given total mortality, these are equivalent [3]. Other disease variables, such as prevalence and disease specific mortality, were derived from this solution, but could not be used as inputs.

In the field of chronic diseases a similar, but simpler, model has been developed and used for purposes of assessing disease epidemiology and the calculation of 'what if'-scenarios [4,5]. This disease model is simpler because, being about chronic diseases such as diabetes only, remission can be ignored. This simplification allows analytical solutions of the differential equations to be used, instead of requiring a numerical approximation [6].

For the Global Burden of Disease 2000 study it was decided to develop a new computer model, called 'DisMod II', which would serve the same purposes as the original DisMod, but would have enhanced usability, such as an interactive graphical interface. An important new feature was to allow for a wider range of disease inputs than the three transition hazards used in DisMod (incidence, remission and case fatality). In particular, prevalence and disease specific mortality would be potential inputs in the new model.

To facilitate interactive use of such a model, speed of computation is essential, and therefore an analytical solution of the differential equations was preferred over a numerical one. Here we report a set of equations that represent the analytical solution of the differential equations. This set of equations forms the computational basis of DisMod II.

The problem and a solution
Conceptual model
The conceptual model of DisMod II, like the original DisMod, is that of a multi-state life table, depicted in figure 1. The model describes a single disease, together with mortality from all other causes. Healthy people, defined as people unaffected by the disease being modeled, are subject to an incidence hazard, and may become diseased. When diseased they are subject to a hazard of dying from the disease, the case fatality, and to a hazard of recovery from the disease, called remission. Both healthy and diseased people are subject to the same mortality hazard from all other causes.

This mortality from all other causes poses a problem: it is an input to the disease model, but often it is not known. It could be calculated from the total mortality rate and the disease specific mortality, but frequently the disease specific mortality is not known.

To get around this problem we use the property that hazards are unaffected by the presence or absence of other hazards that act on the same population. If it is assumed that mortality from all other causes is independent of the disease, i.e., that it is the same for healthy and diseased people, this implies that the transition hazards for incidence, remission and case fatality are not affected by the value of the 'all other causes' mortality. Therefore we can set the value of mortality from all other causes to 0 (i.e., leave it out of the equations) and still derive the right values for the disease rates. Disease prevalence, when reported as a proportion of the total population, will also be unaffected [6].

The assumption of independence of the mortality from all other causes implies that the disease-specific mortality in the model stands for all excess mortality caused by the disease, which is not necessarily the same as the disease-specific mortality reported by national statistical offices. This definition of disease-specific mortality complies with the methodology of burden of disease studies, which aims to attribute all excess mortality to the disease.

One of the DisMod II outputs – disease duration – is affected by mortality from all other causes. DisMod II therefore calculates results in two steps. First it calculates the numbers of people in the three states 'healthy', 'diseased' and 'dead from the disease' for all ages and derives disease rates such as incidence, prevalence and mortality. Next, from the disease specific mortality rate it calculates the mortality rate for all other causes, thus making it possible to calculate disease duration.

DisMod II calculates the disease starting with a cohort of 1000 people at age 0, and working up to the highest age considered. Within an age interval transition hazards are assumed constant, and to minimize the impact of this assumption the calculation is done in 1-year age intervals.

Basic equations
For a single cohort the following three differential equations describe the conceptual model when 'all other causes' mortality is ignored:
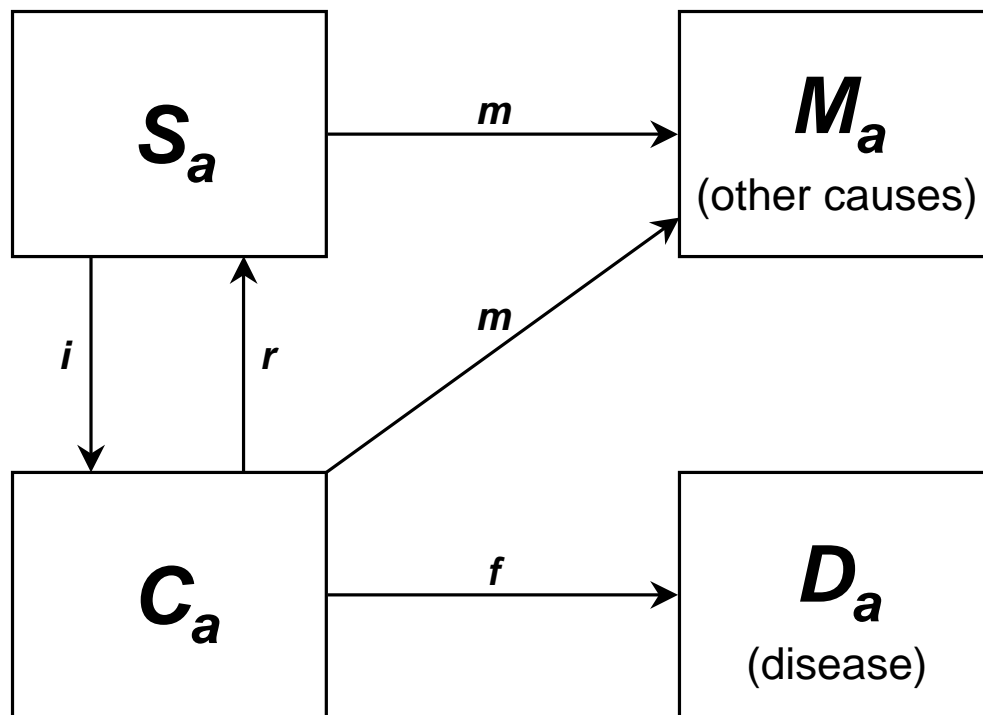
Figure 1
The conceptual disease model *S*: number of healthy people (i.e., without the disease under consideration); *C*: number of diseased people; *D*: number of people dead from the disease; and *M*: number of people dead from all other causes, with *a* an age subscript. There are four transition hazards: *i*: incidence, *r*: remission, *f*: case fatality, and *m*: all other mortality.

$$(1) \quad \frac{dS_a}{da} = -i_a S_a + r_a C_a$$

$$(2) \quad \frac{dC_a}{da} = -(f_a + r_a)C_a + i_a S_a$$

$$(3) \quad \frac{dD_a}{da} = f_a C_a$$

where the three model parameters, representing the transition hazards, are:

i: incidence

r: remission

f: case fatality

and the three states are:

S_a : Number of healthy people at age *a*

C_a : Number of diseased people at age *a*

D_a : Number of dead people at age *a*

Equations that express the number of people in each of the states *S*, *C*, and *D* at age *a* as a function of the parameters *i*, *r*, and *f* were derived from equations 1–3 using Maple V [7]. To simplify the derived equations we first define a number of intermediate variables:

$$l_a = i_a + r_a + f_a$$

$$q_a = \sqrt{i_a^2 + 2i_a r_a - 2i_a f_a + r_a^2 + 2f_a r_a + f_a^2}$$

$$w_a = e^{\frac{-(l_a + q_a)}{2}}$$

$$v_a = e^{\frac{-(l_a - q_a)}{2}}$$

Using these intermediate variables the derived equations then become:

$$(4) \quad S_a = \frac{2(v_a - w_a)\left[S_{a-1}(f_a + r_a) + C_{a-1}r_a\right] + S_{a-1}\left[v_a(q_a - l_a) + w_a(q_a + l_a)\right]}{2q_a}$$

$$(5) \quad C_a = -\frac{(v_a - w_a)\left\{2\left[(f_a + r_a)(S_{a-1} + C_{a-1}) - l_a S_{a-1}\right] - C_{a-1}l_a\right\} - C_{a-1}q_a(v_a + w_a)}{2q_a}$$

$$(6) \quad D_a = \frac{(v_a - w_a)\left\{2f_a C_{a-1} - l_a(S_{a-1} + C_{a-1})\right\} - q_a(S_{a-1} + C_{a-1})(v_a + w_a) + 2q_a(S_{a-1} + C_{a-1} + D_{a-1})}{2q_a}$$

Prevalence and mortality

From the resulting numbers of people in *S*, *C* and *D* for all ages the age specific prevalence proportion and mortality rates are calculated. First, for each age interval person years at risk (PY_a) are calculated:

$$(7) \quad PY_a = 0.5(S_a + C_a + S_{a+1} + C_{a+1})$$

The prevalence proportion c_a then becomes:

$$(8) \quad c_a = 0.5\left(\frac{C_a + C_{a+1}}{PY_a}\right)$$

and the mortality rate b_a is:

$$(9) \quad b_a = \frac{D_{a+1} - D_a}{PY_a}$$

Disease duration

The age-specific mortality rates allow derivation of mortality from all other causes (*m*), needed to calculate disease duration. The equations below describe the expected duration of disease for a person who became incident in the age interval [*a*,*a* + 1), while taking mortality from all other causes *m* into account. We define:

$\beta_a = r_a + f_a + m_a$: the total hazard to leave the diseased state C_a

$\gamma_{a,d}$: probability to be in the diseased state after duration *d*, $\gamma_{a,0} = 1.0$

$x_{a,d}$: contribution of duration in the interval [*d* - 1,*d*) to the total duration in the disease state after incidence in [*a*, *a* + 1): X_a

Then, for incidence in the interval [*a*, *a* + 1) the equations for that first year are:

$$(10) \quad \gamma_{a,1} = \begin{cases} \gamma_{a,0} \dfrac{i_a}{i_a - \beta_a} \dfrac{\left(e^{-\beta_a} - e^{-i_a}\right)}{\left(1 - e^{-i_a}\right)} & \text{if } \beta_a \neq i_a \\[2em] \gamma_{a,0} i_a \dfrac{e^{-i_a}}{\left(1 - e^{-i_a}\right)} & \text{otherwise} \end{cases}$$

and

$$(11) \quad x_{a,1} = \begin{cases} \dfrac{1}{\beta_a} - \dfrac{i_a}{\beta_a(i_a - \beta_a)} \dfrac{\left(e^{-\beta_a} - e^{-i_a}\right)}{\left(1 - e^{-i_a}\right)} & \text{if } \beta_a \neq i_a \\[2em] \dfrac{1}{\beta_a} - \dfrac{1}{e^{\beta_a} - 1} & \text{otherwise} \end{cases}$$

For durations in subsequent years [*a* + *k*, *a* + *k* + 1), *k* = 1,2,3..., the following two equations apply:

$$(12) \quad \gamma_{a,k+1} = \gamma_{a,k} e^{-\beta_{a+k}}$$

and

$$(13) \quad x_{a,k+1} = \frac{\gamma_{a,k}}{\beta}\left(1 - e^{-\beta_{a+k}}\right)$$

Total duration X_a for incidence in [*a*, *a* + 1) then becomes:

$$(14) \quad X_a = \sum_{k=1}^{\infty} x_{a,k}$$

Implementation

Availability

The equations above are implemented as a software package, designed for use by epidemiologists and public health scientists. Users combine the available data and their own expert knowledge interactively to produce best estimates of the epidemiology of the disease. To this end DisMod II comes with a graphical interface and a number of features for fitting curves to and interpolation of the in-

put data. There is an extensive online help, including a tutorial to help users to get started. The software runs on Windows 95 and higher, and is available from the website of WHO http://www.who.int/evidence/dismod/.

Input variables
An explicit aim of the development of DisMod II was to allow for a wider range of input variables than the three transition hazards in the original DisMod. Equations 4–6 allow calculation of the numbers of people in the three states when the three transition hazards incidence, remission, and case fatality are known. Often these transition hazards are not observed, but, for example, prevalence and disease-specific mortality are. Allowing for prevalence and mortality as inputs directly would require rewriting equations 4–6 accordingly, but this may not be tractable.

In those cases the analytical solution presented here is supplemented by an iterative optimization method, the 'downhill simplex method' [8]. This is an optimization method in multiple dimensions, which in this case are the three transitions hazards. Starting at the lowest age group, values for the transition hazards are inserted in equations 4–6, and all output variables are calculated. A loss function then evaluates the difference between the input variables and the corresponding output, and based on this evaluation a different set of values of the transition hazards is inserted. This procedure is iterated until the loss function reaches a minimum, and the optimization moves to next age group.

Because of this combination of analytical and numerical methods DisMod II accepts, in addition to the transition hazards incidence, remission and case fatality (or its equivalent relative risk for total mortality), the following disease input variables: incidence as a population rate (with total population in the denominator instead of person years at risk), prevalence, duration, and mortality. Because of the two-step calculation procedure duration is a valid input only when it is short, preferably less than one year.

When the input variables do not consist of the three transition hazards incidence, remission and case fatality, they may be (and often are) internally inconsistent. In that case the 'downhill simplex' optimization procedure will adjust the values of the input variables such that they are internally consistent while staying as close as possible to the original values. The user can influence the outcome by applying different weights to the input variables: an input variable with a higher weight will remain closer to the original value. The same procedure applies when more than three inputs are available, i.e., when the model is overidentified.

Generally, at least three disease input variables are needed to calculate the full disease epidemiology; the exception is when case fatality and relative risk for total mortality are given. Case fatality and relative risk for total mortality contain the same information, given total mortality, and therefore count as only one input when they are both included.

In addition to the disease input variables, DisMod II needs total mortality rates and population numbers for the population under study. All input variables are by age, and calculations are done separately for men and women.

Trends in disease epidemiology
Equations 1–3 describe a life table cohort, which, when interpreted as a description of a cross-sectional population, implies an assumption of steady state for the disease. However, it is possible to include past trends on the transition hazards incidence, remission, and case fatality in the calculation. DisMod II then switches to a fully dynamic calculation mode: it still uses equations 4–6, but for each age separately. In life table mode, disease variables at age a depend on variables at age $a - 1$; in dynamic mode, variables at age a and time t depend on variables at age $a - 1$ and time $t - 1$ (which, because of the trends, have not the same values as those at age $a - 1$ and time t). DisMod II still tries to reproduce the currently observed input variables, but taking past trends into account will result in different values for the unobserved variables. Dynamic mode requires considerably longer computation time than life table mode.

Uncertainty
Uncertainty intervals for the output can be obtained by specifying distributions (and parameters) for the disease input variables. A number of distributions are available, among them Poisson, binomial, and normal. DisMod II uses these distributions and associated parameters in a monte carlo simulation (or parametric bootstrapping [9]). For each of the input variables a value is randomly chosen from its distribution, and the model is calculated. This procedure is repeated a large (user specified) number of times, resulting in distributions for all output variables. From the distributions of the output variables uncertainty intervals are derived. This too may take considerable computation time.

Even if the input variables were internally consistent, the randomly sampled values from their distributions will not be. This causes DisMod II to adjust the input variables to output values that are internally consistent, which has an impact on the width of the uncertainty intervals. Randomly sampled values from the distributions of the input variables will cause the adjustments of the output to have a distribution as well. Consequently the uncertainty inter-

Table 1: Asthma prevalence, remission and mortality rates (per 1000 population) by age and sex, Victoria 1996*

Age (years)	Males			Females		
	Prevalence	Remission	Mortality	Prevalence	Remission	Mortality
0–4	66.00	46.00	0.00	45.00	46.00	0.00
5–14	132.00	79.00	0.01	94.40	79.00	0.00
15–24	63.00	75.00	0.01	62.50	75.00	0.01
25–34	46.00	27.50	0.00	59.90	27.50	0.00
35–44	46.00	12.00	0.01	67.70	12.00	0.01
45–54	48.00	14.00	0.01	78.80	14.00	0.03
55–64	46.00	29.50	0.06	79.70	29.50	0.04
65–74	38.00	35.00	0.14	70.90	35.00	0.12
75+	30.00	30.77	0.51	63.60	30.77	0.44

*See text for references

vals do not just reflect the sampling variability of the individual variables (which is why they are not called 'confidence intervals').

An example

We illustrate DisMod II with the example of asthma from the Victorian Burden of Disease study. Deaths for Victoria in 1996 are from the Australian Bureau of Statistics [10]. Prevalence is based on a number of Australian studies, and male-to-female ratios for children and adults. Remission rates are based on a follow-up study from the United States [11].

In table 1 the resulting DisMod II input prevalence, remission and mortality for asthma is presented for males and females. We assumed that at birth the prevalence of asthma is zero. For the calculation of uncertainty intervals we assumed prevalence to have a binomial distribution, and remission and mortality a Poisson distribution.

Tables 2 (males) and 3 (females) present a selection of the corresponding DisMod II output. Shown are incidence, prevalence, remission, case fatality and mortality, each including 95% uncertainty intervals. Note that the output prevalence, remission and mortality are not identical to their input values in table 1: the observed input variables are not necessarily internally consistent, while the calculated output always is. The calculation also smoothes the age pattern of the variables.

Discussion

Epidemiology is first and foremost an empirical science. But the emphasis on observations, while certainly justified, does not preclude that using more theoretical tools like DisMod II may contribute as well.

We see two main applications for DisMod II. The first is supplementing incomplete data, and this was illustrated by our asthma example. With data for prevalence, remission and mortality, it is possible to calculate the complete epidemiology of asthma, including back-calculating the incidence. This application is useful when data are scarce but an estimate of disease epidemiology is urgently needed. This is a common situation in burden of disease studies.

Supplementing incomplete data is not a fully automatic process, however. For example, available data tend to come in wide age intervals. From the point of view of DisMod II, the differences in value between adjacent age groups are discontinuities, which may be impossible to resolve without very extreme values of one or more variables. In particular, when back-calculating incidence from, among others, prevalence, such discontinuities may result in huge spikes in the back-calculated incidence.

It is the responsibility of the user to guard against such 'solutions'. DisMod II tries to help by showing graphs of the input and output variables, and by providing interpolation and smoothing algorithms. But in the end it is the user who has to decide whether the outcomes are acceptable, and if not, what strategy is needed to resolve the problem (smoothing the input, using different weights for the input variables, etc.). Using DisMod II for this purpose is an interactive exercise.

The second application is checking for internal consistency of existing estimates. Empirical observation of the variables that describe the epidemiology of a disease is subject to measurement error, which may affect variables differently. For example, for a disease with a gradual onset, such as dementia, it is much harder to estimate incidence than prevalence. Measurement error may go undetected, but if we know it exists it may be possible to account for it. Checking the internal consistency of the es-

Table 2: DisMod II outputs, males Asthma incidence, prevalence, remission, case fatality and mortality rates per 1000 population (95% uncertainty interval)

Age (years)	Incidence	Prevalence	Remission	Case fatality	Mortality
0–4	21.86 (17.07, 26.65)	50.57 (37.96, 63.25)	45.92 (44.40, 47.43)	0.02 (0.00, 0.26)	0.00 (0.00, 0.00)
5–14	14.94 (12.05, 17.85)	124.30 (108.17 140.62)	81.73 (76.99, 86.50)	0.05 (0.00, 0.56)	0.01 (0.00, 0.03)
15–24	1.62 (0.83, 2.45)	82.96 (65.43, 100.83)	86.18 (75.73, 96.86)	0.10 (0.01, 0.60)	0.01 (0.00, 0.04)
25–34	0.16 (0.00, 0.67)	54.48 (41.11, 67.98)	28.31 (24.08, 32.51)	0.07 (0.00, 0.57)	0.00 (0.00, 0.06)
35–44	0.51 (0.00, 1.05)	47.82 (37.72, 57.93)	12.13 (10.10, 14.10)	0.19 (0.01, 0.69)	0.01 (0.00, 0.19)
45–54	0.84 (0.19, 1.49)	48.95 (39.66, 58.27)	13.39 (11.33, 15.44)	0.35 (0.02, 0.88)	0.02 (0.00, 0.37)
55–64	1.14 (0.42, 1.87)	47.34 (36.78, 58.03)	28.89 (24.98, 32.79)	1.38 (0.40, 2.36)	0.07 (0.00, 0.56)
65–74	0.78 (0.21, 1.37)	40.75 (29.24, 52.30)	35.47 (31.62, 39.30)	4.00 (0.58, 7.44)	0.16 (0.00, 0.67)
75+	0.89 (0.34, 1.47)	32.76 (22.59, 43.12)	32.07 (29.76, 34.36)	14.70 (1.02, 28.88)	0.48 (0.00, 0.99)

*See text for references

Table 3: DisMod II outputs, females Asthma incidence, prevalence, remission, case fatality and mortality rates per 1000 population (95% uncertainty interval)

Age (years)	Incidence	Prevalence	Remission	Case fatality	Mortality
0–4	15.17 (11.05, 19.37)	34.79 (24.08, 45.57)	45.87 (44.24, 47.49)	0.01 (0.00, 0.49)	0.00 (0.00, 0.33)
5–14	10.34 (8.06, 12.63)	90.69 (73.56, 107.92)	80.45 (75.25, 85.70)	0.01 (0.00, 0.52)	0.00 (0.00, 0.50)
15–24	3.12 (1.99, 4.25)	67.31 (50.60, 84.29)	76.07 (69.13, 83.14)	0.08 (0.00, 0.58)	0.01 (0.00, 0.49)
25–34	1.79 (0.91, 2.66)	61.01 (46.99, 75.23)	27.49 (24.99, 29.98)	0.03 (0.00, 0.54)	0.00 (0.00, 0.50)
35–44	1.76 (0.83, 2.69)	67.17 (54.71, 79.74)	12.00 (10.99, 12.99)	0.21 (0.01, 0.72)	0.01 (0.00, 0.49)
45–54	2.25 (1.10, 3.41)	76.94 (64.20, 89.74)	13.47 (12.31, 14.62)	0.36 (0.02, 0.87)	0.03 (0.00, 0.36)
55–64	2.44 (1.24, 3.64)	79.69 (66.30, 93.07)	29.12 (26.66, 31.58)	0.53 (0.08, 1.05)	0.04 (0.00, 0.41)
65–74	1.94 (0.93, 2.96)	72.62 (59.09, 86.25)	35.19 (31.93, 38.44)	1.89 (1.03, 2.75)	0.14 (0.00, 0.64)
75+	2.11 (1.18, 3.06)	65.00 (52.31, 77.74)	32.00 (29.20, 34.80)	6.45 (3.96, 8.95)	0.42 (0.00, 0.93)

timates with DisMod II may help to detect the existence of measurement error.

This application also carries certain difficulties. Inconsistency of cross-sectional variables describing a disease may be real or deceptive [12]. Real inconsistency may be due to the combination of measurements from different sources, or to measurement and sampling error. Deceptive inconsistency is due to past trends in disease epidemiology. The problem is that it is not possible to distinguish between the two, unless the epidemiology of the disease (including the past) is fully known, and in that case there is little need for a tool like DisMod II.

Without complete knowledge of the disease epidemiology this dilemma can be solved only by expert judgement. When it is unlikely that trends in the past have existed, or when a sensitivity analysis shows that reasonable past trends are unable to explain the inconsistency, the expert may decide that most of the inconsistency is real, and also which variable is most likely to be in error.

The message here is that DisMod II is a tool for experts, who should carefully weigh all available evidence, of which the DisMod II output is only a part.

All this assumes that the conceptual model underlying DisMod II is applicable to a wide range of diseases with very different epidemiology. While most aspects of the conceptual model are a matter of definition, this is not true for survival in the diseased state. This survival is piece-wise exponentially distributed: exponential within each year of age, with (possibly) a different hazard for each age. However, a check against survival data that at the individual level were lognormally distributed, showed that on the population level DisMod II was able to reproduce the data very well [12].

One kind of disease for which the conceptual model may not be applicable is an infectious disease that confers immunity. DisMod II assumes that those who remit go back to the pool of susceptibles, which in the case of acquired immunity clearly is not appropriate.

Where epidemiologic data are incomplete or in doubt, DisMod II may prove to be helpful. Thanks to the analytical solution described here it is fast enough to allow interactive use. When a single disease is to be studied, customized tools may be more appropriate, but when the epidemiology of a large range of diseases must be assessed, as in burden of disease studies, the generic DisMod II disease model will be a useful tool.

Competing interests
None declared.

Authors' contributions
JJB developed the model and drafted the manuscript. GJvO contributed to the mathematics. TV provided the asthma data. CJLM contributed to the mathematics. All authors commented on earlier drafts and read and approved the final manuscript.

Acknowledgements

References
1. Murray CJ and Lopez AD **Regional patterns of disability-free life expectancy and disability-adjusted life expectancy: Global Burden of Disease Study** *Lancet* 1997, **349:**1347-1352
2. Murray CJ and Lopez AD **Quantifying disability: data, methods and results** *Bull World Health Organ* 1994, **72:**481-494
3. Murray CJ and Lopez AD **Global and regional descriptive epidemiology of disability: incidence, prevalence, health expectancies and years lived with disability** *The Global Burden of Disease (Edited by: Murray CJ, Lopez AD) Boston: Harvard School of Public Health* 1996, 201-246
4. Barendregt JJ, Baan CA and Bonneux L **An indirect estimate of the incidence of non-insulin-dependent diabetes mellitus** *Epidemiology* 2000, **11:**274-279
5. Barendregt JJ, Bonneux L and van der Maas PJ **The health care costs of smoking** *N Eng J Med* 1997, **337:**1052-1057
6. Barendregt JJ, van Oortmarssen GJ, van Hout BA, van den Bosch JM and Bonneux L **Coping with multiple morbidity in a life table** *Mathematical Population Studies* 1998, **7:**29-49
7. Char B, Geddes K, Gonnet G, Leong B, Monagan M and Watt S **First Leaves, a Tutorial Introduction to Maple V** *New York: Springer-Verlag* 1991,
8. Press WH, Teukolsky SA, Vetterling WT and Flannery BP **Numerical Recipes in Fortran 77: The Art of Scientific Computing** *Cambridge: Cambridge University Press* 1992,
9. Efron B and Tibshirani RJ **An introduction to the bootstrap** *London: Chapman & Hall Inc* 1993,
10. Victorian Government Department of Human Services, Public Health and Development Division **The Victorian Burden of Disease Study: Mortality** *Melbourne* 1999,
11. Victorian Government Department of Human Services, Public Health and Development Division **The Victorian Burden of Disease Study: Morbidity** *Melbourne* 1999,
12. Kruijshaar ME, Barendregt JJ and Hoeymans N **The use of models in the estimation of disease epidemiology** *Bull World Health Organ* 2002, **80:**622-628