Public Health Monograph Series
No. 35
ISSN 1178-7139

COST OFF-SETS USED IN BODE^3 MULTISTATE LIFETABLE MODELS

Burden of Disease Epidemiology, Equity and Cost-Effectiveness Programme (BODE^3)

Technical Report: Number 15

Giorgi Kvizhinadze
Nhung Nghiem
June Atkinson
Tony Blakely

July 2016

A technical report published by the Department of Public Health,
University of Otago, Wellington

ISBN 978-0-9876663-9-0

Acknowledgements
We thank other BODE^3 team colleagues for comments on early versions of this work. This Programme receives funding support from the Health Research Council of New Zealand.

Further details can be found at www.uow.otago.ac.nz/bode3-info.html .

Competing Interests
The authors have no competing interests.

Table of Contents

List of Tables

Introduction

This Technical Report outlines the method for using linked New Zealand health data, with unit costs attached for most events, to generate costs for the BODE[3] multistate lifetables (MSLTs). These MSLTs are macrosimulation models, with parallel proportions of the cohort (overlapping; independent probabilities) being in multiple disease states. To harness the rich New Zealand data on costs, for cost offsets in the MSLTs, we calculate a cost for every New Zealander of a given sex and age that is the 'base cost', for people neither in the last six months of life nor with any 'model-related diseases'. (Regarding the latter, for the Tobacco MSLT this would be no tobacco-related diseases.) Then excess costs for those in the last six months of life, first year of diagnosis, and prevalent disease are estimated (see below for more details). It should be noted that this costing schema is designed to both 'suit' the MSLTs, and harness the rich New Zealand health data.

This Technical Report was prepared in July 2016, and represents a consolidation of methods used to that point. Further information on the data used in the costing can be found elsewhere.[1-3] Of note, the actual costs will continue to 'evolve' or 'improve' as the underlying data is improved. However, the method below should be constant in the future.

Disease cost inputs

We calculated costs such as 'the health system costs for males age 57 in their first year post diagnosis with a stroke', and more generally these five costs:

A. Annual cost to the New Zealand health system for being alive for a given sex and age, and *not* in the last six months of life and *not* concurrently alive with one of the modeled diseases (i.e. tobacco-related in this Report). All members of the cohort alive (or dying in this cycle) are assigned (half) this cost; it is the base cost.
B. Excess cost to A, for being in the last six months of life if dying of a disease *other than* one of the modeled diseases (i.e. dying of a non-tobacco-related disease).
C. Excess cost to A for being in first year of diagnosis of a tobacco-related disease.
D. Excess cost to A for being alive with a given disease, and neither in the first year of diagnosis nor in the last six months of life if dying of that tobacco-related disease.
E. Excess cost to A for being in the last six months of life if dying of the tobacco-related disease.

To calculate above five streams of costs we took the following steps outlined below, and as per structure shown in Table 1:

1. Extract from HealthTracker (i.e. linked New Zealand health data, with event costs merged) total costs and person time (months) not in the last 6 months of life for everybody, regardless of their cause of death – group All_0 – and total costs and person times in the last 6 months of life for everybody, regardless of their disease – group All_6.
2. From the group All_0 extract total costs and person times for diseases of interest for:
    a. first year of diagnosis - group total costs and person times, $C_{i,C}$ and $C_{i,PT}$, for diseases of interest for being alive with a given disease i,
    b. not in the first year of diagnosis - group $D_{i,C}$ and $D_{i,PT}$

c. total costs and person times for everybody without the modeled diseases - group A_0.

Note that $C_{i,C}$ and $D_{i,C}$ for given disease i is not mutually of the same entities for other values of i, i.e. diseases are not mutually exclusive. But A_0 is mutually exclusive of B, C, D and E costs.

3. From the group All_6 extract total costs and person times for diseases of interest E_{6i} of group E_6 and total costs and person times for not diseases of interest group B_6. No person can contribute to more than one of the E_6 and B_6 – they are mutually exclusive as one person can only die of one disease.

Table 1: Data extraction structure to subset observations, with attendant costs and person times

Step 1	Step 2	Output
Extract not last 6 months of life - group All_0	First year Disease 1	group C_0
	First year Disease 2	
	First year Disease i	
	First year Disease n (where n= total number of tobacco-related diseases)	
	Subsequent years (i.e. prevalent) Disease 1	group D_0
	Subsequent years (i.e. prevalent)Disease 2	
	Subsequent years (i.e. prevalent)Disease i	
	Subsequent years (i.e. prevalent)Disease n	
	Pre-first diagnosis of any Disease i	group A_0
	Post-maximum cure time of any Disease i	
	Step 3	
Last 6 months of life-group All_6	Last 6 months of life died from Disease 1	group E_6
	Last 6 months of life died from Disease 2	
	Last 6 months of life died from Disease i	
	Last 6 months of life died from Disease n	
	Last 6 months of life died from non-tobacco-related causes causes	group B_6

The A costs per month are calculated simply by taking the ratio of total costs over total person times in group A_0.

$$A = \frac{A_{0,C}}{A_{0,PT}}$$

The B costs take the ratio of total costs over total person time in group B_6, then subtract off the A cost so as to give an 'excess cost', namely

$$B = \frac{B_{6,C}}{B_{6,PT}} - A \quad (1)$$

The E costs take the ratio of total costs over total person time in group E_6 ,and then subtract A, for each disease i, namely:

$$E_i = \frac{E_{i,C}}{E_{i,PT}} - A \quad (2)$$

Importantly, the excess disease costs (C and D costs) are <u>not</u> for mutually exclusive groups of people. That is, when calculating the excess costs to A for people in their first year of diagnosis with CHD, some of these people may also have new/prevalent other disease such as stroke and cancer. If we do not adjust for this, the model will overestimate the total health system cost. (This is very similar to the need to identify separately and independently the comorbidity of each disease.)

We first calculate for each sex by age group a scalar for person time for C and D costs jointly. The scalar is the ratio of 'true' person time for people in C and D groups, divided by the 'naïve' overestimated person time for C and D is one just sums C and D person time for each disease separately considered. Namely:

$$s_{PT} = \frac{(All_{0,PT} - A_{0,PT})}{\sum_1^n (C_{i,PT} + D_{i,PT})}$$

Where:

$All_{0,PT}$ = the total 'true' person time not within six months of death

$A_{0,PT}$ = the total person time not within six months of death for those with no tobacco-related diseases

$C_{i,PT}$ = the person time not within six months of death for first year of diagnosis of disease i, with n diseases in total

$D_{i,PT}$ = the person time not within six months not within first year of diagnosis (i.e. 'prevalent disease) for disease i, with n diseases in total.

For example, if there was 20,0000 person years in total not within the last six months of life, of which 10,000 was person time free of any tobacco-related disease, the numerator would be 20,000 – 10,000 = 10,000. And the sum across each tobacco-related disease considered separately of all person time not within six months of death was 15,000 (due to many people with two or more diseases simultaneously, 15,000 > 10,000). Thus the scalar is 10,000/15,000 = 0.667. This scalar applied to all C and D person time will make then comparable to 'unique' individuals.

Next, we estimate the total 'excess' costs for C and D groups, denoted with $C_{tot\_ex,C}$ and $D_{tot\_ex,C}$ respectively. At this point, it is important to note that the MSLT (where these costs are destined as inputs) assigns A costs to the 'alive' proportion of the cohort in the main lifetable, ***not*** within each disease. Given that some of the cohort may reside in two or more disease states, the unique number of individuals contributing person time to the disease states will be less than the naïve sum of person-time across disease states – by an amount given by S_{PT} above. Therefore, to remove the 'A costs' to give the total excess costs requires subtracting off S_{PT}-scaled A costs, namely:

$$C_{tot\_ex,C} = All_{0,C} - S_{PT} \times A \times C_{PT} \quad (3)$$

for all diseases and

$$C_{i,tot\_ex,C} = C_{i,C} - S_{PT} \times A \times C_{i,PT}$$

by disease. And likewise for D costs.

$$D_{tot\_ex,C} = All_{0,C} - S_{PT} \times A \times D_{PT} \quad (4)$$

and

$$D_{i,tot\_ex,C} = D_{i,C} - S_{PT} \times A \times D_{i,PT}$$

A scalar was then calculated for the **excess** total costs, namely:

$$s_C = \frac{(All_{0,C} - A \times All_{0,PT})}{C_{tot\_ex} + D_{tot\_ex}} \quad (5).$$

There were two final steps to estimate each of the excess C and D costs. It is useful to keep in mind how these costs will be used in the MSLT. Namely, they will be costs scaled down for comorbidity, excess to the A costs for the number of person years for unique individuals (not total person time which includes duplicated individuals).

The first step is estimating the total excess costs by disease i and phase (C,D), scaled down for comorbidity:

$s_c \times C_{i,tot\_ex,C}$, $s_c \times D_{i,tot\_ex,C}$

Second, this all needs to be divided by the observed person time: $C_{i,PT}$.

Accordingly, the monthly cost by disease and phase can be calculated as:

$C_i = \frac{s_c \times C_{i,tot\_ex,C}}{C_{i,PT}}$ (6)

Likewise, the D costs are:

$D_i = \frac{s_c \times D_{i,tot\_ex,C}}{D_{i,PT}}$ (7)

(Workings and spreadsheets are available from the authors on request.)

Logic check

As a high-level calibration/logic check, we checked that the above method returns all costs observed for total population. The MSLT model cost assignment requires the following equation to hold:

$A \times All_{PT} + \sum_{i=1}^{n}[IR_i \times All_{PT} \times C_i + PR_i \times All_{PT} \times D_i] + B \times All_{PT} \times p_{OC} + \sum_{i=1}^{n} E_i \times All_{PT} \times p_i = All_{0,C} + All_{6,C}$ (8)

Where IR_i and PR_i are incidence and prevalence rates of disease i, p_{OC} is probability of death from diseases not modelled and p_i is probability of death from i disease

Assuming $All_{PT} = All_{0,PT} + All_{6,PT}$, we can write equation (8) as follows:

$$A \times All_{0,PT} + A \times All_{6,PT} + \sum_{i=1}^{n}[IR_i \times All_{PT} \times C_i + PR_i \times All_{PT} \times D_i] + B \times All_{PT} \times p_{OC} + \sum_{i=1}^{n} E_i \times All_{PT} \times p_i = All_{0,C} + All_{6,C}$$

Noting that, $\times All_{6,PT} = A \times All_{PT} \times (p_{OC} + \sum_{i=1}^{n} p_i)$, we will prove (8) by showing that

$A \times All_{0,PT} + \sum_{i=1}^{n}[IR_i \times All_{PT} \times C_i + PR_i \times All_{PT} \times D_i] = All_{0,C}$ (9)

and

$B \times All_{PT} \times p_{OC} + \sum_{i=1}^{n} E_i \times All_{PT} \times p_i = All_{6,C} - A \times All_{PT} \times (p_{OC} + \sum_{i=1}^{n} p_i)$ (10)

Equation (9) can be written as:

$$\sum_{i=1}^{n}\left[C_{i,PT} \times C_i + D_{i,PT} \times D_i\right] = All_{0,C} - A \times All_{0,PT}$$

Substituting C_i and D_i with (6) and (7) respectively, we obtain

$$s_C \sum_{i=1}^{n}\left[C_{i,tot\_ex} + D_{i,tot\_ex}\right] = All_{0,C} - A \times All_{0,PT}$$

Which is equivalent to equation (5), hence equation (9) holds.

For equation (10), since

$$B \times All_{PT} \times p_{OC} \approx B \times B_{6,PT} = B_{6,C} - A \times B_{6,PT}$$

following from (1)

$$\sum_{i=1}^{n} E_i \times All_{PT} \times p_i \approx \sum_{i=1}^{n} E_i \times E_{i,PT} = \sum_{i=1}^{n}(E_{i,C} - A \times E_{i,PT})$$

from (2)

and

$$All_{6,C} = B_{6,C} + \sum_{i=1}^{n} E_{i,C}$$

the equation (10) can be written as follows:

$$B_{6,C} - A \times B_{6,PT} + \sum_{i=1}^{n}(E_{i,C} - A \times E_{i,PT}) = B_{6,C} + \sum_{i=1}^{n} E_{i,C} - A \times B_{6,PT} - A \times \sum_{i=1}^{n} E_{i,PT}$$

This completes the proof of equation (10), hence equation (8) holds.

References

1. Blakely, T., et al., Updated New Zealand health system cost estimates from health events by sex, age and proximity to death: further improvements in the age of ‘big data’ N Z Med J, 2015. **128** (1422): p. 13-23.

2. Blakely, T., et al., Health system costs by sex, age and proximity to death, and implications for estimation of future expenditure NZMJ, 2014. **127**(1393): p. 12-25.

3. Blakely, T., et al., Patterns of Cancer Care Costs in a Country With Detailed Individual Data. Medical Care, 2015. **53**(4): p. 302-309.