Public Health Monograph Series
No.30
ISSN 1178-7139

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

Version 2.1

Burden of Disease Epidemiology, Equity and Cost-Effectiveness Programme

Technical Report: Number 3

Principal Authors:
Tony Blakely
Rachel Foster
Nick Wilson

BODE^3 Co-Investigators and contributing authors:
Roy Costilla, Martin Tobias, Diana Sarfati, Des O'Dea, Melissa McLeod, Matthew Soeberg, David Hadorn, Ken Richardson, Andrew Simpson, June Atkinson, Giorgi Kvizhinadze, Nhung Nghiem, Theo Vos, Jan Barendregt, Linda Cobiac, Lucie Collinson, Amber Pearson

December 2012

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

ISBN 978-0-9876663-5-2

Recommended citation:

Blakely T, Foster R, Wilson N, and BODE^3 Team. *Burden of Disease Epidemiology, Equity and Cost-Effectiveness (BODE^3) Study Protocol.* Version 2.0. Wellington: Department of Public Health, University of Otago, Wellington, December 2012.

Acknowledgements

The Burden of Disease Epidemiology, Equity and Cost-Effectiveness (BODE^3) Programme is funded by the Health Research Council of New Zealand (10/248).

BODE^3 is conducted in collaboration with:

- the University of Queensland Burden of Disease and Cost-effectiveness Centre, more specifically the ACE-Prevention (Australia) Programme.
- the Ministry of Health, particularly:
  - the Burden of Disease study
  - HealthTracker
  - numerous senior advisors

Abbreviations

ABC-CBA	Aotearoa Burden of Cancer and Comparative Benefit Assessment study
ACC	Accident Compensation Corporation
ACE	Assessing Cost-Effectiveness
AIHW	Australian Institute of Health and Welfare
BDS	Burden of Disease study (NZ for 2006 or 2011; or Australia)
BMI	Body Mass Index
BODE^3	Burden of Disease Epidemiology, Equity and Cost-Effectiveness Programme
CHHP	Community Heart Health Programmes
CI	Confidence Interval
CIAG	Cancer Interventions Advisory Group
CPI	Consumer Price Index
CRA	Comparative Risk Assessment
CVD	Cardiovascular disease
DALY	Disability-adjusted Life Year
DCIS	Disease Costs and Impacts Study
DES	Discrete Event Simulation
DHB	District Health Board
DRG	Diagnosis-Related Group
DW	Disability Weight
GBD	Global Burden of Disease (usually the 2010 version)
GCEA	Generalised Cost-Effectiveness Analysis
GP	General Practitioner
GST	Goods and Services Tax
HALY	Health-Adjusted Life Year
HRC	Health Research Council (of New Zealand)
HTA	Health Technology Assessments
ICD	International Classification of Diseases
ICER	Incremental Cost-Effectiveness Ratio
NGO	Non-Governmental Organisation
NHB	National Health Board
NHI	National Health Index
NHL	Non-Hodgkin's Lymphoma
NICE	UK National Institute for Health and Clinical Excellence
NRT	Nicotine Replacement Therapy
NZACE	New Zealand Assessment Cost-Effectiveness Prevention Study
NZCMS	New Zealand Census-Mortality Study
NZHIS	New Zealand Health Information Services
PAG	Programme Advisory Group
PHO	Primary Health Organisation
PSA	Prostate Specific Antigen, or Probabilistic Sensitivity Analysis [depending on context]
QALY	Quality Adjusted Life Year
RR	Rate Ratio

SIGN	Scottish Intercollegiate Guidelines Network
RCT	Randomised Controlled Trial
WHO	World Health Organization
YLDs	Years of Life in Disability
YLLs	Years of Life Lost

Glossary

@RISK	Risk analysis software from Palisade that works with Microsoft Excel for Windows. Used in most ACE-Prevention (Australia) models for all uncertainty analyses. The BODE^3 Team has used @RISK or **Ersatz** (see entry) for some uncertainty analyses in NZACE-Prevention.
Average cost-effectiveness	The ratio of difference in costs for intervention compared to base-case (e.g., partial null), to the difference in utility or health consequences (e.g., QALYs, **HALYs** gained) between intervention and base-case. The slope of the line from the origin to the interventions x-y coordinate on a cost-effectiveness plane.
Bottom-up costing	Bottom-up costing means determining the number of units of a particular resource used, then multiplying it by the unit cost for that resource, and then aggregating.
Cost-effectiveness expansion pathway	The curve on a cost-effectiveness plane formed by joining the origin to the intervention with the lowest average cost-effectiveness, then the next lowest, and so on. Any intervention above and to the northwest of the curve is "dominated".
Cost offsets	These are future health system costs incurred or averted by an intervention that prevents, or reduces severity of, disease in the future or prolongs life that would not have been incurred/ averted without the intervention under consideration. (Using Drummond et als' framework[1], these are S1 (and usually S3) costs). Sum of all future disease and population health system costs (see **Costs: health system**; may be negative) that are altered due to the epidemiological impacts of the interventions (e.g. reducing future disease incidence, curing people who live longer). These costs may be either positive (additional costs incurred) or negative (cost savings due to costs averted), but for simplicity the term cost offsets is used for both scenarios.
Costs: health system	For the purposes of this Protocol and BODE^3 analyses, costs <u>other</u> than the direct/intervention costs that 'routinely' occur for people with given disease or wellness states. S1 and S3 costs using Drummond's framework. In BODE^3 linked health administrative datasets held by the Ministry of Health (i.e. 'HealthTracker') will generally be the primary source of health system costs. In BODE^3 it is useful to conceptualise four types: 1. **Disease health system cost for those not within 6 months of death**. Costs per unit person-time for people with the

	disease(s) directly addressed by the intervention under assessment. For cancers, these are further broken down by month post diagnosis, often to be coincident with each cancer's assumed disease model times (i.e. number of months in diagnosis and treatment, remission, etc). 2. **Disease health system cost for those within 6 months of death from the given disease**. Costs per unit person-time for people about to die from the disease(s) directly addressed by the intervention under assessment. For cancers, the six month period may be changed to match the pre-terminal and terminal phase durations of a given type of cancer. 3. **Population health system cost for those not within 6 months of death of any cause.** The average cost of a citizen of a given sex and age. Note that these costs are sometimes referred to as 'unrelated health system costs'. 4. **Population health system cost for those within 6 months of death of any cause.** The average cost of a citizen of a given sex and age, within six months of death from any cause. Note that the disease health system costs include all health system costs for someone with a given disease, which includes the unrelated health system costs (see **costs: unrelated**) of 'average' comorbidities for someone of the same sex and age. Thus, if modelling (many) common diseases, the assumption that population health system costs apply to those without the diseases under assessment may become implausible, requiring a 'non-diseased' cost to be estimated rather than just using the population-wide sex/age average (e.g., for heart disease).
Costs: intervention	The costs incurred directly by the intervention under assessment; for instance, those related to implementing and monitoring the intervention, and in some cases also the set-up costs. Where the comparator is a partial null or current standard of care, these costs will be incremental to the comparator. For example, the intervention costs of patient navigators is not just that of patient navigators, but the difference in costs for patient navigators compared to whatever pre-existing coordination services were in place in the current standard of care comparator. Where two interventions are directly compared, the cost of *each* intervention is calculated. In some cases, the direct cost of each intervention (e.g. different chemotherapy regimens) will be only those costs that are additional to a partial null or standard care costs that are the same regardless of the intervention applied and thus do not need to be costed (e.g. care of cancer patients

	that is not affected by the choice of chemotherapy). The cost difference *between* the two interventions informs the incremental cost-effectiveness ratio.
Costs: unrelated	Costs (usually) to the health system (i.e. S1 and S3 per Drummond's framework) that are unrelated to the disease under question. Some economic decision models exclude unrelated health care costs. BODE^3 includes them for two reasons: it is usually impossible to clearly define which costs are, and are not, attributable to the disease(s) in question, and; the BODE^3 perspective is life-long and health system wide. (See **Costs: health system**).
Cure time	(See **statistical cure time**).
Disability adjusted life year (DALY)	A type of health adjusted life year (HALY) used in burden of disease studies that assess the cross-sectional or prevalent burden of disease. The sum of years of life lost (YLLs) and years of life lived with disability (YLDs). The health metric terminology used in BODE^3 is that of HALYs. (See **health adjusted life years**).
Ersatz	Ersatz is an uncertainty and bootstrap add-in for Microsoft Excel, developed by Jan Barendregt (EpiGear International Pty Ltd). The BODE^3 Team may use this instead of **@RISK** (see entry) in future work. See: www.epigear.com/
Excess mortality rate (EMR)	The excess mortality for a given disease (usually from a given cancer) compared to expected population mortality (i.e. that from life tables). Cancer EMRs in BODE^3 are determined using cancer registry data linked to mortality data, life tables, and Poisson regression models. They are used to parameterise economic decision models.
Excess health system cost	The excess health system cost for a given disease compared to the expected population health system cost (see **costs: health system**).
Health adjusted life years (HALYs)	The remaining expected life expectancy, weighted for quality of life or health status. Quality adjusted life years (QALYs) are one form of HALYs. The HALYs used in BODE^3 are the same in principle as a QALY, except: **disability weights** (DWs) rather than utilities are used to adjust for quality of life; the maximum HALYs that can be awarded for any given sex by age group is not 1.0, but rather 1 minus the **population morbidity** (or **pYLD**). **DALYs** are also a form of HALY, but are usually reserved for cross-sectional or prevalent quantification of a population's burden of disease; economic decision modelling requires prospective modelling of HALYs, streamed into the future. Thus, to avoid confusion with DALYs as measured in a burden of

	disease study, we use the term 'HALYs gained' rather than 'DALYs averted'. DALYs also tend to measure years of life lost (YLLs) against an external or model life table, not the population's own life table or life expectancy as for HALYs in economic decision modelling.
Heterogeneity	Heterogeneity is where costs, consequences and cost-effectiveness may vary by 'fixed' socio-demographics such as sex, age, ethnicity and socioeconomic position, and other 'fixed' population characteristics (e.g., cancer stage). It differs from uncertainty, where there is uncertainty in estimates due to one of model structure, input parameter and stochastic uncertainty. Heterogeneity is of direct interest. For example, varying cost-effectiveness by age is critical to inform decision making. Also, and importantly in BODE^3, it is likely that ethnic and socioeconomic equity considerations will be largely understood through heterogeneity on these variables.
Incremental cost	The difference in cost for intervention A compared with intervention B. When calculating the ICER, the difference in **net cost** between the two interventions is used.
Incremental cost-effectiveness ratio (ICER)	The ratio of difference in costs for intervention A compared to intervention B, to the difference in utility or health gain (e.g., QALY, HALYs gained) between treatment A and treatment B. The slope of the line from intervention A's to intervention B's coordinates on a cost-effectiveness plane (assuming that intervention B's consequences and costs have been calculated as though intervention A has already been applied).
Input parameter uncertainty	(Also see entry under "**Uncertainty"**) One of three forms of uncertainty in economic decision modelling. It arises due to uncertainty in the input parameters to the model. For example, there may be uncertainty about the benefit of a new treatment, often expressed as a confidence interval about an effect estimate such as a rate ratio. Of note, the confidence interval captures random error about the input parameter; often there will be likely residual systematic error (ie, confounding, selection or information biases that are thought to still be present in systematic reviews, meta-analyses, etc). BODE^3 will on occasion attempt to include combined estimates of random and systematic error about input parameters, using quantitative bias analysis techniques (in addition to usual methods for measuring random error).[2 3] There are well specified distributional forms for input parameter uncertainty – at least the random error component. For

	example, log normal for rate ratios, gamma for costs, etc.[2 4]
Input parameter uncertainty analysis ("uncertainty analysis" for short)	Input parameter uncertainty effects on total cost, health consequence and cost-effectiveness uncertainty will be quantified using Monte Carlo simulations, with each iteration selecting values from the distribution about each input parameter with probability proportional to the density function, using distributional forms that best capture **input parameter uncertainty**.
Intervention pathway	(See **Cost-effectiveness expansion pathway**)
Macro-costing	Macro-costing uses cost estimates for units of input and output that are large relative to the intervention being analysed.[5] For example, macro-costing uses cost estimates for hospital stays or doctor visits rather than for the procedures and professional time expended during these encounters.
Macro-simulation	A mathematical or economic decision model of states of groups of people (cohorts) to estimate cost effectiveness. Often a Markov state-transition model. NZACE-Prevention makes extensive use of multistate lifetables. Macro-simulation can incorporate and estimate outputs by **heterogeneity** and **parameter uncertainty**. But it neither includes nor estimates individual-level or **stochastic uncertainty.**
Micro-costing	Micro-costing requires the direct enumeration and costing out of every input consumed in the treatment of a particular patient. [5] Micro-costing is laborious to implement.
Micro-simulation	A mathematical or economic decision model of individuals to estimate cost effectiveness. Incorporates and estimate outputs by **heterogeneity, parameter uncertainty** and **stochastic uncertainty**. More computationally demanding than macro-simulation, but often required or beneficial in the presence of population heterogeneity, complexity, future costs/epidemiology contingent on past experience (i.e. memory required), interactions with other subjects, etc. Requires Monte Carlo simulation. There are many types of micro-simulation, including: • Markov micro-simulation, where individuals (as opposed to cohorts) are sampled thousands of times and randomly walk through a Markov structure (with the advantage of memory retained) • Discrete event simulation, where time to event rather than (usually) state transitions are modelled. • Dynamic and other models that allow interactions (e.g. infectious diseases modelling).
Model structure	(See entry under **Uncertainty**).

uncertainty	
Net cost	[Costs: intervention] + [Cost offsets] The final net cost, including both intervention and health system costs. The latter may be either costs incurred or costs averted (cost savings). This should not be confused with **incremental cost**, where the net cost of one intervention is compared with the net cost of another intervention.
Partial Null	The 'base-case' scenario against which proposed intervention costs and health consequences are compared. Often it is simply our best estimate of the current (and projected into future) state of disease incidence and prevalence, preventive programmes, treatment coverage, etc. However, if we are to undertake economic decision modelling about interventions that are currently in place, or responsible for some of the projected future 'business as usual' scenario, we need to remove the current/projected effects and cost of these interventions from the base-case scenario. For example, in a comparison of cardiovascular disease prevention and treatment programmes that includes comparing current practice with cholesterol lowering drugs with alternative use of thiazide diuretics, a new 'null' base-case scenario needs to be constructed. Usually it is a 'partial null' as the costs and health consequences of current/projected interventions that impact only the domain of interest are stripped out of the base-case model – not the costs and consequences of all health system interventions.[6]
Population morbidity	The average or expected level of population morbidity, by sex and age. The average of all years of life lived with disability (YLDs) at a given sex and age; sometimes referred to as prevalent YLDs (**pYLDs**). Alternatively, the weighted average disability weight (DW) at a given sex and age, where the weights are the prevalence of disease (or combinations of diseases). These pYLDs give the envelope of total possible health adjusted life yeas (**HALYs**). For example, if the pYLD is 0.2 at older ages, the maximum possible HALY for a year of life at that age is 1 – 0.2 = 0.8, not 1.0.
Probabilistic analysis	Probabilistic analysis is used in **input parameter uncertainty analysis** in BODE^3. To capture the full uncertainty of an important input parameter, a distribution around the best estimate value is usually defined. The uncertainty is then propagated through the model using simulation techniques (see **input parameter uncertainty analysis**).

Probabilistic sensitivity analyses (PSA)	(See entry under **Sensitivity analysis**).
pYLD	Prevalent years of life lived with disability (See entry under **Population morbidity**).
Scenario analysis	A type of sensitivity analysis (see entry under **Sensitivity analysis**). In BODE^3 the term scenario analysis is reserved for input parameters or model structure when there is insufficient information to specify a probabilistic distributions, but it is still meaningful (if not essential) to conduct 'what if' analyses. For example, one might rerun a model assuming the effect of an intervention wanes to zero in 10 years, as opposed to persisting for life. Alternatively, one might determine what amount of waning of the intervention effect is necessary to render the intervention cost-ineffective. There is overlap between input parameter and model structure uncertainty analysis.
Sensitivity analysis[a]	Sensitivity analysis has a variety of meanings. In BODE^3, we use this term to refer to changes or 'sensitivity' of modelling outputs (e.g., HALYs gained, costs, ICER) to model structure and inputs. There are two important types of sensitivity analysis: 1. **Sensitivity to model structure assumptions**. Here sensitivity analysis is the testing of scenarios around variations in key design features and structural assumptions of the model (e.g., varying the discount rate, adding states for relapse, including or excluding certain costs [although the latter two might be considered model extensions – but there is a grey zone between what might be called sensitivity analyses and model extensions/variations]). 2. **Sensitivity to input parameter uncertainty**. Here sensitivity analysis is the determination of changes in the HALYs gained, costs or ICER arising from **input parameter uncertainty**. Results of input parameter uncertainty sensitivity analyses are often shown as a Tornado plot. There are two common ways to quantify this: a. **One-way (input parameter) sensitivity analyses**. Determination of sensitivity or change in output variable for a meaningful unit change (e.g., 1 standard deviation) in each input variable, where each sensitivity analysis is for only one variable

[a] Terminology around sensitivity analysis is not uniform across practitioners. The terms and definitions used in this Glossary are influenced by schema in Briggs et al (2006) and that used by Jan Barendregt (course teaching notes).

	at a time. b. **Probabilistic sensitivity analyses**. Probabilistic analyses must be undertaken first, generating multiple sampled values of input parameters and multiple values of output variables across all iterations. The probabilistic sensitivity analysis is then the correlation between each input parameter (that has uncertainty and sampling from a probability distribution across iterations) and the output parameters. It will differ from one-way sensitivity if there are correlations between parameters in their impact on outputs. (Also see **Scenario analysis** entry).
Statistical cure time	The time post diagnosis when excess mortality has (essentially) reached zero, or the relative survival curve has (essentially) become horizontal. That is, death from the cancer is negligible, and all survivors are assumed to be cured.
Stochastic uncertainty	(See entry under **Uncertainty**).
Top-down costing	Top-down costing is perhaps best thought of in terms of aggregate Vote:Health (and private) outlays, broken down by main expenditure categories or for a specific condition. This approach is appropriate where it is not important to know with precision what the cost drivers are. An advantage of top-down costing is that one may know the total health expenditure (e.g., Vote-Health plus total private health expenditure from survey data), giving a known 'envelope' to work within.
Uncertainty	There is often considerable uncertainty in costs, consequences and cost-effectiveness estimates. We identify three sources of this uncertainty: 1. **Model structure uncertainty** arises due to uncertainty about the assumptions implicit in setting the model structure. For example, the disease model, discount rates, etc. In BODE^3, model structure uncertainty will be assessed by **sensitivity analysis** about given alternatives. 2. **Input parameter uncertainty** arises due to uncertainty in the true value of parameters input into the model, such as uncertainty about the consequences of a new treatment. Within BODE^3, **"input parameter uncertainty analysis"** (and on occasion **scenario analysis**) will be used to address this source of uncertainty. 3. **Stochastic uncertainty** arises due to random variability in individual experiences (e.g., time to event). We will not usually be directly concerned with stochastic uncertainty, as either: the macro-simulation models used in BODE^3 are estimating population-level expectations with input parameter uncertainty

	capturing likely variation about mean estimates, or; micro-simulation models when used allow for heterogeneity and other complexities, rather than describing individual-level variation per se.
Uncertainty analysis	Abbreviation for "**input parameter uncertainty analysis**".

Table of Contents

List of Tables

List of Figures

Executive Summary

This Protocol is for the Burden of Disease Epidemiology, Equity and Cost-Effectiveness programme (BODE^3), funded by the Health Research Council of New Zealand (HRC) for the period 2010-15. The aim of the Burden of Disease Epidemiology, Equity & Cost-Effectiveness Programme (BODE^3) is:

> To build capacity and academic rigour in New Zealand in the estimation of disease burden, cost-effectiveness and equity impacts of proposed interventions, and undertake a range of such assessments.

The objectives are:

1. To estimate the impact (total & equity-related) and cost-effectiveness of cancer control interventions using Markov time dependent macro- and micro-simulation models and discrete event simulation (Aotearoa Burden of Cancer and Comparative Benefit Assessment study; ABC-CBA).
2. To estimate the impact (total & equity-related) and cost-effectiveness of preventive interventions using multistate lifetables and micro-simulation models (NZ-Assessing Cost-Effectiveness: Prevention; NZACE-Prevention).
3. To build capacity and academic rigour in disease, intervention, equity, uncertainty and cost-effectiveness modelling.

This report is Version 2.0 of the BODE^3 Protocol, published in December 2012. The Protocol will periodically be updated to include on-going methodological developments.

The protocol is in three Parts. **Part I** describes the overall principles that apply to the whole programme. These include:

- A general approach that uses burden of disease methods. This means that epidemiological models across interventions will be comparable, as they share similarly derived epidemiological parameters.
- Emphasis on a health system perspective, in that consequences (health adjusted life years gained (HALYs)) and costs are health-related and (usually) health system costs.
- A focus on allocative efficiency in economic evaluation.
- The use of probabilistic analysis methods to model input parameter uncertainty (e.g., uncertainty around costs and intervention effectiveness).
- A focus on equity analyses, by which we mean the comparative impact (costs, health consequences and cost-effectiveness) of interventions by ethnicity and socioeconomic position (and by sex and age). A key facilitator of such a focus is the rich New Zealand data by ethnicity and socioeconomic position that allows heterogeneity to be specified and modelled.
- The processes that BODE^3 will use to:
    - select interventions to model.

  - assess the likely health consequences, and uncertainty, of the selected interventions (e.g., evidence hierarchies, review and expert consensus methodologies).
  - assess the likely costs of interventions, including the direct 'upfront' costs of the intervention to the health sector (and occasionally more widely) and the cost-offsets that may be incurred or accrued in the future by preventing death and prolonging life.
- The modelling approaches to be used in BODE^3, namely:
  - multistate lifetables in NZACE-Prevention.
  - Markov models and discrete event simulation (DES) in both NZACE-Prevention and ABC-CBA.
- The minimum outputs that BODE^3 will generate:
  - total net cost of the intervention, with uncertainty.
  - total HALYs gained by the intervention, with uncertainty.
  - estimates of cost-effectiveness, with uncertainty.
  - a description of the relative magnitude of disease burden (both DALY and (where relevant) comparative risk assessment (CRA) output) that the intervention is addressing, capitalising on the burden of disease study (BDS) underpinnings of BODE^3.
  - and estimates of ethnic and socioeconomic equity impacts.

Part II provides detail specific to the ABC-CBA methods and models. ABC-CBA gathers together a wide range of New Zealand and international data on cancer incidence, survival, disease models and quality of life in the form of large and heterogeneous (by cancer-site, sex, age, ethnicity, deprivation and sometimes stage/sub-type of cancer) models (Markov or discrete event simulation (DES)) that represent the business as usual scenario for incident cancers in 2011. This baseline model is extended out into the future for incident cases in years 2012 and beyond. Laid over the states of the model are disability weights and average health system costs for being in that state (the latter being sourced from Ministry of Health "HealthTracker" data if feasible).

Having developed these large and heterogeneous baseline models, cancer control interventions are then modelled by changing key parameters from the baseline model for sub-populations of interest. That is:

- treatment interventions are modelled by changing cancer mortality and disability weight parameters (and possibly time spent in various states), and the occurrence of sequelae.
- preventive interventions are modelled by changes in the future incidence rate.
- screening and early diagnosis interventions are modelled by changes in the stage or severity distribution at diagnosis, and incidence rates themselves.
- palliative and supportive care interventions are modelled by changes in disability weights.

It is assumed that many (if not most) interventions can be modelled by extracting portions of the baseline Markov cancer model and specifying input parameters (and uncertainty) directly to this model structure. However, it is also envisaged that:

- Disease models will need specific modification for some interventions (e.g., the addition for serial relapse and remission stages if the treatment has varying effectiveness dependent on previous history of relapses (i.e., adding more 'memory' to the Markov model, or additional competing events in DES)).
- Due to either computationally over-whelming heterogeneity (i.e., just too many states for probabilistic analyses and simulation), or non-linearity in the model (e.g., aggressiveness of tumour inversely proportional to time since last 'clear' screening test, where that relationship is not efficiently captured by adding further Markov states), discrete event simulation (DES) or other microsimulation modelling approaches may be used. An over-riding principle, however, of ensuring comparability of epidemiological and cost parameters with the baseline data will be adhered to.

Part III provides summary detail on the NZACE-Prevention model. The NZACE-Prevention model is largely based on that used in ACE-Prevention in Australia, which has its existing study protocol.[7 8] Therefore, Part III of the protocol is brief, focusing on issues such as the adaptation of multi-state lifetables to the New Zealand setting, and incorporating output from the parallel NZ BDS revision. Given the established nature of ACE methodology, specific pieces of additional work have already been undertaken for NZACE-Prevention and are published separately on the BODE^3 website (www.uow.otago.ac.nz/BODE3-info.html), namely reports detailing options for selecting risk factors, then interventions per se, to model in NZACE-Prevention.[9 10] Of note is that minor differences will apply to some of the methods of NZACE-Prevention compared to ACE-Prevention (Australia) in the domains of what is included in the "health perspective", and how certain costs are managed (e.g., health costs relating to extra life lived as a result of the intervention(s); and set-up costs). NZACE modelling will also include extensions beyond the Excel-based multistate lifetables to Markov models in TreeAge and micro-simulation models in R.

There are also a number of additional BODE^3 Technical Reports that more fully detail methods that could not be addressed in this overview Protocol. These Technical Reports should been seen as subsidiary to this Protocol. They are also available at the BODE^3 website (www.uow.otago.ac.nz/BODE3-info.html). At the time of this Version 2.0 of the Protocol, 12 additional Technical Reports included:

1. What are the Priority Health Risk Factors for Researching Preventive Interventions as Part of NZACE-Prevention?[9]
2. Possible NZACE-Prevention Interventions for Stakeholders to Critique.[10]
3. Burden of Disease Epidemiology, Equity and Cost-Effectiveness (BODE^3) Study Protocol. Version 1.0. [Now superseded by this Version 2.0 of the Protocol]
4. Projected NZ Life Tables.[11]
5. Incorporating Ethnic and Deprivation Variation to Cancer Incidence Estimates over 2006-2026 for ABC-CBA.[12]

6. Costing of Pharmaceuticals in New Zealand for Health Economic Studies: Backgrounder and Protocol for Costing.[13]
7. Literature Search and Data Synthesis Methods for Estimating Inputs for Health Economic Modelling.[14]
8. Price Elasticities for Health Economic Modelling of Food Pricing Interventions in Australia and New Zealand.[15]
9. Modelling Options for ABC-CBA.[16]
10. Cancer Excess Mortality Rates Over 2006-2026 for ABC-CBA. [17]
11. Determination of Effect Size for Modelling in BODE^3: A Worked Example of the Effect of Reducing Dietary Saturated Fat Intake on Cardiovascular Events.[18]
12. Protocol for Direct Costing of Health Sector Interventions for Economic Modelling (Including Event Pathways).[19]

PART I: OVERALL BODE3 PRINCIPLES AND PROTOCOL

1 Key principles

1.1 General approach

BODE^3 is a Health Research Council of New Zealand funded programme, running from 2010 to 2015. The stated aim of BODE^3 is:

> To build capacity and academic rigour in New Zealand in the estimation of disease burden, cost-effectiveness and equity impacts of proposed interventions, and undertake a range of such assessments.

The objectives during 2010 to 2015 are:

1. To estimate the impact (total & equity-related) and cost-effectiveness of cancer control interventions using Markov time dependent macro- and micro-simulation models and discrete event simulation (Aotearoa Burden of Cancer and Comparative Benefit Assessment study; ABC-CBA).
2. To estimate the impact (total & equity-related) and cost-effectiveness of preventive interventions using multistate lifetables and micro-simulation models (NZ-Assessing Cost-Effectiveness: Prevention; NZACE-Prevention).
3. To build capacity and academic rigour in disease, intervention, equity, uncertainty and cost-effectiveness modelling.

Beyond 2015, and perhaps added on during 2010 to 2015, there may be extensions to include other domains (e.g., mental health).

Stated briefly, both ABC-CBA and NZACE-Prevention will calculate the disease impact and cost-effectiveness of interventions by propagating the impacts of the interventions through core (and on occasion additional) Markov and multistate lifetable models, with costing done both external to the models (i.e., intervention costs per se) and internal to the models (i.e., health care system costs, and resultant cost offsets). The models will be created using epidemiological parameters (e.g., incidence of disease and sequelae, disease stage at presentation, survival, mortality and survival rates) from a variety of sources, such as the parallel NZ Burden of Disease study (BDS) being conducted by the Ministry of Health, cancer registry data, etc. The modelling of the effects of interventions will be done by altering at least one of these parameters, or by using the baseline models as a departure point (or 'inventory' of epidemiological and costing parameters) to undertake more particular modelling specific to the intervention (e.g., additional states for serial relapse in a Markov model, or discrete event simulation for a cancer screening programme).

This Protocol builds on the Australian ACE-Prevention Protocol, and is consistent with a recent ISPOR series of papers on best practice in economic decision modelling.[20-26] Where issues are not covered in this protocol, these external sources will provide guidance.

There are also a number of additional BODE^3 Technical Reports that more fully detail methods that could not be addressed in this overview Protocol. These Technical Reports should been seen as subsidiary to this Protocol. They are also available at the BODE^3 website (uow.otago.ac.nz/BODE3-info.html). At the time of this Version 2.0 of the Protocol, 12 additional Technical Reports include:

1. What are the Priority Health Risk Factors for Researching Preventive Interventions as Part of NZACE-Prevention?[9]
2. Possible NZACE-Prevention Interventions for Stakeholders to Critique.[10]
3. Burden of Disease Epidemiology, Equity and Cost-Effectiveness (BODE^3) Study Protocol. Version 1.0. [Now superseded by this Version 2.0 of the Protocol]
4. Projected NZ Life Tables.[11]
5. Incorporating Ethnic and Deprivation Variation to Cancer Incidence Estimates over 2006-2026 for ABC-CBA.[12]
6. Costing of Pharmaceuticals in New Zealand for Health Economic Studies: Backgrounder and Protocol for Costing.[13]
7. Literature Search and Data Synthesis Methods for Estimating Inputs for Health Economic Modelling.[14]
8. Price Elasticities for Health Economic Modelling of Food Pricing Interventions in Australia and New Zealand.[15]
9. Modelling Options for ABC-CBA.[16]
10. Cancer Excess Mortality Rates Over 2006-2026 for ABC-CBA. [17]
11. Determination of Effect Size for Modelling in BODE^3: A Worked Example of the Effect of Reducing Dietary Saturated Fat Intake on Cardiovascular Events.[18]
12. Protocol for Direct Costing of Health Sector Interventions for Economic Modelling (Including Event Pathways).[19]

1.2 Overarching Framework of BODE^3

1.2.1 Allocative efficiency

BODE^3 will be evaluating allocative efficiency. Allocative efficiency answers questions of 'what to do', ideally across the health sector but more often within domains such as prevention or disease groupings (e.g., cardiovascular disease). For example, should we introduce a new drug for colon cancer or increase palliative care nursing? This also means that when we investigate drug treatments, where feasible analysis will preferentially focus on classes of drugs (e.g., receptor-specific drugs for breast cancer) rather than single drugs (e.g., trastuzumab).

Technical efficiency answer questions of 'how to do it'. For example, which regimen of a new drug to use for colon cancer. BODE^3 will not be addressing technical efficiency in the first instance.

1.2.2 National decision context

BODE^3 will be primarily addressing questions about the national allocation of health resources. However, the boundaries with DHB-level down decision-making are inevitably blurred in a health system where much of the decision-making resides at the DHB-level. Key non-academic audiences for BODE^3 findings will include researchers, the Ministry of Health, the National Health Board, the National Health Committee, PHARMAC, and national organisations (e.g., Medical Colleges, National Heart Foundation).

1.2.3 Balance of rigour, relevance and process

NZACE-Prevention draws from the ACE-Prevention (Australia) project, which emphasises a balance of technical rigour (what matters to academics), relevance (what matters to policy makers), and due process (what matters to stakeholders).[27] BODE^3 will follow this balance.

1.2.4 Study perspective for health consequences

Consequences, that is health impacts sometimes more loosely referred to as benefits, will follow a cost-utility framework. Healthy life-years gained (or potentially lost) for a given intervention will be quantified using a hybrid measure of mortality and morbidity, namely health-adjusted life-years (HALYs) gained. Whilst the HALYs gained in BODE^3 have a burden of disease foundation, they are actually very similar to QALYs with two differences:

- disability weights (DWs) rather than utilities are used to adjust for quality of life.
- the maximum HALYs that can be awarded for any given sex by age group is not 1.0, but rather 1 minus the population morbidity (or pYLD).

Regarding the utility versus DW aspect, the utility component of a QALY is often disease or survey specific. This may be appropriate if one is just assessing interventions for (say) end-stage renal failure. However, an overarching goal of BODE^3 is to make comparisons across multiple interventions, risk factors and disease states, with the greatest comparability possible.

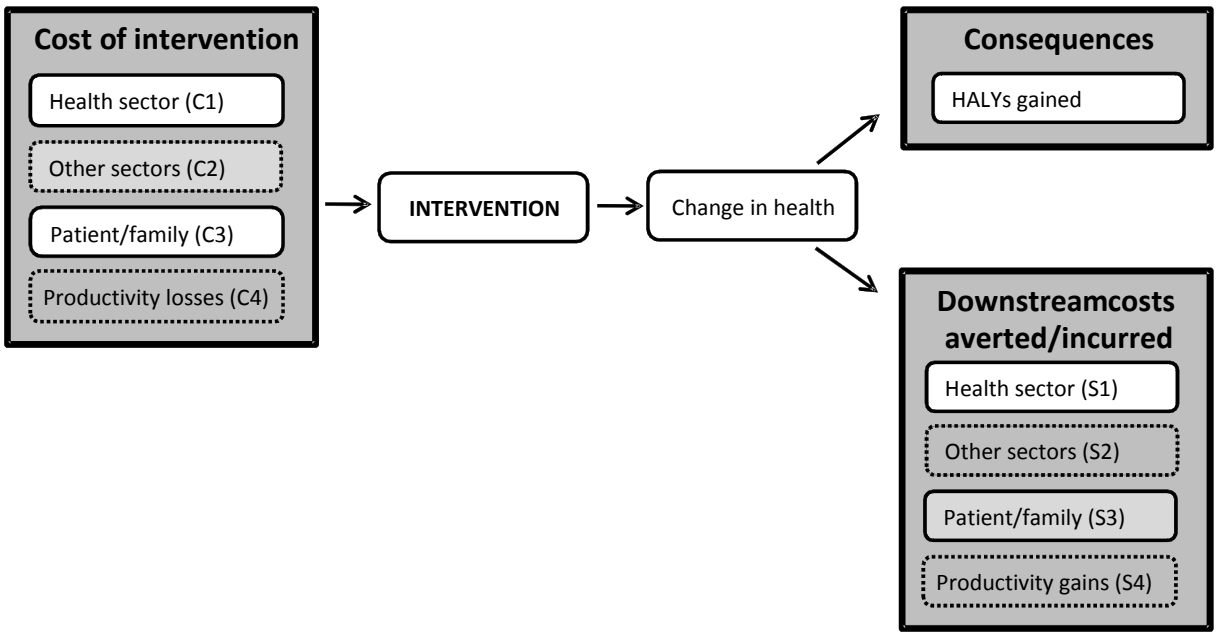
Consequences beyond directly measurable mortality and morbidity effects to the individual are not included. Thus, issues such as the impact on partner or family/whanau wellbeing from an intervention directly affecting just one person are not quantified. Nor are issues such as reassurance from a true negative screening test (a potentially positive consequence of an intervention). Such 'extra-HALY' aspects of the consequences of an intervention, if assessed as substantial, will be flagged in discussion.

1.2.5 Study perspective for costing

Costing in BODE^3 is primarily from the health system perspective. However, there are some variations as outlined below.

Drummond et al (2005)[2] propose a framework for costing (and consequences) that is represented below in Figure 1. From the costing perspective, the first distinction is between the direct costs of the intervention ('C' costs in figure) and the future costs incurred/averted due to flow-on changes in disease incidence and prevalence ('S' costs). Both C and S costs can then be disaggregated into four further costs: 1. health sector (including government and private funded, and voluntary); 2. other sectors (e.g., road traffic safety; again could be government, private or voluntary); 3. patient and family (e.g., co-payments, cost of travelling to clinics); and 4. productivity.

Figure 1: Components of economic decision modelling in BODE^3 – adapted from Figure 2.1 of Drummond et al (2005)[1]. Components shaded in white boxes are routinely in scope in BODE^3, and in half-tone boxes are included either as practicable or as scenarios analyses – see text.



There are many (often subtle) variations in what costs are included in an economic analysis. A health funder perspective would include only C1 and S1 costs, although there may be variation on whether just Government funded activity or Government plus private and voluntary sector activity is included. Furthermore, there are boundary issues to consider between what is the health sector, and what is not. For example, disability support and occupational safety and health.

A health system perspective would normally extend out to include both C1 and C3 and S1 and S3 costs. Within C3 costs, though, issues then arise as to whether one includes just direct patient co-payments and costs (e.g., petrol for travel), or one additionally includes a cost for patient/participant time taken up by the intervention (which is an opportunity cost). Including a costing of time raises further issues, for example: how to cost it (average wage, or leisure time cost?); full attribution, or partial (e.g., a patient engaged in a treatment may also derive value from reading during that time); and ensuring that the utility component of health consequences is not also quantifying time considerations.

In prevention, the intervention cost is often external to the health sector (i.e., C2 costs). For example, a Ministry of Transport may fund breathalysers and traffic safety campaigns. It would seem to be nonsense to exclude these costs that are so obviously and directly part of the intervention – unless perhaps you were a policy-maker with concerns only about the government Vote:Health budget. However, what about the cost to the tobacco industry of improved health warning labelling on products? Should that be included in the analysis? Possibly 'yes' if a classical welfare approach was being taken. But probably not for other welfare approaches (e.g., if a society had moderately developed "consumer rights" norms and legislation relating to informed choices). And probably definitely not from a health system point of view.

There will also often be pragmatic limitations on costing too. For example, determining the future stream of S3 health system costs as well as S1 health system costs may be difficult.

In BODE^3 we will adopt a **base-case** model primarily from a health system perspective (Table 1), and then possibly assess **variations** about this base case scenario.

Table 1: BODE^3 base-case approach to costing

Costs of intervention	**Health system downstream costs (that when summed give cost-offsets)**
C1: Health sector, including: • Government costs to Ministry of Health (including disability support and Government-funded primary care), DHBs, ACC. • Voluntary and NGO costs, such as Cancer Society and Heart Foundation costs of running a health education or supportive care programme, and recognising that costs will often either be: reimbursed by subcontract to Government agency; or passed on to (or raised from) patients/public. • Private health care and other directly health-related out of pocket expenditure. **C3: Patient/family**, including: • Patient co-payments and out-of-pocket costs, such as prescription fees, doctors' fees, physiotherapy, etc. • Direct travel costs (e.g. petrol). But excluding costing of patient/family time.	**S1: Health sector**, including: • Principally Vote:Health (i.e., government) costs as captured by HealthTracker. But excluding due to pragmatic reasons many of the (likely) smaller S1 costs. **S3: Patient/family**, including only: • Other costs as captured by HealthTracker, including average co-payments for primary care. But excluding due to pragmatic reasons many potential S3 costs.

In addition to this base-case costing, the most likely alternative scenario will additionally include C2 costs (i.e. other sector). This would be most likely for preventive interventions

that involve the non-health care sector. Whether it is just government-funded or wider C2 costs will be explicitly stated in scenario analyses. (S2 cost offsets also arise. For example, anti-acute psychosis drugs may reduce downstream "Law and Order" costs. It is unlikely that we will include these).

Regarding S1 (and S3) costs, they will include both related and unrelated health care costs (see Section 4.5). The rationale for not including productivity costs is further discussed in Section 4.

The methods for costing are considered in detail in Section 4 and in other BODE^3 Technical Reports.[13 19] Briefly, the costs (more strictly the opportunity costs) are to be measured in 'resource cost' terms; i.e., the economic or 'dollar' costs of the resources involved, approximating market prices.

The costs of the intervention will be costed directly, using activity costing as the default position. A top-down approach will be used to estimate downstream costs averted and incurred (i.e. the future stream of health system costs that manifest as cost offsets). This will be undertaken using HealthTracker data from the Ministry of Health. These costs will include those averted by reduced incidence of disease and sequelae, as well as future costs incurred by improved survival. See Section 4.2 for further details.

1.2.6 Reference year

The reference year is 2011, the year for which the baseline epidemiological models are constructed. For example, the prevalence of various risk factors is matched as closely as possible to the state of New Zealand in 2011. Likewise real costs are in 2011 values. 2006 is the reference year for the current BDS revision, although projections to 2011 for some (but not all) epidemiological parameters will be provided.

1.2.7 Target population

The default target population is the resident New Zealand population alive in 2011 who are potential recipients of the intervention. For treatment interventions in ABC-CBA, this means the newly incident cases in 2011. (See Section 1.3.2 for a discussion of why prevalent cases at 2011 probably do not need to be modelled). For preventive interventions, this means the population identified for the preventive measure in 2011. However, many interventions will be targeted at subpopulations, such as an appropriate age group for a specific screening programme.

Note that a closed cohort approach will be used. That is, only the usually resident 2011 population will be included in models. Put another way, people born after 2011 or immigrating to New Zealand are not included. (Also note that we will not be able to allow for emigration out of New Zealand from our closed cohort either).

Allowing for the examination of heterogeneity across subpopulations in costs and/or consequences, and hence cost-effectiveness, is a critical feature of both ABC-CBA and

NZACE-Prevention. Both models will, where feasible, disaggregate the population into strata of sex, age, ethnicity, deprivation and such like, allowing highly specific 'business as usual' parameterisation (e.g., cancer survival by age by sex by ethnicity by stage group), as well as the ability to simulate interventions through selected subpopulations. Indeed, a major aim of BODE^3 is to quantify equity impacts, and this will be achieved by leveraging off these disaggregated strata allowing modelling of consequences and costs separately by subpopulation.

1.2.8 Time horizons

There are two time horizons to consider: the time over which the intervention is applied (and hence costs of the intervention itself accrue); and the time over which both consequences and costs averted/incurred are tracked and summed.

Intervention time frame. The time frame for the application of the intervention will follow how the intervention would be applied in real life, with the default position being a time-frame of one year (i.e., the reference year 2011). For instance, a three month nicotine replacement therapy (NRT) could be modelled as applying to eligible members of the 2011 New Zealand population if this is consistent with the interventions specification; it might be, however, that NRT is one component of a more comprehensive five year programme.

There are challenging study design interactions with the intervention time horizon (e.g., comparability of prevalent versus incident disease, time-lags of prevention/screening programmes and attribution of consequences in future years). These issues need to be carefully considered in the final determination of time horizons. Therefore, we defer further discussion of intervention time frames to a more detailed consideration below under Study Design (Section 1.3.2, page 15).

Follow-up time frame. The 2011 population will be followed up for consequences and costs (averted and incurred specifically due to the intervention) till death or 110 years of age.

1.2.9 Defining the intervention

In modelling interventions we fully specify all activities (i.e., who does what, to whom, when, how many times, where and how often?). The default position is to assume that the intervention is fully implemented and operating at full potential (i.e., in 'steady-state' operation). However, set-up and time delays will be modelled when appropriate. Whether or not to include set-up costs and phases will depend on how important they are for cost-effectiveness, and will be clearly documented in evaluations.

1.2.10 Defining the comparator

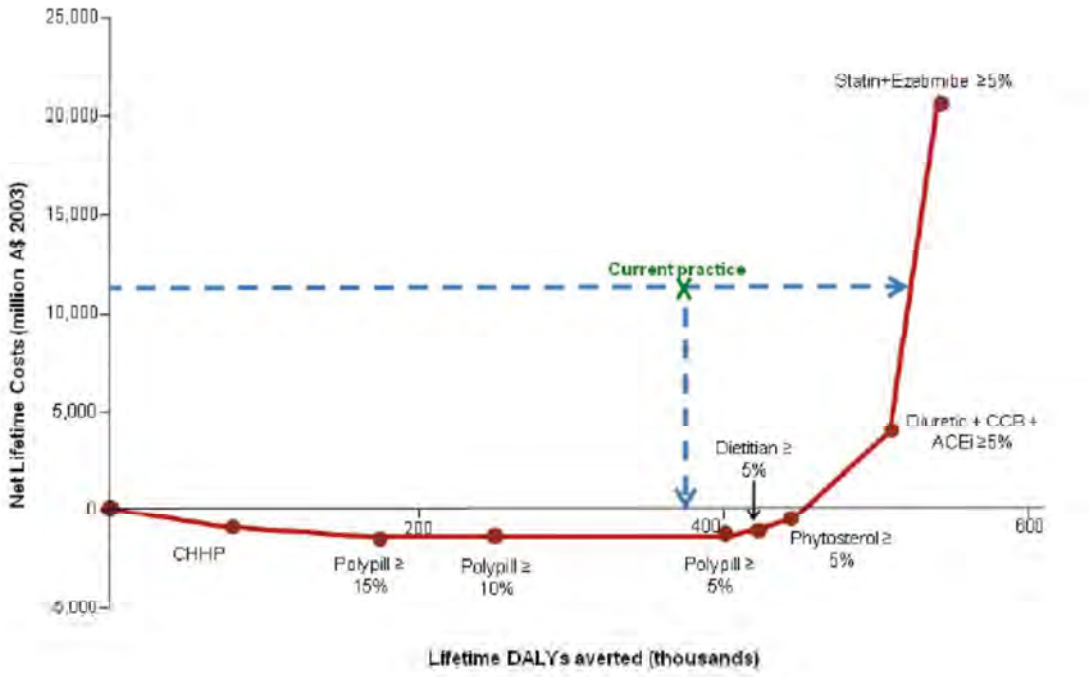
One of two comparators will be used. First, 'current practice' or 'business as usual'. Here, any effect of an intervention is assumed to be over and above the cumulative effect of current interventions, and acting on the baseline of future projected disease incidence, duration and survival parameters. This 'current practice' comparator is adequate for

interventions that occur in addition to the current array of interventions. It gives rise to incremental cost-effectiveness ratios (ICER).

A problem, however, with current practice as the comparator is that inefficient current practice will make many alternative interventions look very cost-effective. For example, if a very expensive and moderately effective drug is currently being used for osteoarthritis, an incremental analysis that substitutes this drug with a cheaper but equally effective intervention will be cost saving with no change in HALYs. However, if one considered the comparator as no use of any drugs for osteoarthritis, both drugs may be cost ineffective.

An alternative approach, therefore, is to use a 'do nothing' or 'partial null' comparator, sometimes known as generalised cost-effectiveness analysis (GCEA).[6 28] This is the preferred option if the policy-maker wants to know what the cost-effectiveness of both current practice and an 'ideal' combination of interventions might be had the health system not have evolved the way it has in recent decades. Figure 2 below is a good example. It shows cholesterol and blood pressure lowering medications including a polypill (a low-cost combination of three generic blood-pressure-lowering drugs and a cholesterol-lowering drug (statin) in a single pill) compared with current practice, from the ACE-Prevention (Australia) project.[27] Current practice is shown as costing about $12 billion AUD, averting about 380,000 DALYs (equivalent to HALYs gained), for a cost-effectiveness of about $32,000 AUD per DALY averted (which is cost-effective for a common threshold of about $50,000 per DALY). Importantly, the comparator is a revised model state, whereby the effects of currently used interventions (e.g., including more expensive cholesterol lowering drugs) are taken out of the model by revising mortality and incidence rates for CVD as though these existing interventions were not occurring. That is, a 'partial null' of no existing interventions. The advantage of this approach is that one can now see the cost-effectiveness of both 'current practice' as opposed to what might be alternative packages, namely the comparison of average cost-effectiveness ratios. The graph shows the same DALYs averted with current practice could be achieved with Community Heart Health Programmes (CHHP) and polypills, and actually *save* about $1 to $2 billion AUD (as opposed to the current $12 billion AUD *cost* of current practice). That is, the 'partial null' approach allows a full evaluation of current practice against alternative practices, potentially identifying major areas for practice change that may free-up resources for use elsewhere.

Figure 2: Intervention pathway of the most cost-effective interventions for blood pressure- and cholesterol-lowering interventions, including the polypill, compared to current practice. Source: Vos et al (2010)[27]



Polypill for those at ≥5%, ≥10% or ≥15% five-year cardiovascular disease risk is considered. The polypill is assumed to cost $200 p.a. per person.

BODE^3 will use a 'partial null' comparator where justified, and otherwise a 'current practice' comparator. The choice of comparator will be explicitly stated.

1.2.11 Adherence to principles and base-case models stated in the Protocol

Peer reviewers or editors may wish to alter our BODE^3 default approaches on a case by case basis. Wherever possible, this will be resisted so as to ensure as much comparability across published outputs. Reviewers and editors will be alerted to this Protocol.

1.3 Study Design

1.3.1 Models

Version 1.0 of this Protocol [29] placed emphasis on pre-configuration of all cancer and prevention data in 'baseline' ABC-CBA and NZACE-Prevention are Markov and multistate lifetable models, respectively. That is, macrosimulation models that describe the New Zealand 2011 population now and into the future, assuming 'business as usual'. However, we have altered our emphasis now to be more about a 'common databank of input parameters' wherever possible. Some interventions will be able to be modelled with the simple baseline models, but many will not (e.g. a relapse state or a complication might need to be added, etc.).

Many states in the models, or starting points for DES, will be specified for the New Zealand population in 2011, disaggregated by sex, age (single or five-year age group), ethnicity (Māori, non-Māori initially), and deprivation (three groups: deciles 1-3, 4-7, 8-10). For ABC-CBA, there may be additional states or competing events for stage or severity of cancer – especially those cancers suitable for screening and early detection programmes (i.e., cervix, breast and colorectal). Starting distributions of the NZ population in 2011 are derived from SNZ population projections, CancerTrends[30] 2011 SNZ population projections and cancer projections data[31 32], and the NZ-BDS revision (in progress). Transition probabilities and time to event are used to stream the simulated population in each stratum of the initial 2011 population over time, transitioning to various disease states: cured, alive, living with sequelae, and such like. Baseline transition probabilities and time to event are taken from the NZ-BDS and survival analyses conducted on cancer registry data, and other existing New Zealand data. However, it will often be necessary to incorporate data from elsewhere; this will be fully documented in Appendices to specific evaluations.

There are many key decisions to make in any modelling project that balance parsimony with complexity.[33-36] Issues include:

1. *Whether to model a single age group, or multiple age groups.*
   Understanding heterogeneity across age groups is an important feature of BODE^3, and we have sufficient data to do so. Therefore, our default approach is to model all age groups considered eligible for the intervention, using the 2011 population distribution of age (i.e., the population alive in 2011 is used for multi-cohort modelling). This will often be achieved by 'multiple cohort Markov modelling'.

2. *Whether to include prevalent cases of disease in 2011, or just incident cases of disease in 2011 that are prevalent thereafter.*
   For ABC-CBA at least, our default position will be to model incident cancer cases in 2011 only. Treatments for people in the second year or later post-diagnosis will be captured as the population ages. This will mean that due to discounting the absolute costs and consequences will be less than if prevalent cases in 2011 were modelled, but the relativities (and hence the ICER) will be unaffected. This issue, and its interaction with other model design assumptions, is considered further in Section 1.3.2 below.

3. *Whether to include incident cases in future years (i.e. 2012 and beyond) as the 2011 population ages.*
   This depends on the analysis question. For treatment, supportive care and palliation it makes little difference to relativities, and thus parsimony dictates just using incident cases in 2011. For preventive interventions, however, one obviously needs to model future changes in incidence. The approach in both ABC-CBA and NZACE-Prevention is to model changes in future incidence among the population alive in 2011. Again, this issue is considered further in Section 1.3.2 below.

All baseline or business as usual models (i.e., before any intervention is modelled) require projections into the future. For example, projected cancer incidence and survival for the remainder of the 2011 population's life. As a general principle, where there is sufficient data (e.g., cancer incidence projections[31 32]) or strong theoretical expectation (e.g., steady improvements in ovarian cancer survival across multiple countries; phasing of the tobacco epidemic associated with delayed phasing of disease rates) we will model changes in such epidemiological parameters for 15 years out to 2026, then assume they are constant thereafter. Specific details or projections will be described later in this protocol, and in subsequent reports and papers.

1.3.2 Interacting consideration of time horizon for application of intervention, and comparability across interventions

There are two key issues to consider:

- Treatment, supportive care and palliative interventions for prevalent and incident cases of disease.
- Preventive and screening and early detection interventions and attribution of future consequences.

1.3.2.1 Treatment, supportive care and palliative interventions for prevalent and incident cases of disease

Should one model incident cases just in 2011, or incident cases in all subsequent years? So long as the discount rates are the same for consequences (i.e., HALYs) and costs, and treatment costs (real dollars, non-discounted) and treatment effectiveness do not change much in the future, the ICER for the 2011 and future incident cohorts will be the same.[36] An exception to this generalisation is if the ICER varies markedly by age, as future incident cases will arise from a modified age distribution (especially if it is only the 2011 population that is modelled out into the future). However, we will explore this by analyses directly of heterogeneity (i.e., examining ICER by age group).

For a treatment applied to the 2011 population, does it matter whether one includes incident, prevalent or both incident and prevalent cases within the population during that year? Quite often, 'yes'. If the disease in question has a reasonably long duration (i.e., years to decades) such that average age differs between incident and prevalent cases, and is progressive such that disease severity increases with duration, and treatment effectiveness (or cost) varies with age and or severity, then the ICER among incident and prevalent cases may vary (often substantially).[36] The exact amount of variation is context specific. Sometimes the cost-effectiveness may be better among prevalent cases; for example, more severe cases of CVD may be more likely to benefit in terms of deaths prevented than less severe CVD. Sometimes the cost-effectiveness may be worse among prevalent cases; for example, a new treatment for osteoarthritis may generate less health gain among older people with more severe disease.

This difference in ICER between prevalent and incident cases can be thought of as heterogeneity of the ICER across subpopulations. Both ABC-CBA and NZACE-Prevention

models will be finely disaggregated by sex, age, ethnicity (and sometimes deprivation and stage). The models are less prone to produce different ICERs for incident and prevalent cases if results are determined by subpopulation. Treatments are more likely to be modelled in ABC-CBA than in NZACE-Prevention. Cancer is often not a long duration (i.e., many decades) chronic and progressive disease, although exceptions such as leukaemia, breast and prostate cancer definitely exist. Otherwise, cancer is often a rapidly progressive disease (be it to death or cure). Thus, this incident versus prevalent disease ICER issue is not too concerning within BODE^3.

Perhaps more important in ABC-CBA are two more subtle questions:

- Do we just model cancers incident in 2011, and assume that results apply also to prevalent cancers in 2011?
- For interventions that often occur in later years of the disease course (e.g., palliative care, treatment of relapse), do we just model the costs and consequences of these interventions in out-years (e.g., 2021) as some of the cases incident in 2011 become eligible?

These two questions are actually different sides of the same coin, and thus have the same answer. Most cancers have curative treatment applied in the first year or so; therefore, there will not often be the need to consider prevalent cases. Considering supportive or palliative care, costs and consequences will occur together at some number of years post diagnosis. Given the same discount rate for costs and consequences, the ICER should be much the same for (say) palliative care applied to those needing it in 2011 (i.e., arising from many previous annual cohorts of incident cases) compared to that arising in future years among the 2011 incident cases.

Thus, for the issues considered above it appears that we are justified in *modelling only incident cases in 2011 for treatment, supportive care and palliative interventions.*

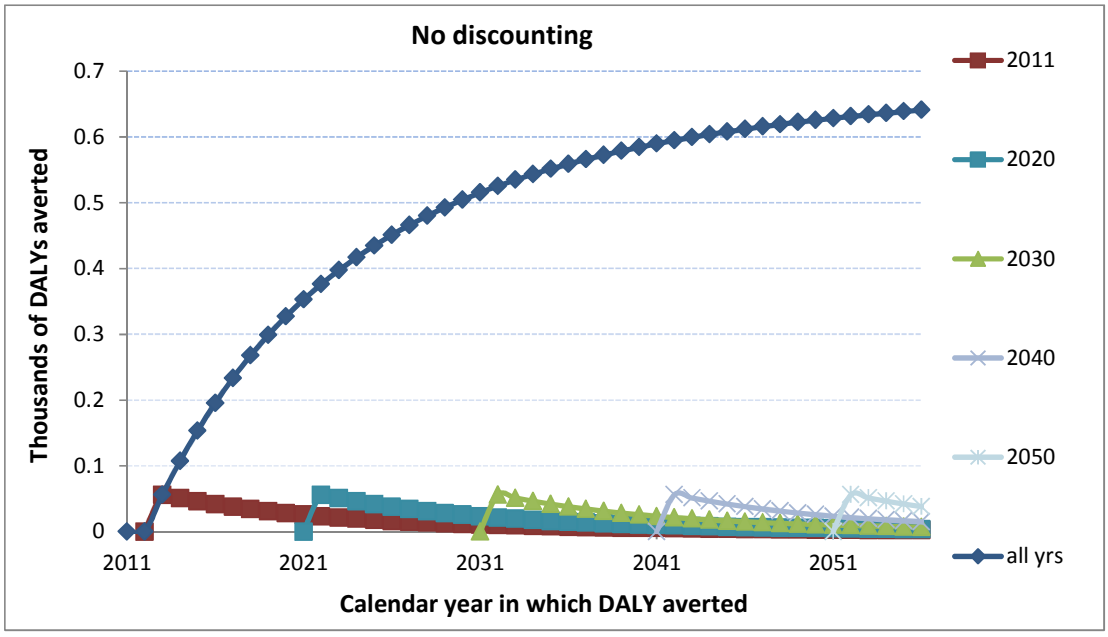
1.3.2.2 Preventive and screening interventions and attribution of future consequences

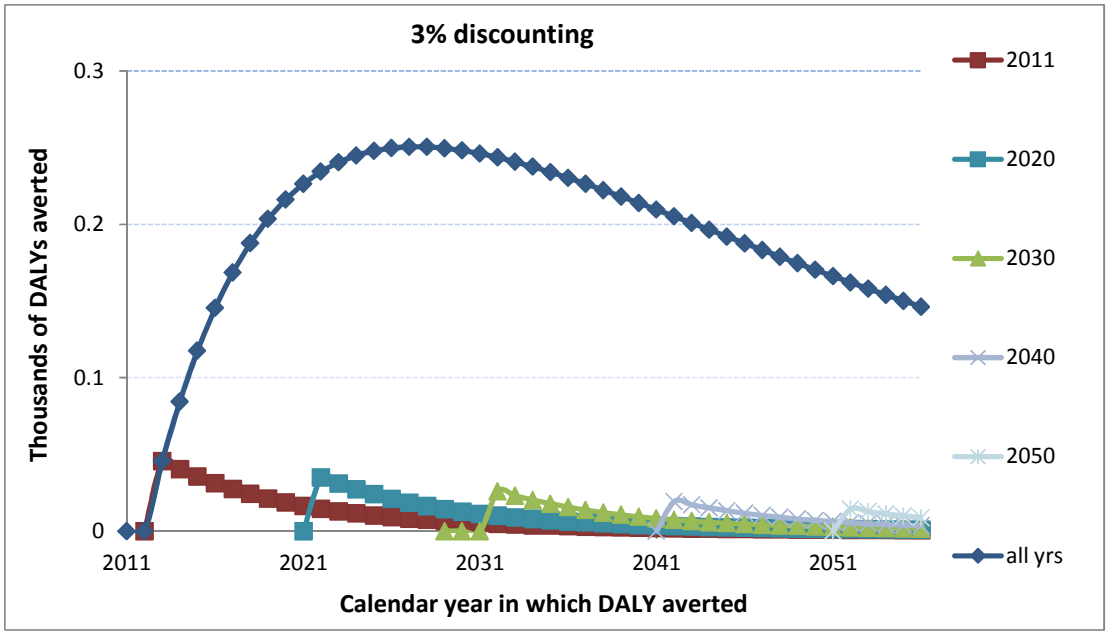
The second issue identified at the outset of this Section was the time horizon over which to apply a prevention or screening intervention. There are two scenarios to consider. First, if the intervention only needs to occur in one year (such as changing a legislation), then one theoretically only needs to model the direct costs of the intervention as incurred in base-year 2011 and sum the HALYs gained and costs incurred/averted thereafter (allowing for time-lags and discounting). (That said, such single year interventions also have monitoring, enforcement or other costs in out years).

Second, some preventive interventions need to be applied continuously to maintain the effect. For example, a future 30% reduction in cancer X mortality requiring screening every 3 years, or a sustained healthy eating campaign that is required to maintain dietary changes. We consider this now.

Figure 3 schematically shows HALYs gained by some preventive campaign applied sequentially in subsequent years to the same (aging) population (series), shown by subsequent calendar year in which the HALY were actually gained (x-axis). The future stream of HALYs gained is assumed to be a log normal function of time for each single calendar year in which the intervention was applied. For example, and shown in the graph as the 2011 series, one can plot the future stream of HALYs gained caused by the intervention in 2011 only. Figure 3 shows such individual-year-HALY-attributions for every tenth year only to avoid clutter. The total HALYs gained in each calendar year post-2011 is also shown (i.e., the 'all yrs' series), which is the sum across all annual applications of the intervention. Note that no allowance has been made for attrition (death mainly) of the starting population, which would mean that the all years line in the 'no discounting' sub-figure would actually peak and decline well before 2056, and the peak in the 3% discounting sub-figure would occur earlier than 2025.

Figure 3: Schematic example of HALYs gained by annual cohort in which preventive intervention applied (i.e., line series), presented by calendar year in which the HALY was actually gained





There are three options for calculating the ICER in the above scenario. First, one could assume the intervention was at steady state and applied to 2011 only. The HALYs gained would be that of the 2011 line only in Figure 3. When it is possible to calculate the stream of HALYs gained in the future due to the intervention effect in one year only, this is viable. The ICER will be the same as that modelled for the programme applied over many years (so long as the discount rate of costs and consequences are the same, and intervention effectiveness and cost structures are similar in the future). However, it is often not possible to estimate the HALYs gained from just one year of the preventive interventions, when the actual intervention or screening programme needs to run continuously. For example, reductions in cancer mortality are usually estimated for a long-running programme as a whole, not allocated per annum of the programme and strung out over future years.

The second option is to simply model the intervention as applying indefinitely from 2011 (or until all members of the simulation population have died), and calculate all of the (discounted) costs and consequences over this same intervention time horizon. *This modelling indefinitely into the future will be our default intervention time frame option for preventive programmes.*

The third option is to model the intervention as running for, say, 10 years (as in WHO-CHOICE [6]) to 20 years, allowing for specification of an aggregate time lag to full programme effectiveness and decline after 10-20 years. However, this is still computationally problematic as one needs to model the decrease in effectiveness after stopping the intervention.

Regardless, the ICER will theoretically be much the same for the above three options with a closed cohort, unless future disease incidence (due to other secular trends) is projected to change markedly, or future (real) cost structures and programme effectiveness are projected

to change markedly. Any such differences in ICER between approaches will reduce further with discounting and attrition of the base-year cohort. The choice of intervention time frame is driven more by pragmatic and empiric considerations.

1.3.3 Consequences: general measurement[b]

We measure the size of the health gain associated with each intervention in 'health-adjusted life years' where we value the loss of health due to non-fatal health states with the appropriate disability weight(s) (see Section 1.3.5 below) used to estimate disability-adjusted life years (DALYs) in burden of disease studies. However, there are important differences between DALYs calculated in burden of disease studies and HALYs gained calculated in disease and economic modelling, such as BODE^3. First, in a burden of disease study, the health status of a population is estimated in a particular year. It is, therefore, a cross-sectional measure. Economic evaluation methods always have a time dimension. Health gain is calculated as the difference in mortality and morbidity outcomes between a comparator and the intervention option over a defined period of time (the 'time horizon').

Second, in burden of disease studies the DALY is constructed as a health gap measure, i.e., an ideal is set (everyone ought to live into old age free of disease) and contrasted with the current health status of a population. Thus, years of life lost, the mortality component of the DALY, are calculated as the difference between age at death and a standard life expectancy at that age for each death. In the economic analyses of BODE^3, we do not use the standard life table to give a value to loss of healthy life. Instead, we track a target population over time and count the health-adjusted years of life lived in intervention and comparator scenarios assuming realistic mortality risks as people age (i.e., the population's own lifetables and its future projections[37] – not an external 'ideal' standard; the population's own lifetable concept will be extended to sex by ethnic (and occasionally by deprivation) specific lifetables, and is described in more detail below in Section 1.3.4, and in BODE^3 Technical Report 4.[37]

This quantification of HALYs also includes an adjustment for expected levels of disability by age and sex for conditions not immediately affected by the intervention of interest. In other words, extra years of life gained from an intervention are counted as less than full years taking into account the probability that the person will suffer from osteoarthritis, dementia, hip fracture or any other condition as they age. That is, we allow for co-morbidities. Operationally, this involves assuming an average disability weight (DW) by sex and year of age, using averages from an appropriate BDS (which will be the New Zealand BDS for 2011 in due course). We allow for expected background disabilities in order to measure realistic health gains, rather than hypothetical health gains assessed against perfect health. Actual DWs used in BODE^3 by sex, age are described in more detail in Section 1.3.7 below.

[b] This section is taken from the ACE Prevention protocol, with minor amendments only.

1.3.4 Consequences: allowing for background mortality rates

Background mortality rates from lifetables are transformed into transition probabilities or time to event distributions to 'death from other causes' in Markov, multistate lifetable and DES models (having first subtracted away the mortality rate(s) for the disease(s) addressed by the intervention).

Two questions arise at this point:

1. Should we allow for varying mortality rates not only by sex and age, but also by ethnicity and deprivation?
2. Should we allow for varying mortality rates into the future?

Our answer to both questions is 'yes', as to not allow for these variations is equivalent to assuming no variation in mortality by ethnicity and deprivation, and no variation over time in mortality rates. Both assumptions are clearly false. However, there are major equity implications that arise from using socio-demographic specific back ground mortality rates (Māori, for example, stand to gain fewer HALYs). This is being worked through under Objective 3 of the programme, and as part of PhD; subsequent versions of this Protocol will likely include specific sections and recommendations on 'equity analyses'.

Regarding ethnic and deprivation-specific lifetables, and hence mortality rates, for base-year 2011, we undertook the following process:

- Official Māori and non-Māori complete lifetables for 2011 were sourced from Statistics New Zealand.
- Abridged (i.e., age <1, 1-4, 5-9,…. 85+ year increments only) lifetables for Māori and non-Māori each disaggregated by quintile of NZDep were sourced from the Ministry of Health.
- The mortality rates from the abridged lifetables for each quintile of deprivation (within ethnic and sex groups, and age groups) were then submitted to a simple linear regression with quintiles coded as 0.1, 0.3, 0.5, 0.7 and 0.9. Thus, the intercept was the expected mortality rate for the least deprived (i.e., 0th percentile) for any sex by age by ethnic groups, and the slope was expected increase in mortality from the least to most deprived (i.e., 100th percentile).
- The slope and intercepts from these models were then used to estimate the mortality rate at the mid-point of the three deprivation groupings used in BODE^3 (i.e., deciles 1-3, 4-7 and 8-10), and hence to estimate mortality rate ratios for these three groups. (Rate ratios above the age 85 were assumed to linearly approach 1.0).
- These rate ratios were then applied to the Māori and non-Māori lifetables to generate 12 complete (i.e., single year of age) lifetables for 2011. That is, sex (2) by ethnic group (2) by deprivation (3) lifetables.

Further details on this method, and the actual lifetables, can be obtained from the BODE^3 website www.uow.otago.ac.nz/bode3-info.html and technical report 4.

Regarding annual percentage changes into the future for mortality rates, we used two sources of information. First, Statistics New Zealand has used cohort lifetables[38] and age period cohort modelling to estimate future annual percentage changes in mortality (averaged over ages). The 2007 estimates for medium scenario projections term projections hovered around 2% per annum reductions in mortality rates for the four combinations of sex by ethnicity. Second, Blakely et al (2010) noted linear trends in the long-run trends in life expectancy for Māori and non-Māori since about 1900 to 2006.[39] Trends vary decade by decade (e.g., flat in the 1960s and 1970s for non-Māori, and likewise in the 1980s and 1990s for Māori), but over the long-run the trends in both ethnic groups are remarkably linear. Assuming such a linear trend continues until 2026, annual percentage changes in mortality rates of about 1.5% and 2.5% for non-Māori and Māori are required. Simply averaging these two sources of estimates gives a 1.75% annual reduction for non-Māori, and 2.25% for Māori. These average estimates have been applied to the above lifetables, giving calendar-year specific lifetables to 2026, and then assumed to be constant beyond 2026.[37]

(Elsewhere, we have calculated ethnicity by income by smoking status lifetables.[40] Smoking-specific expected mortality rates may be required for some simulation models in the future in BODE^3, but we will not routinely include smoking strata in the baseline model).

1.3.5 Consequences: allowing for disease morbidity

New or updated disability weights (DWs) from the GBD 2010 were made available in late-2012.[41] Unlike the previous DWs from an earlier GBD, that were based on person trade-off and expert calibration (e.g. [42]), the new DWs are derived with a different methodology – pair-wise comparisons. Survey respondents (n=15,000) in five countries (Bangladesh, Indonesia, Peru, Tanzania and USA) and 16,000 internet survey respondents were asked to rate which of two hypothetical individuals with different heath states was healthier. The data were then analysed with probit regression to rank the 215 health states, and they in turn were assigned to a score between 0 and 1 (where 0 is no disability and 1 is (equivalent to) death) using 30 anchor states that had been subjected to further questioning about the health benefits of different life saving or disease prevention programmes. Uncertainty intervals – using an underlying logit-normal distribution – were estimated with bootstrap sampling. Descriptors of each health state, and their mean DW and 95% uncertainty estimates, are all reproduced in Salomon et al.[41]

Of note, there was high concordance in the ranking of health states across countries, suggesting a global assignment of DWs is viable. There was a moderate to strong correlation of the old with new DWs for those states that were consistent, except beneath a DW of 0.20. In this relatively mild range of disability states – where most of the new DWs reside – the new DWs tended to be considerably lower than the old DWs, and with a reduced correlation with the old DWs. This issue will undoubtedly be the subject of research and scrutiny in future years.

Further, there are some specific instances of surprisingly low DWs (e.g. moderate and profound hearing loss new DWs 0.02 and 0.03, compared to 0.12 and 0.33 in old DWs). As

part of the New Zealand BDS 2006, some of the new DWs were subjected to further New Zealand-specific expert consideration and modification (personal communication, Martin Tobias, Ministry of Health, September 2012). The New Zealand- specific revisions are shown in Table 2 below.

Table 2: Summary of New Zealand Ministry of Health modifications to GBD 2010 disability weights (DWs)

Condition **Health States**	**'New' GBD Text Descriptor**	**'New' GBD 2010 Mean DW (95% UI)**	**Comment on GBD Text Descriptor**	**Anchors for NZ Modification (U/L = upper/lower)**	**NZ Modified Mean DW**
Intellectual Disability / mental retardation					
Severe	"Has low intelligence and cannot speak more than a few words, needs help with most basic activities of daily activities, and can do only simple tasks under close supervision"	0.126 (0.085 to 0.176) ['Profound' as well; 0.157, 0.107 to 0.221]	Does not fully capture the need for total care required	U: Severe dementia 0.438 L: Autism 0.259	0.348 (mid-point)
Moderate	"Has low intelligence and is slow in learning to speak and do simple tasks. As an adult, the person requires a lot of supervision to work productively live independently and raise children."	0.080 (0.053 to 0.114)	DW differs from health states with similar severities of cognitive and behavioural disabilities	U: Autism 0.259 L:Traumatic brain injury, moderate long term 0.224 (Mild dementia considered too severe)	0.240 (mid-point)
Mild	"Has low intelligence and is slow in learning at school. As an adult, the person can work at simple supervised jobs and live independently, but often needs help to raise children."	0.031 (0.018 to 0.049)	Considered to be the equivalent of the moderate level (code17) described within GBD	Not applicable	0.080
Hearing Loss					
Severe	"Has great difficulty in hearing in any situation or in using a phone."	0.032 (0.018 to 0.051)	Does not capture the psychological component associated with the reduced social contact (less severe than psychiatric	U: Motor and cognitive impairments, moderate 0.221	0.188 (mid-point)

Condition **Health States**	**'New' GBD Text Descriptor**	**'New' GBD 2010 Mean DW (95% UI)**	**Comment on GBD Text Descriptor**	**Anchors for NZ Modification (U/L = upper/lower)**	**NZ Modified Mean DW**
			disorder so not captured elsewhere)	L: Moderate anxiety 0.149	
Moderate	"Has difficulty hearing a normal voice and great difficulty following a conversation in a noisy environment."	0.023 (0.013 to 0.038)	As above	U: Motor and cognitive impairments, mild 0.054 L: Motor impairment, mild 0.012 (Closer to upper level recommended)	0.05
Mild	"Has difficulty hearing a conversation in a noisy environment but no other hearing problems."	0.005 (0.002 to 0.012)	As above	Relation to original DW and to and severity scale	0.010
Vision loss					
Severe	"Has severe vision loss, which causes difficulty in all daily activities, some emotional impact (for example worry), and some difficulty going outside the home without assistance."	0.191 (0.129 to 0.269)	Description of health state adequate, but DW is low in comparison to other health states with similar level of disability. Changed also to maintain relationship with hearing loss	U: Parkinson's Disease moderate 0.263 L: Hearing loss severe 0.188	0.225 (mid-point)
Moderate	"Has vision problems that make it difficult to recognize faces or objects across the room."	0.033 (0.020 to 0.052)	As above	U: Stroke, moderate long term 0.076 L: Hearing loss, moderate 0.05	0.060 (maintains relationship with hearing loss)
Mild	"Has some difficulty with distance vision, for example reading signs, but no other	0.004 (0.001 to 0.010)	As above	U: Parkinson's Disease, mild 0.011 L: Hearing loss, mild	0.011 (maintains relationship

Condition Health States	'New' GBD Text Descriptor	'New' GBD 2010 Mean DW (95% UI)	Comment on GBD Text Descriptor	Anchors for NZ Modification (U/L = upper/lower)	NZ Modified Mean DW
	problems with eyesight."			0.01	with hearing loss)
Heart failure					
Severe	"Is short of breath and feels tired when at rest. The person avoids any physical activity, for fear of worsening the breathing problems."	0.186 (0.128 to 0.261)	Description of health state inadequate,such that DW is low in comparison to other health states with similar level of disability e.g. COPD (Severe COPD descriptor = "Has cough, wheezing and shortness of breath all the time. The person has great difficulty walking even short distances or climbing any stairs, feels tired when at rest, and is anxious.")	COPD, severe 0.383 (equivalent)	0.383
Moderate	"Is short of breath and easily tires with minimal physical activity, such as walking only a short distance. The person feels comfortable at rest but avoids moderate activity."	0.070 (0.044 to 0.102)	As above	COPD, moderate 0.192 (equivalent)	0.192
Mild	"Is short of breath and easily tires with moderate physical activity, such as walking uphill or more than a quarter-mile on level ground. The person feels comfortable at rest or during activities requiring less effort."	0.037 (0.021 to 0.058)	Description adequate and now in keeping with case definition	No change	0.037
Atrial fibrillation [equivalent GBD 2010 health state =	"Has periods of rapid and irregular heartbeats and occasional fainting."	0.145 (0.097 to 0.205)	Quiescent phase: use heart failure, mild Active phase: use cardiac conduction disorders 0.145	Active phase no more than 10% of time expert advice of some	0.048 or less (0.037 x 0.9 + 0.145 x 0.1)

Condition **Health States**	**'New' GBD Text Descriptor**	**'New' GBD 2010 Mean DW (95% UI)**	**Comment on GBD Text Descriptor**	**Anchors for NZ Modification (U/L = upper/lower)**	**NZ Modified Mean DW**
"Cardiac conduction disorders and cardiac dysrhythmias"				panel members; to be researched further by literature review	
Attention Deficit Hyperactivity Disorder	"Is hyperactive and has difficulty concentrating, remembering things and completing tasks."	0.049 (0.031 to 0.074)	Description fails to capture the severe behavioural difficulties and cognitive/learning impacts. (Prevalence case definition implies severe case)	U: Contact disorder 0.236 L: Asperger's 0.110 (close to lower level recommended)	0.110
Infertility	"Wants to have a child and has a fertile partner, but the couple cannot conceive."	0.011 (0.005 to 0.021)	Description fails to capture the psychological component (though unclear as to severity distribution of linked psychological distress)	U: Anxiety disorder, mild 0.030 L: Existing primary infertility 0.011	0.020 (mid-point)

These caveats and modifications noted, the new DWs are based on a coherent system of estimation, using lay populations and sophisticated analytical methods. As of 2012, they are clearly the 'best current' DWs available for use in BODE^3. Whether BODE^3 uses the New Zealand modifications will be decided on a case by case basis, documented, and subjected to sensitivity analyses between New Zealand and GBD variants.

1.3.6 Allowing for two or more disability weights

A person with a particular cancer may also have other diseases or co-morbidities, with their own attendant impact on quality of life. Considering individuals for now, this creates challenges in assigning a combined DW to someone with two or more diseases and hence two or more DWs. Additionally, everyone will be assigned an average background DW for average comorbidities (see Section 1.3.7 below). For example, if someone has another disease with a DW of 0.40, and their current cancer state has a DW of 0.30, what is the combined DW for this individual? A value of 0.70 (i.e., 0.40 + 0.30) is one possibility, but two problems arise with this system:

- For someone with multiple diseases and a simple sum of DWs greater than 1.0, do we assign them a DW > 1.0 (i.e., a state worse than death)?
- Thinking in terms of capacity to benefit in intervention modelling, do we assume that the reduction in DW is just that for the disease under question, or some fraction of the individual's remaining quality of life (i.e., 1 – DW)?

The usual way of accommodating multiple DWs is to let the total disability weight = $1 - \prod_i (1-DW_i)$. In the case above, this would be $1 – (1-0.3)\times(1-0.4) = 0.58$. That is, the 'independent' DW from cancer added another $0.30 \times 0.60 = 0.18$ to the already existing 0.40 DW. It is this marginal change in the DW that needs to be captured by simulation modelling.

1.3.7 Consequences: allowing for population morbidity

As stated above in Section 1.3.3, it is unrealistic to assume that an average year of life saved at age 85 is a year of life in perfect health, or more specifically a year of life with full utility or no disability. National burden of disease studies can be used to estimate the average DW (weighted by disease(s)) by single year of age (by sex), or what has been termed the 'prevalent YLD' in the ACE-Prevention (Australia) project.[27] They are typically close to 0.0 through most years, but increase to 0.2 to 0.4 at older ages (e.g., 75 to 95 years of age) as (co)morbidities tend to increase.

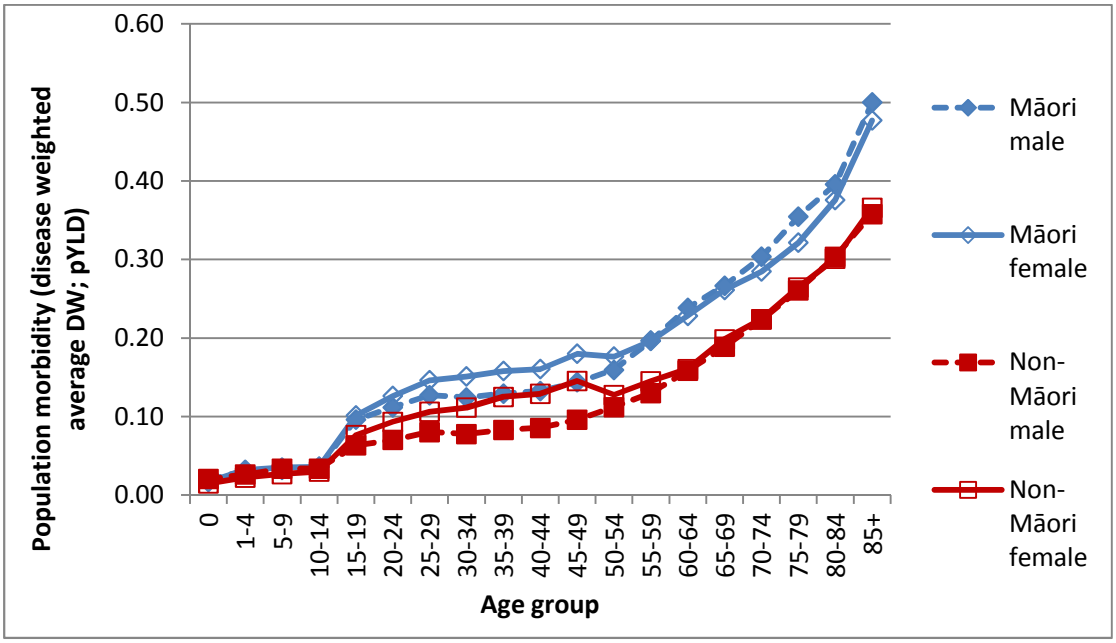
Table 3 and Figure 4 show the population morbidity (or pYLDs) for 2006 for the New Zealand BDS (personal communication, Martin Tobias and Deepa Weerasekera, Ministry of Health, October 2012). These pYLDs were calculated using New Zealand prevalence data, with 2010 revised disability weights applied[41] (with some revisions by a New Zealand expert advisory group as described above). The new DWs from the 2010 GBD tend to be lower than the previous DWs (which were largely the Dutch ones[42 43]). However, the just completed New Zealand BDS probably elicited more prevalent disease than – say – the 2003 Australian BDS.[44] As a net consequence, these 'new' New Zealand pYLDs are similar to the 2003

Australian pYLDs (which used old [higher] DWs, but probably did not elicit as much prevalent disease).

Table 3: Population morbidity (average disability weight, weighted by prevalence of disease(s); prevalent YLDs) by sex and age from the New Zealand Burden of Disease Study (year 2006, but will be used for base year 2011 in BODE^3)

Age	Males					
	Total	Māori	Non-Maori	Total	Māori	Non-Maori
0	0.020	0.020	0.021	0.015	0.017	0.015
1-4	0.028	0.031	0.026	0.025	0.032	0.022
5-9	0.033	0.032	0.034	0.029	0.035	0.027
10-14	0.034	0.036	0.034	0.031	0.036	0.030
15-19	0.070	0.096	0.063	0.082	0.101	0.076
20-24	0.019	0.112	0.070	0.099	0.126	0.093
25-29	0.088	0.127	0.080	0.113	0.146	0.106
30-34	0.085	0.124	0.078	0.118	0.151	0.112
35-39	0.089	0.129	0.083	0.130	0.158	0.125
40-44	0.091	0.133	0.086	0.133	0.160	0.129
45-49	0.101	0.144	0.096	0.149	0.180	0.145
50-54	0.117	0.159	0.112	0.133	0.176	0.128
55-59	0.136	0.197	0.130	0.150	0.196	0.145
60-64	0.164	0.238	0.158	0.166	0.228	0.161
65-69	0.194	0.267	0.189	0.203	0.261	0.199
70-74	0.228	0.304	0.224	0.228	0.285	0.224
75-79	0.251	0.355	0.261	0.267	0.321	0.265
80-84	0.306	0.396	0.304	0.303	0.376	0.301
85+	0.369	0.500	0.358	0.368	0.477	0.366

Figure 4: Population morbidity (average disability weight, weighted by prevalence of disease(s); prevalent YLDs) by sex and age from the New Zealand Burden of Disease Study (year 2006)



Conceptually, it is important to capture quality of life in economic decision models (hence QALYs and HALYs), and to allow for background morbidity that might limit the full envelope of health that someone might enjoy. However, the difference between the old and new GBD DWs, and the variation often seen across context and study in utilities from a range of measures (e.g. EQ5D, SF36, etc), clearly demonstrates that the *actual* value assigned is uncertain – and hence any aggregated pYLD. This is both a conceptual and measurement limitation. Valuation of quality of life does vary depending on how it is framed and measured. Thus, the DWs used in BODE^3 are just one set of possible weights. We believe they have overall advantages over alternatives (e.g. based on 30,000 survey respondents in five countries and through internet surveys; calculated to give one coherent set), but they are still just one possible set to use.

These caveats issued, the average DWs by sex and age are then used to estimate *healthy* life years gained, in that expected years of life gained (e.g., due to prevention of a CVD death) are adjusted for disability. For example, if a 77 year old non-Māori has death delayed due to some preventive intervention, each year of life gained is not equal to 1.0 HALYs, but rather about 0.645 (i.e. 1-0.355) HALYs as the expected population morbidity at this age is 0.25 (see Table 3).

Of note, the population morbidity varies by both age and ethnicity; Māori have higher pYLDs due to higher prevalence of disease/comorbidity. These means that the HALY envelope for health gain among Māori is less than for non-Māori. We are subjecting this, and other heterogeneous parameters, to scrutiny with respect to impacts on equity elsewhere in the BODE^3 Programme.

The 'new' DWs in the GBD 2010[41] also have uncertainty about them (an important advantage over previous work). If the Ministry of Health produces the above pYLDs with uncertainty, we may also include uncertainty in the pYLDs in BODE^3. However, there are three reasons why including uncertainty in pYLDs may be unwarranted or problematic:

- To properly determine uncertainty in pYLDs would require simulation of the New Zealand population with correlations (of unknown magnitude) between all the disease DW uncertainties. (Not allowing for such correlation will lead to underestimation of uncertainty in pYLDs).
- The pYLDs 'only' specify the limit in potential HALYs, and only make a substantial difference at older ages. Uncertainty in other parameters (most importantly intervention effectiveness and cost), not baseline parameters like pYLDs and disease incidence, will be the most influential drivers of uncertainty in incremental cost, HALYs and the ICERs. Given the structural uncertainty in quality of life estimation, it is not clear that the effort (and computational complexity and time) required to model uncertainty in pYLDs is warranted.
- The mathematical complexity involved in assigning 'HALY rewards' in economic decision models – particularly DES – if the pYLDs are subjected to uncertainty is probably not warranted (in our current assessment).

Finally, we will usually assume that the pYLDs are the expected morbidity of people without the disease(s) specific to the intervention. However, for common diseases (e.g. CVD) it may be necessary to estimate the pYLD for only those people without CVD, and therefore not just use the population-wide pYLD.

1.3.8 Discounting

Discounting is standard practice for economic evaluation to incorporate time preferences for current and future costs and benefits. In BODE^3, discounting will be applied to both costs and benefits (HALYs). The default discount rate will be 3% per annum.

Use of a discount rate of 3% p.a. is in line with ACE-Prevention (Australia). This is also the rate recommended by a consensus panel of health economists in the USA for cost-effectiveness analysis.[5] Use of this 3% figure will optimise comparisons with existing international work. We will also include other discount rates in sensitivity analyses (i.e., definitely 0%, and 6%), and on occasion a discount rate of 3.5% p.a. for consistency with PHARMAC. In some cases it may be appropriate to consider a threshold analysis (e.g., to determine at what discount rate an intervention becomes no longer cost-saving or cost-effective).

1.3.9 Uncertainty analysis[c]

There is usually considerable uncertainty in the outputs of economic decision modelling. There are different frameworks for considering uncertainty.[4 45] They may be thought of at three levels:

- Model structure uncertainty.
- Input parameter uncertainty.
- Stochastic uncertainty.

Model structure uncertainty is where the structure of the model may not be a good approximation of reality. This is difficult to incorporate explicitly into analyses, although it is still important and must be considered in interpretation. It may also be possible to consider in model structure uncertainty analyses (sometimes referred to as “sensitivity analyses” for short [see Glossary]), where the model structure is altered for a different set of assumptions. For example, additional remission and relapse stages might be added to the model, to facilitate increased ‘memory’ in modelling as to previous disease progression. (Some practitioners might prefer to not call this sensitivity analyses, but ‘model elaboration’ or some such term). Other model assumptions that may be assessed include the intervention time frame (e.g., where it is inappropriate to model just 2011 as steady-state), duration of effect of an intervention, etc.

[c] Many of terms used in this section are also defined in the Glossary.

Stochastic uncertainty, or what is sometimes referred to as first-order uncertainty or random variability, is simply the variability in responsiveness and cost between different individuals. Modelling of stochastic uncertainty is necessary in micro-simulation models, but not macro-simulation models.

Thus, *input parameter uncertainty* is the major focus in BODE^3.

We will use Monte Carlo simulation to model input parameter uncertainty. This involves thousands of iterations of the model calculations. Each iteration involves a random draw of a value from the probability distribution specified for each input parameter about which there is uncertainty. For example, we may specify a normal distribution about the natural logarithm of a rate ratio (ln(RR)) for change in a mortality rate. Each draw has a maximum probability of being close to the best or central estimate of the ln(RR), but it might also be one, two or more standard deviations above or below the central estimate with a probability proportional to that under the normal curve. Having run thousands of iterations, we then have thousands of (paired) values of both the cost and HALY, giving an uncertainty range for the cost-effectiveness ratio (or the net benefit, etc.). A 95% uncertainty interval would be bounded by the 2.5 and 97.5 percentile values across all iterations.

It is then possible to determine which of those input parameters with uncertainty confer the most total uncertainty in costs, consequence (i.e., HALYs gained) and cost-effectiveness estimates. Those input parameters (with their associated uncertainty) that determine the greatest amount of output parameter uncertainty can then be prioritised for further attention and research. Tornado plots are often used for this type of analysis.

Note that heterogeneity is a separate issue, and relates to expectations of costs and consequences that systematically vary by variables such as sex, ethnicity, age, and so on, and sometimes by variables we do not understand but we assume exist (e.g., tumour progression by some genotype). The BODE^3 models are highly disaggregated, and hence allow modelling of cost-effectiveness for many (heterogeneous) populations. But parameter uncertainty will still exist within sub-populations, requiring uncertainty analyses by sub-population.

There are many input parameters for the modelling about which there is likely to be uncertainty. For assessing total parameter uncertainty, it is the uncertainty about the *difference* between the baseline (or partial null) to intervention scenario that matters most. For example, the difference in disease mortality between no intervention and intervention A. Or the difference in downstream costs between no intervention and intervention A. In this instance, one may not need to allow for uncertainty in the baseline or partial null parameters (i.e., they are treated as certain or 'fixed'), and only allow for uncertainty in the intervention scenario (which is equivalent to just allowing for uncertainty in the difference between the partial null and intervention).

However, there may be instances where it is important to model uncertainty in both the partial null and intervention states. For example, the difference in mortality rates between

the partial null and intervention scenario may vary in proportion to the partial null mortality rate (which is also prone to uncertainty). That is, the expected value (and distribution) of the parameter estimate for the intervention is correlated with the partial null mortality rate (which is uncertain). However, it is unlikely that this situation will both arise and be of a magnitude sufficient enough to worry about on many occasions. For example, in specifying a partial null one is essentially calculating back with the same intervention assumptions (e.g., effectiveness, adherence, coverage) as for the interventions that will then be evaluated. Thus, it is the same uncertainty that is being used to strip the model back to the partial null, as it is for then modelling these interventions. Therefore, we do not anticipate having to often explicitly model parameter uncertainty for both partial null and intervention states – just the intervention parameters themselves.

In health economics, analysts usually use confidence intervals from other research studies or meta-analyses to specify probability distributions about input parameters. This focus on only random variation may be due to the heavy reliance on randomised trial evidence where, theoretically, systematic error is less problematic than in observational studies. Epidemiologists on the other hand are increasingly focusing on systematic error (i.e., selection, information and confounding biases), and using methods (e.g., quantitative bias analyses, including probabilistic sensitivity analysis methods) to address residual systematic error.[2 3 46] A methodological objective of this BODE^3 programme will be to bring these two approaches together, and more explicitly model both random and systematic error.

1.3.10 Equity analyses

Our main focus will be Māori vs non-Māori comparisons, but differences by level of deprivation, gender and age may also be analysed. An aim will be to ascertain the relative advantages and disadvantages of various methods for equity incorporation, both for academic and policy communities, ranging from simpler to more complex methods (as ordered below):

- Separate modelling by social group, e.g., Māori-specific patient navigator programme. If the 'targeted' programme is cost-effective, then it is both equity promoting and cost-effective on a total utilitarian perspective.[47] However, there will often be equity-efficiency trade-offs, hence the options below.
- Presenting HALYs gained separately by social group. This can be presented as a ratio to the 'standard' total DALY from a BDS for that social group (as a marker of 'need'), to determine whether HALYs gained are accruing more (or less) to those with the greatest disease burden. (At the time of writing this version of the Protocol, it seems that this option may be the most useful. It draws on the rich heterogeneity intrinsic in BODE^3 models, and the strengths of being linked to a BDS).
- We will trial measures of cost expressed per unit change in absolute difference in HALYs gained. For example, if the total population cost was $1 million, but for Māori 0.1 DALY per capita (age-standardised) was averted compared to 0.08 for non-Māori, then this equity-change ratio is $1m/ per 0.02 change in DALY difference between Māori and non-Māori.
- Although not fully developed, equity-weighted benefit measures (either QALYs (quality adjusted life years) or HALYs) are the most commonly anticipated method of quantitatively

incorporating equity.[1 47-56] However, empirical analyses are rare – probably due to a lack of high-quality data, which is where BODE^3 has an advantage. We will make a contribution to this literature by examining what happens to cost per HALY rankings if HALYs gained for (say) Māori/young are weighted by 10% to 50% more than non-Māori/old.

In sum, our equity work will respond to the emerging call internationally for explicit equity methods that inform rather than obfuscate.[57] We will test the utility of such methods with policy-makers. The final shape of our approach to equity analyses is not yet known, and is the focus of a PhD within the programme.

1.3.11 Effectiveness and consequences of interventions, and classification of the strength of evidence

The determination of effectiveness of interventions, and hence impact on consequences in the modelling, is of critical importance. Exact methods of information synthesis (e.g., literature reviews, meta-analyses, expert consensus) and parameterisation for modelling will be detailed later in this protocol (Section 3) and in a subsidiary Technical Report[14]. Here, three general issues are considered:

- How is the intervention's impact on HALYs gained in the model going to be parameterised, and what 'link' models are required to do this?
- General approach to synthesising evidence.
- Classification of strength of evidence, and implications for uncertainty analysis, sensitivity analysis and second stage filters.

1.3.11.1 How is effectiveness going to be parameterised?

Cost-utility analysis captures health consequences of interventions through changes in the QALYs or HALYs. Changes in the HALYs are driven by changes in years to live and any change in disability weights. These in turn are driven by changes in transition probabilities (or time to event) due to changes in death, survival and cure rates, disease incidence, remission, and so on, produced by the intervention.

However, the majority of the research evidence on effectiveness of interventions does not use these types of variables as the outcome. For example, much evidence of preventive programme effectiveness uses changes in risk factors, physiological markers or other intermediary outcomes. This requires 'link models' to convert research evidence into parameter changes for the economic decision models of BODE^3. The comparative risk assessment (CRA) models used in BDS are one example, whereby changes in risk factors are 'converted' to changes in incidence rates.

Such link models will often be needed, and may be intervention specific. Hence, the specification of the intervention, and how it will be parameterised in the model, is a critical first step to informing *what* research findings are going to be synthesised (and hence the nature of the literature search strategy).

1.3.11.2 General approach to synthesising evidence

A high quality systematic review and meta-analysis may take one analyst 6 months to a year. We do not have such resources to apply to each key input variable for modelling. Rather, a parsimonious and efficient approach will be required, guided by principles such as:

- making use of the growing number of international clearing houses for quality-appraised systematic reviews
- utilising evidence synthesis from ACE-Prevention (Australia), the UK National Institute for Health and Clinical Excellence (NICE) and elsewhere
- using other published systematic reviews that meet quality appraisal criteria.

Where there is an appropriate good-quality systematic review available, we may update it, and where necessary tailor it to the NZ context. We will perform meta-analysis, as needed, to determine single point estimates (with variance) for the intervention's effectiveness or other relevant parameters. On occasion, we may also use quantitative bias analysis methods to allow for likely residual systematic error in point estimates (i.e., likely residual measurement error, confounding or selection bias).

Where there is not a suitable systematic review available, we will conduct an appropriate review *de novo*.

Given that many of the interventions that will be assessed in BODE^3 will not have randomised controlled trial evidence, there will not always be a strong evidence base (see following section). We may need to seek and quantify expert opinion. In such cases, there will be greater parameter uncertainty about the effect size of an intervention.

1.3.11.3 Strength of evidence, and implications for evaluation

Table 4 below presents the classification of strength of evidence used in ACE-Prevention (Australia), with some minor modifications for application to BODE^3 (consideration of confidence intervals rather than p values; mention of quantitative bias analysis). We will use this classification to structure our approach to evidence synthesis. It provides a useful overview of what constitutes evidence, and what constitutes better evidence.

As has been described in the context of the SIGN (Scottish Intercollegiate Guidelines Network) grading system, grading aims to differentiate between results based on strong evidence and those based on weak evidence. This does not reflect the importance of the finding, but is a measure of the accuracy of the estimates. For the policy-maker, the evidence grading will form part of the 'picture' in understanding the likelihood that the predicted outcome will be achieved if the recommendation is implemented. Of note, and as found in the ACE-Prevention (Australia) report,[27] it is often the interventions with level IV, indirect or parallel evidence that have the greatest health impacts and best cost-effectiveness.

Table 4: Approach to classification of strength of evidence in BODE^3. Source: Vos et al (2010) [27]

Conventional approach based on epidemiological study design *Evidence* from level I-III study designs*	**Additional categories to be utilised in BODE^3** *Evidence from level IV studies, indirect or parallel evidence and/or from epidemiological modelling using a mixture of study designs*
A. 'Sufficient evidence of effectiveness' *Effectiveness is demonstrated by sufficient evidence from well-designed research that the effect:* • is unlikely to be due to chance (e.g. 95% CI excludes the null);and • is unlikely to be due to bias, e.g. evidence from: – a level l study design; – several good quality level ll studies; or – several high quality level lll-1 or III-2 studies (from which effects of bias and confounding can be reasonably excluded on the basis of the design and analysis, and/or quantitative bias analysis).	**B. 'Likely to be effective'** *Effectiveness results are based on:* • sound theoretical rationale and programme logic; and • level IV studies, indirect† or parallel‡ evidence for outcomes; or • epidemiological modelling to the desired outcome using a mix of evidence types or levels. The effect is *unlikely* to be due to chance (the final uncertainty interval does not include zero, and there is no evidence of systematic bias in the supporting studies and/or quantitative bias analysis suggests bias an unlikely explanation).
C. 'Limited evidence of effectiveness' *Effectiveness is demonstrated by limited evidence from studies of varying quality that*: • the effect is probably not due to chance (e.g. 90% CI excludes the null); but • bias, while not certainly an explanation for the effect, cannot be excluded as a possible explanation (e.g., evidence from: – one level II study of uncertain or indifferent quality; – one level III-1 or III-2 study of high quality; – several level III-1 or III-2 studies of insufficiently high quality to rule out bias as a possible explanation; or – a sizeable number of level III-3 studies of good quality and consistent in suggesting an effect.)	**D. 'May be effective'** *Effectiveness results are based on:* • sound theoretical rationale and programme logic; or • level IV studies, indirect† or parallel‡ evidence for outcomes; or • epidemiological modelling to the desired outcome using a mix of evidence types or levels. The effect is <u>probably not</u> due to chance. But bias, while not certainly an explanation for the effect, cannot be excluded as a possible explanation. Would benefit from further research and/or pilot studies before implementation.
E. 'Inconclusive evidence of effectiveness' *Inadequate evidence due to insufficient or inadequate quality research.* No position could be reached on the presence or absence of an effect of the intervention (e.g. no evidence from level I or level II studies, and level III studies are available but they are few and of poor quality).	**F. 'No evidence of effectiveness'** No position could be reached on the likely credentials of this intervention. Further research may be warranted.

*Evidence classifications based on those of the Australian National Health and Medical Research Council.
I - Evidence obtained from a systematic review of all relevant randomised controlled trials.
II – Evidence obtained from at least one properly designed randomised controlled trial.
III-1 – Evidence obtained from well-designed pseudo-randomised controlled trials (alternate allocation or some other method).
III-2 – Evidence obtained from comparative studies with concurrent controls and allocation not randomised: cohort studies; case-control studies; or interrupted time series with a control group.
III-3 – Evidence obtained from comparative studies without concurrent controls: historical control studies, two or more single-arm studies, or interrupted time series without a parallel control group.
IV – Evidence obtained from case series, either post-test, or pre-test/post-test outcomes.
†Information that strongly suggest that the evidence exists (e.g. a high and continued investment in food advertising is indirect evidence that there is positive (but propriety) evidence that food advertisement increases sales of those products).
‡Evidence of intervention effectiveness for another public health issue using similar strategies (e.g. the role of social marketing, regulation or behavioural change initiatives in tobacco control, sun exposure, speeding).

1.3.12 Extrapolating intervention effects over time

Trials often only measure outcomes over a limited time period, while epidemiological and economic decision models need estimates of the true impact on disease outcomes and costs well into the future. This is a major problem in economic decision modelling. For example, does an obesity prevention programme in childhood alter people's future BMI trend or track, or does any effect completely disappear in 10 years? Does a new treatment producing 30% lower mortality in the first year result in 30% lower mortality continuously into the future, or only in the first year?

One option is to limit the modelling to the duration of the trial or observational data, but this does not adequately reflect reality (e.g., health improvements are likely to be underestimated). The alternative is to make necessary assumptions about the impact beyond the duration of the available evidence (i.e., to assume either a continued impact over time, a lessening of the impact over a period beyond the known impact time from underlying studies, or the abrupt disappearance of the impact).

The choice of whether to model future attenuation (or not) of treatment effectiveness will depend on:

- the intervention in question;
- discussions with technical experts; and
- the most plausible way of modelling.

Often, however, there is no clear choice and the solution we adopt is to present results as discrete scenarios using different choices as a sensitivity analysis. For instance, ACE-Prevention (Australia) assumed a best estimate of an annual decay of the impact of GP-mediated physical activity interventions of 50% - but then varied this between 0% and 100% in sensitivity analyses.

1.4 Advisory committees

Final autonomy about what BODE^3 evaluates, and how BODE^3 evaluates it, rests with the Director, co-Directors and named investigators of BODE^3. However, BODE^3 also has strong collaborative links with the Ministry of Health and the sector, and a range of advisory groups (Figure 5). The Programme Advisory Group (PAG) will meet annually to:

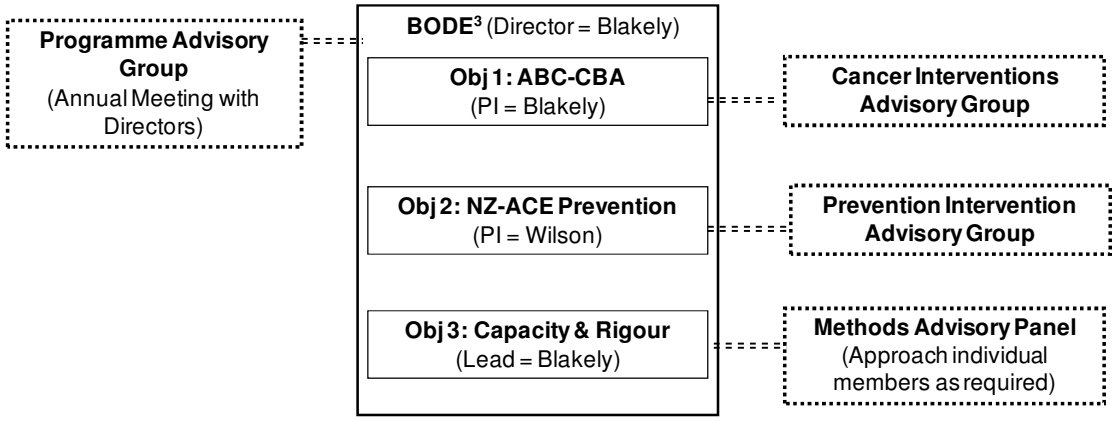
- Review and monitor strategic direction
- Advise on sector interface
- Provide overall advice on process
- Provide specific advice on aspects of prioritisation beyond the economic decision model (or what were termed 2nd stage filters in ACE-Prevention).

The membership of the PAG will include representatives from: Ministry of Health– including National Health Board (NHB); University of Otago; PHARMAC; and the National Health Committee. At the inaugural meeting of PAG in January 2011, members included:

- Deborah Roche (Deputy Director General, Strategy and System Performance Directorate; MoH)
- Teresa Wall (Deputy Director General, Māori Health Directorate; MoH)
- Dr Sharon Kletchko (General Manager Planning and Funding; Nelson Marlborough District Health Board)
- Rico Schoeler (Manager, Analysis and Assessment; PHARMAC)
- Scott Metcalf (Senior Advisor; PHARMAC)
- Kelvin Moffatt (Director, National Services Purchasing, National Health Board)
- Bridget Robson (Director, Te Ropu Rangahau Hauora a Eru Pomare; UOW)
- Prof Richard Edwards (HoD Public Health; UOW)
- Dr Darren Hunt (Deputy Director of Public Health; MoH)

At this inaugural meeting, advice was received on general issues, second stage filters, and intervention selection criteria. Content of this version of the Protocol reflects that advice.

Figure 5: Programme structure and advisory groups



At the time of writing this version of the Protocol, PAG is acting as the Prevention Intervention Advisory Group.

1.5 Other criteria for policy-making, or second-stage filters

The main focus of BODE^3 will be technical analysis (epidemiological and economic decision modelling, and quantification (where possible) of uncertainty and equity) and building capacity. The outputs will be academically relevant, but of course we also hope relevant and useful for policy-making in New Zealand (and perhaps elsewhere). Some institutions have formalised the role of cost-effectiveness analysis in decision making (e.g., NHS through NICE in the UK; PHARMAC in New Zealand). But no institution uses *only* cost-effectiveness analysis in decision making, and nor should it. There are other considerations, for example: equity (e.g., socioeconomic and ethnic, those in poor health compared to those in good health), age (e.g., the 'fair innings' argument [58]), total cost of the intervention, total burden of disease being addressed, total health impact, capacity of sector to implement intervention, political will, rule of rescue, and societal values.

In New Zealand these other considerations have been formalised by at least three influential institutions: the Ministry of Health, the National Health Committee and PHARMAC. The Ministry of Health's suggested criteria for prioritisation on the "Best Use of Available Resources" (2005) include three high-level principles: effectiveness, equity and value for money. Concepts of whānau ora from He Korowai Oranga [59] are also adopted, to ensure that Māori health is taken into account. For example, "Whānau ora means considering effectiveness, value for money and equity for Māori from a Māori perspective. It also recognises that prioritisation processes should enable Māori to participate in and contribute to strategies for Māori health improvement, and foster the development of Māori capacity to participate in the health and disability sector." [page iv, 60] Other considerations identified by the Ministry of Health included:

- the acceptability of the proposal, including the degree of acceptability to, and participation by, Mäori, other population groups and other stakeholders
- the ethical dimensions of the proposal
- the impact on the sector
- the ability to manage potential risks
- other legislative requirements.

PHARMAC[61] and the National Health Committee (NHC) [62] also have criteria, however they are similar to, precede, and form the basis of the above Ministry of Health criteria.

The ACE-Prevention (Australia) project developed 'second stage filters' that were applied to their main analyses, and have been included in both academic and policy outputs.[27] The core filters used in all ACE studies are:

- capacity of the intervention to reduce inequity
- acceptability to stakeholders
- feasibility of implementation
- strength of the evidence base.[d]

The ACE-Prevention (Australia) Project Steering Committee specified additional filters:

- sustainability; and
- potential for other consequences (side effects).

The Indigenous Steering Committee in ACE-Prevention (Australia) also specified two additional filters:

- cultural security; and

[d] The strength of the evidence base should be reflected in the uncertainty estimates about cost-effectiveness. Nevertheless, a separate itemisation of strength of evidence is often useful in its own right.

- community health gain.

It will be routine in BODE^3 for ethnic inequalities (and probably socioeconomic inequalities), to be explicitly and quantitatively included in analyses. Other equity considerations (e.g., age, gender, severity of illness) are beyond the scope of 'routine' BODE^3 output in the first instance.

Beyond BODE^3's focus on equity analyses, it is a moot issue how much the research should explicitly consider all these decision criteria. One could argue that it is not the role of researchers, but that of policy institutions and elected representatives. Alternatively, one could argue that as part of dissemination, translation and maximising the impact of research, researchers should actively engage in providing information for all decision criteria. Following consultation with the BODE^3 PAG, at the minimum BODE^3 outputs will provide:

- total estimated cost of the intervention, with uncertainty
- total estimated HALYs gained by the intervention, with uncertainty
- estimates of cost-effectiveness, with uncertainty
- a description of the relative magnitude of disease burden (both DALYs from the BDS and (where relevant) comparative risk assessment (CRA) output) that the intervention is addressing, capitalising on the burden of disease study (BDS) underpinnings of BODE^3
- and estimates of ethnic and socioeconomic equity impacts.

2 Selection of interventions to evaluate

The selection of interventions to evaluate is critical. For example, selecting very similar interventions for a narrow range of diseases would not achieve the goals of BODE^3 of comparing interventions across the health services and disease spectrum. There are perhaps three over-riding principles to consider:

1. Relevance. Selected interventions should inform decision-making in the next two to five years. This does not mean 'only select interventions that are likely to be directly considered by policy-makers in the near future'; there will also be a need for good comparators or benchmarks, and the need for academic innovation and agenda-setting.
2. Academic leadership. Academia has a position in society that allows it to lead policy thinking, and propose interventions that may be currently beyond what is considered viable by policy makers and society at large.
3. Academic rigour. The work must be academically rigorous. With respect to selection, this means selecting interventions that can be reliably specified and plausibly parameterised in terms of best estimates and uncertainty. To be clear, this does not mean excluding interventions that don't have (say) randomised trial evidence.

These three over-riding principles were reinforced by the BODE^3 PAG. Other general considerations included:

- Costs of interventions may rapidly change. Therefore, it may be warranted to revisit evaluations at times in the future.
- Developing infrastructure that allows rapid evaluations is highly desirable (and indeed is a motivation for BODE^3).
- Initially selecting some evaluations that can be rapidly undertaken as 'pilot' cases, perhaps focused around a risk factor or disease cluster.
- Retaining the focus on generality and allocative efficiency. For example, considering the viability of evaluations of classes of treatments rather than specific treatments per se.

The remainder of this section considers specific ABC-CBA and NZACE-Prevention criteria and processes for selecting interventions to evaluate, and a brief conclusion on issues to consider when 'specifying' the intervention.

2.1 ABC-CBA

2.1.1 Cancer intervention selection criteria

Individual interventions for ABC-CBA will first be selected on the basis of three key criteria:

1. Interventions which are likely to have a substantial impact on cost and/or cancer burden, and for which the evidence on programme effectiveness is likely to be in categories A, B, C or D shown in Table 4 (page 35).
    - For interventions already implemented (i.e., the policy question is potential disinvestment), weighting will be given to those interventions that are unlikely to

be cost-effective compared to standard thresholds or likely to be dominated by (new) alternative programmes addressing the same problem.

2. Interventions that do not meet the criteria above, but for other reasons (e.g., public pressure, political lobbying) are highly likely to require active decision making in the next 2-5 years, and where economic decision modelling is likely to be useful and influential in that decision making process.
3. Interventions satisfying either criteria 1 or 2 will be further weighted for inclusion if they are likely to have substantial equity relevance, where:
    - Dimensions of social group equity – in order of priority – are ethnic (Māori:non-Māori mostly), socioeconomic and regional.
    - Equity relevance includes:
        - Interventions designed for particular social groups (e.g., Māori patient navigators)
        - Total population interventions that address a cancer that varies in incidence, survival or mortality between social groups
        - Total population interventions that are likely to have varying cost-effectiveness between social groups.

The final set of cancer interventions will include:

- a range of interventions across the cancer spectrum from prevention through to palliation, and across a range of cancer sites;
- some interventions that can be treated as validation case studies (e.g., drugs previously assessed by PHARMAC, prevention interventions also assessed within NZACE-Prevention prevention)
- some interventions for cancers that specifically address Māori / non-Māori health inequalities
- for the first round of analyses, clusters of interventions for the same cancer (to facilitate assessment (and validation) of methods during the early stages)

2.1.2 Cancer intervention selection process

The selection process will be primarily undertaken by the BODE^3 researchers and the Cancer Interventions Advisory Group (CIAG; Figure 5 page 37). Additional input for interventions to be included will be sought from PAG and others as appropriate.

Members of the Cancer Interventions Advisory Group (CIAG) will be asked to individually identify interventions that meet the above criteria. The actual process of selecting and finalising the interventions will be a series of iterations backwards and forwards between the BODE^3 researchers and the CIAG:

1. Initial horizon scanning and preliminary selection. [Completed in 2009/10 as part of the HRC submission. See Table 5 below.]

2. *BODE^3 researchers*. Summary, assessment and analysis of the preliminary interventions in Table 5 below against the above criteria. Tighter specification (and options) of the actual interventions.[e] Possible addition of other interventions.
3. *CIAG*. A meeting of the CIAG will be held to discuss the above assessment and analysis. Particular input for Criteria 2 above (i.e., likely policy relevant in next 2-5 years). Second round of scanning and suggestion of potential interventions (e.g., based on recent clinical conference updates). Revised and prioritised list agreed.
4. *BODE^3 researchers*. Final assessment and analysis of intervention selections, and recommendations to CIAG. Consult more widely on this final proposal with PAG and others (as agreed by CIAG).
5. *CIAG*. Minor modifications only. Approval of list. Discuss and agree to role of CIAG as potential consultative body (e.g., technical advice on parameters for some interventions) during actual analysis, interpretation and dissemination.

A second round of this selection process above will be undertaken to select interventions for modelling in years 3 to 5 of the programme (i.e., the above applies only to the first round of selection).

Table 5: Preliminary selection of cancer control interventions – the ten non-italicised interventions were prioritised and included as 'preliminary' in the HRC application.

Domain	Intervention	Comments
Prevention	1. Doubling number of calls to Quitline (for Māori and non-Māori analysed separately)	Cross-reference with NZACE-Prevention. Includes Māori-focused mass media campaign.
	2. Reducing tobacco imports by 10 % per yr for 10 yrs → <2% smoking prevalence.	Equates to assessment of total tobacco-cancer burden. Societal costing necessary.
	3. Ensuring target 80% of smokers receive ABC (Ask, Brief Advice, Cessation).	Good trial of a systems intervention. Will be a challenge to cost.
	4. *Low-dose aspirin of cancer prevention*	*Possible to compare with NZACE-Prevention*
Screening/early detection	5. Colorectal cancer screening programme.	Validation with results from Ministry, and possibly MoDCONZ project (led by A/P Sarfati).
	6. Instigation of prostate cancer screening using PSA testing.	*Useful test of ABC-CBA ability to triage interventions. Shifts in disability weights (or utilities) will be important.*
Diagnosis and Treatment	7. Trastuzumab (Herceptin) for 12 months for early breast cancer.	Lots of data, and validation with PHARMAC and Ministry.
	8. Bevacizumab (Avastin™) for lung cancer.	Increasing amount of data, validation with PHARMAC, lung cancer theme for

[e] Specification should include: the target population; specific technologies used; the type of personnel delivering the service or treatment; the site of delivery; whether the service is bundled or piggy-backed with other services; and the timing of the intervention.

Domain	Intervention	Comments
		equity analysis.
	9. Ensuring all rectal cancer patients receive surgery in specialist unit/team.	Little scrutiny of effectiveness of surgery compared to drug therapy.
	10. As above, but for lung cancer patients.	*Ditto.*
Support and rehabilitation	11. Improved adult survivorship care, as per late effects programme used for adolescents.	System intervention. Under consideration.
Palliation	12. Radiation for prostate cancer bone pain, including fractionation options.	RCTs emerging on radiotherapy.
	13. Community care (e.g. Liverpool Pathways of Care model)	*Currently being considered. Some evidence from UK.*
Pan-Spectrum	14. Scaling up of patient navigators.	Challenging, but current. Trial scenarios that estimate the improvement in HALYs that would be necessary to justify the cost, rather than CEA per se.
	15. Enhanced patient management support and/or IT support (e.g. Map of Medicine)	

2.2 NZACE-Prevention

The selection of interventions in NZACE-Prevention differs from that for ABC-CBA in the following ways. First, domains of priority *risk factors* were selected. This follows the modelling approach in ACE-Prevention (Australia), which uses CRA approaches, making it most sensible and pragmatic to evaluate interventions by domains.

Second, a list of potential preventive interventions were collated for further stakeholder assessment. Unlike ABC-CBA, we did not rely on a specific advisory group for three reasons: the BODE^3 researchers have substantial expertise in preventive interventions; the BODE^3 PAG also has substantial experience with preventive interventions, and we drew on this advisory group as required rather than a specific prevention advisory group; and the previous experience of ACE-Prevention (Australia) researchers (e.g., Vos, Barendregt, Cobiac) included in the NZACE-Prevention team gives a stronger departure point.

2.2.1 Risk factor selection

The process of selection of risk factor topics for NZACE-Prevention is described in detail in an online Background Paper on Risk Factors[9] (www.uow.otago.ac.nz/BODE3-info.html) and elsewhere.[63]

To be considered, a risk factor had to be in the top 15 for causing lost DALYs for high-income countries (from recent WHO work on DALYs).[64 65] The following steps were applied to further select and prioritise the various risk factors:

- The risk factor had to be amendable to at least one preventive intervention for which there was a good evidence-base for effectiveness and likely cost-effectiveness.
- The risk factor had to contribute to health inequalities in the New Zealand setting in terms of the gap between Māori and non-Māori.

- The risk factor was given less priority if study of the effectiveness and cost-effectiveness of preventive interventions would be particularly demanding because of the need for complex new burden of disease data.

The final prioritised risk factor list is shown in Table 6 below.

Table 6: Final prioritised list of the top six major risk factors for evaluation in NZACE-Prevention

Risk factor (prioritised order)	Rationale and comment	Relevance to ongoing research around cancer control (ABC-CBA)
Highest priority		
Tobacco use	A major cause of disease burden and especially of inequalities in the NZ setting.[f]	High overlap given the number of tobacco-related cancers
High blood pressure	A more important cause of lost DALYs than cholesterol, contributes to inequalities, and many effective interventions are available.	Small overlap: salt intake is a joint risk factor for high blood pressure and stomach cancer
High cholesterol	This risk factor was upgraded in priority because interventions appear more promising than for most other risk factors in this list (and there is some overlap with the blood pressure interventions if an absolute risk approach is adopted e.g., for considering a polypill intervention).	Nil
Medium priority		
Alcohol use	This risk factor is important but is complex to study as there are over 200 ICD-10 three-digit disease codes in which alcohol is part of a component cause. Intervention analyses therefore should follow the completion of the NZ Burden of Disease Study revision.	High overlap given the number of alcohol-related cancers
Overweight and obesity	An important risk factor, but there is uncertainty around the persistence of intervention effects.	Some overlap given that obesity is a risk factor for some cancers
Physical inactivity	An important risk factor but the possible impact on health inequalities is indirect and there are uncertainties around the persistence of intervention effects (especially for interventions applied to children).	Some overlap given that physical inactivity and obesity are risk factor for some cancers
Lower priority		
Low fruit and vegetable intake	This risk factor is ranked relatively low as past work may have over-estimated the benefits of its reduction given the findings in a recent and very large cohort study.	Modest overlap given a potential role for low fruit and vegetable intake as a risk factor for certain cancers.
High blood glucose	This risk factor is of relatively lower priority given that interventions addressing blood glucose	Nil

[f] We note however, that ACE-Prevention (Australia) work on tobacco control intervention modelling is unlikely to be completed before mid-2012.

Risk factor (prioritised order)	Rationale and comment	Relevance to ongoing research around cancer control (ABC-CBA)
	directly are not particularly cost-effective. Also this risk factor will be partly addressed by considering other risk factors e.g., "physical inactivity", "overweight and obesity" (see above) and possibly vegetable intake.	

2.2.2 Specification of interventions, and groupings of interventions, for NZACE-Prevention analyses

The process for selecting potential interventions for NZACE-Prevention modelling work is detailed in the *ACE-Prevention Interventions Report*.[10] To summarise, the list of potential interventions was based on considerations of likely effectiveness, likely cost-effectiveness and likely potential to reduce health inequalities. In particular they arose from the following:

- Details from published ACE-Prevention (Australia) interventions (especially the September 2010 Report[27]). Relevant aspects of ACE-Prevention (Australia) work have also been published in the journal literature in such topic areas as: alcohol use,[66 67] overweight and obesity (particularly for children/adolescents), skin cancer,[68] pre-diabetes[69] and physical inactivity.[70]
- Knowledge of the ACE-Prevention (Australia) Team members (particularly in the area of tobacco control) and research around innovative edges of tobacco control.
- Frontier scanning for innovative interventions in cardiovascular disease prevention, particularly from Scandinavian countries (Google Scholar and Medline searches).
- Consideration of the literature around the co-benefits to health from climate change interventions which are likely to benefit cost-effectiveness from a societal perspective (e.g., active transport such as walking or cycling).
- Within-team review of the draft lists of possible interventions.

2.2.3 Process for Stakeholder Critique

Stakeholders (including PAG) were asked to critique about 20 specified interventions within each domain of the three highest priority risk factors: tobacco, high blood pressure and high cholesterol. Stakeholders were particularly asked to focus on the *relevance* of introducing the intervention in the NZ setting as the key criterion for stakeholder consideration. In particular, stakeholders were asked to:

- Tick the **top five most relevant interventions** (next 5 years in NZ), for each of the three domains of risk factors.
- Cross out the **five least relevant interventions** (next 5 years), within each of the three domains.
- Add **critical comment** if they thought there were problems with the argument for considering a particular intervention.
- Provide **new intervention ideas**.

Stakeholders found the task challenging. Feedback included a suggestion to merge some of the interventions across risk factor domains, consistent with many of the risk factors addressing CVD and an absolute risk approach to CVD. At the time of writing Version 1.0 of the Protocol, work on revising the selected preventive interventions is still proceeding. An initial focus on salt reduction interventions has been selected for a range of reasons, including being a suitable domain in which to train staff in the use of ACE multistate lifetables.

2.3 Specification of the Intervention

In order to determine health consequences and costs of an intervention, the intervention needs to be clearly specified. Key steps include:

1. Retrieving the relevant published studies (including those relevant to effectiveness, cost and cost-effectiveness; Section 3).
2. Summarising the basic activities for the intervention and key design issues in each of the key papers. (Note that there is a strong link here with the event pathway specification needed in costing (Section 4.4)).
3. Specifying that version of the intervention where 'best evidence' exists or those which are most appropriate in the New Zealand health service context.

There may be many versions of an intervention evaluated in the literature. It is beyond the scope of BODE^3 to perform extensive evaluations on different versions of a single intervention. Thus, the 'most promising' or 'prototypical' version of the intervention needs to be identified from the outset – namely that version with the best evidence base and which is the most appropriate from a policy standpoint. The weight put on these two factors can be somewhat flexible, but the intervention must be modelled to fit into the current New Zealand health service structure. An example might be where the intervention in question uses pharmacist prescribers to provide a service, but in New Zealand only prescribing by GPs would be acceptable. In this instance, we might model our intervention with GPs, but still use the efficacy data from the trial, along with documenting support for the assumption that GPs are as efficacious as pharmacist prescribers for the task.

3 Assessment of health consequences of interventions

Effect sizes and other relevant parameters of interventions investigated in BODE^3 will be derived from existing literature whenever possible (evidence levels I to III of the Australian National Health and Medical Research Council).However, many interventions will not have been subjected to rigorous evaluation, especially non-clinical interventions. When needed, BODE^3 will consider less robust levels of evidence, such as case series (level IV studies), indirect or parallel evidence and/or epidemiological modelling (see Table 4 page 35). We may at times need to involve expert opinion if existing data are not sufficient. Importantly, BODE^3 will generously specify the uncertainty surrounding such inputs, and additionally identify lack of good evidence as an issue in a brief second-stage filter stage.

These methods apply to determination of the effect size of an intervention, but may equally apply to other parameters of importance to the model. For instance, for some interventions, differences in toxicity may drive costs and outcomes more than difference in effectiveness. Literature searches may also be used to inform other parameters such as quality of life and resource utilisation.

Our aim is to determine the best estimate for input parameters for BODE^3 modelling. Thus, the focus is on finding the best evidence rather than all available evidence.

This section provides a brief outline of the methods that will be used (and need further developing) in BODE^3 for assessing health consequences of interventions, in three parts:

1. Higher quality or levels I to III evidence
2. Indirect and parallel evidence
3. Expert opinion.

3.1 Levels I to III: Literature synthesis

The first choice will be to use existing systematic reviews (including meta-analyses and health technology assessments [HTAs]) where an appropriate, high-quality review exists, and updating as needed. This approach is largely pragmatic. Systematic reviews are highly time and resource intensive. This would not be the best use of the project's resources when organisations such as the Cochrane Collaboration, the Campbell Collaboration and other HTA bodies have the resources to produce systematic reviews to very high standards with many internal quality assurance and peer review procedures and methods aimed at reducing bias. Use of existing systematic reviews is in line with the protocol of ACE-Prevention (Australia).[71] Furthermore, PHARMAC recognises high-quality meta-analyses and systematic reviews of RCTs as being the highest level of evidence.[72]

The approach that will be used for BODE^3 analyses can be conceptualised as an iterative process to which a common research question is applied. The first iteration will be to identify existing systematic reviews from HTA/systematic review sites and literature-indexing databases. If a systematic review is available, the inclusion/exclusion criteria will be applied to determine that it is relevant, and it will be subjected to critical appraisal.

Subsequent iterations will identify relevant literature to update the systematic review, or to conduct a *de novo* systematic review if there is no appropriate existing systematic review.

The literature searches and data extraction will be performed using systematic methods, including: clear definition of the research question; use of an explicit and documented search strategy; application of predetermined inclusion and exclusion criteria, and; critical appraisal of potentially relevant papers to identify the most appropriate research.

The detail of methods used for levels I to III synthesis are described in a separate BODE^3 Technical Report.[14]

3.2 Lower level evidence synthesis

The existence of a body of higher-quality evidence is not a prerequisite for an intervention to be evaluated in BODE^3. Rather, interventions are also selected for their policy and equity relevance (see Sections 2.1 and 2.2). The exact methods of assessing lower level evidence will vary depending on the context and intervention. This Section is therefore more about the principles than the exact method.

3.2.1 Level IV evidence

Level IV evidence is that obtained from case series describing post-test, or pre-test/post-test outcomes.[73] This information will be identified by the literature searches described earlier in this Section and in the Technical Report.[14] Where stronger evidence exists, level IV evidence will be given less weight or disregarded. However, where no stronger evidence is available, level IV evidence may contribute to the criteria for an intervention being considered to be "likely to be effective" or "may be effective" (see Table 4). These terms reflect the lesser certainty around data from level IV findings.

Such studies are not well controlled, and may be more susceptible to bias (particularly confounding and selection biases) and chance findings. Thus, level IV evidence must be considered in the context of other supporting factors such as theoretical rationale and logic.

3.2.2 Indirect and parallel evidence

Along with level IV evidence, indirect and parallel evidence may contribute to the evidence for effectiveness of an intervention (see Table 4). Indirect evidence is information that strongly suggests the intervention is *likely* to be effective, but intervention-specific evaluations have not been done, or at least not in the public arena. For instance, the food industry's willingness to invest in gaining the Health Foundation Food Tick provides indirect evidence that this strategy increases sales of their products. Parallel evidence is provided when effectiveness is established for an intervention that uses similar strategies. For example, the proven effect of tobacco tax and alcohol tax in reducing harm to health provides parallel evidence that an unhealthy food tax may be effective.

The quantification (including uncertainty) of interventions with only indirect or parallel evidence will be context specific. The rationale and workings will be clearly documented.

3.2.3 Expert knowledge and consensus

Given the nature of the interventions being investigated in BODE^3, it is anticipated that for some models expert opinion will need to be drawn upon. Expert opinion is a legitimate source of information for modelling parameters when there is not sufficient data from other valid sources.[74]

Guidelines have recommended various methods including Delphi, modified Delphi and nominal group techniques.[74] As has also been emphasised by others[74] the methods used in BODE^3 will not force a consensus because we want to capture the variability in opinions and investigate (through uncertainty analysis) the impact that this diversity has on the model outcomes. The simplest way to look at this is to average the parameter estimates from the individual experts, and use the lowest and highest estimates to represent the variance for uncertainty analysis. It would also be possible to use the estimated extremes to test 'best' and 'worst' case scenarios in sensitivity analyses. However, these approaches provide little information on the distribution of the uncertainty surrounding the estimate.[75] It may be possible to determine a statistical distribution to the range of values provided (e.g., using bootstrap methods), and test uncertainty e.g., with probabilistic sensitivity analysis.

There is no single consensus on the optimal number of experts from whom opinions are elicited, but existing literature tends to suggest that 8–20 subjects is reasonable.[74 76 77] A larger number than this is likely to incur extra expense and time with no significant change to the findings.[76] Indeed, in early work within BODE^3 (e.g. expert knowledge to specify model structure and parameter values for care coordinators in stage III colon cancer) it was often found that little was gained after the consultation of the first two or three experts.

There is also no consensus on the exact criteria that should be applied in selecting experts, and it is to a large extent a matter of judgement by the principal investigators. The subjects being knowledgeable in the topic area is essential, but may not be sufficient in itself.[77] In line with others' recommendations,[74 77] diversity of experts is considered a key requirement for BODE^3.[77] It has been suggested that subjects should also be respected opinion leaders, be from diverse practice settings and geographical areas, and ideally should include representatives of the key policy-makers and stakeholders in the topic area.[74 77], with focus on incorporating different stakeholder viewpoints, e.g., clinicians, Ministry of Health, DHB, and Māori Health advocacy.

The methods used to identify experts and methods used to elicit their opinions must be documented clearly.

BODE^3 will aim to explore methods of expert knowledge elicitation (EKE). It will, however, also use a parsimony and time efficiency approach. That is:

- Rapid specification of models, with 'generous' uncertainty assigned to input parameters with lower levels of evidence or needing EK elicitation.
- Determination of which input parameters are driving model output uncertainty (e.g. using Tornado plots)

- Return to more formal and rigorous EKE methods for only those input parameters with a major influence of model output uncertainty.

There is an overlap here with expected value of (partial) information here.[4]

4 Assessment of costs

In Section 1.2.5 we stated that the study perspective is (principally) that of the health system, involving C1,C3, S1 and S3 costs as per Drummond's framework.[1] This Section provides more detail on the approach to costing that will be used in BODE^3, in both NZACE-Prevention and ABC-CBA. More detailed Technical Reports regarding cost methods are also available and should be referred to when undertaking costing for BODE^3.[13 19]

4.1 Overview of costing methods

- BODE^3 will treat intervention costs (C1, C2 and C3) and downstream cost offsets (S1 and S3) separately.
- Intervention costs will primarily be estimated by standard activity costing methods using event pathways and patient flowcharts, but other macro- and micro-costing methods will be used as required.
- Health system costs in the baseline model, and thence cost offsets, will primarily be estimated by a top-down approach. This requires a source of total or average costs for each disease state by sex, age, and, if possible, ethnicity and deprivation (and possibly cancer stage or severity). We will principally use routine Ministry of Health data for this purpose (see Section 4.5).

BODE^3 will measure costs in economic terms, i.e., the market costs of the resources consumed, as a substitute for full and proper measurement of opportunity costs. This approach involves three basic steps:

- Identifying what costs are to be included
- Measuring the resources consumed (or saved) with and without the intervention
- Valuing these resources.

Costs are to be measured in 'real' dollars. That is dollars in costs and prices at a specified date. Costs will be valued at the 'market value' of the resources involved. Therefore, if 'purchase cost' data that is more readily available for actual costing deviates from the market value (e.g., because of subsidies on GP consultations or pharmaceuticals) the cost should be adjusted to equate more closely to the 'market value' (e.g., by adding the subsidy amount to the purchase cost). Where there is no market to provide market prices, and the cost is important for overall estimates of cost-effectiveness (e.g., > 10% of net cost), sensitivity analyses about the market value of the item will be undertaken.

Importantly, the intervention will be modelled as part of the *current* New Zealand health system. Even though the project reference year is 2011, from a costing perspective, this simply means that all prices are expressed in 2011 dollars. The organisation of health services must reflect current practice. For instance, if a drug has been recently listed on the Pharmaceutical Schedule, but was not listed in 2011, then this drug should be priced as listed on the Schedule with prices deflated back to 2011. Where subsidies on doctors' visit and prescription fees, for example, have changed since 2011, current subsidy levels will be applied.

The direct costs of the intervention will be calculated by standard activity costing methods based on creating event pathways and patient flowcharts, as was done by ACE-Prevention (Australia) (Section 4.4). A top-down method of costing will be used for valuing cost offsets (Section 4.5).

Ideally, direct costing of the intervention will be disaggregated by sex, age, ethnicity, and deprivation and disease state when costing the intervention. However, this may either be challenging to implement (e.g., inadequate data) or conceptually erroneous (e.g., where an intervention is developed for the whole population, and cannot be divided across individuals).

Ideally, cost offsets may also be disaggregated by population heterogeneity. However, sex by age health system and population costs (and hence cost-offsets) only may be preferred for equity reasons. For example, HealthTracker data may find lower cost offsets for Māori with a given disease than should be the case, if utilisation relative to need is lower among Māori.

4.2 HealthTracker

HealthTracker is a New Zealand data tool that has become available relatively recently. The unique patient identifier (NHI – the National Health Identifier) now has coverage of at least 98% of the New Zealand population. HealthTracker links, by means of the NHI, health-care events occurring to any individual. The data are validated against PHO registers every three months. Reasonable data are available from July 2006, but are better from 2007 onwards. It includes only costs to the Government, and most is claiming data where the Ministry of Health pays for the service. The database contains information on those with an NHI number who have had any contact with health services in the last year. It is estimated that data are missing for about 37 000 people.

Data are linked from a number of existing databases, including: National Minimum Dataset (hospital events); National Non-admitted Patient Collection (outpatient and emergency department events); Laboratory Claims Collection; Pharmaceutical Collection; National Travel Assistance Claims; Primary Health Organisation Enrolment Collection; Mortality Collection; New Zealand Cancer Registry.

Costing data included in HealthTracker includes:

- hospital costs paid by the Ministry or DHBs (case mix cost weights)
- outpatient costs (contracted purchase units)
- GP visits (average capitation cost only, using enrolled capitation costs and funding formula (that includes ethnicity, NZDep, sex and age))
- general medical subsidy for visits to GPs outside of enrolled PHO
- client contract processing system that pays for Disability Support, and Health of Older People (HOP)

- emergency department triage level contracted purchase unit cost for event (costs differ by seriousness of presentation)
- community pharmacy, and more recently hospital pharmacy costs (excluding non-subsidised medications)
- laboratory tests funded by Vote:Health.

HealthTracker can track health-care events 'cross-sectionally', within a given year or other time-period, or 'longitudinally' linking to NHI numbers for earlier years in which NHI coverage was reasonably complete. Alternatively, for cancers, links can be made to the year in which the patient was diagnosed with cancer and was added to the Cancer Registry. Note that Cancer Registry data are generally robust from about 1996 onwards (after registration procedures were reformed in 1994).

HealthTracker is potentially a useful tool for improving cost data coverage, but what advantages it offers over other cost sources is still to be determined. Scoping the full potential of HealthTracker, and determining how it can usefully contribute to epidemiological and economic modelling, is an important aspect of BODE^3.

Outputs and processes for HealthTracker will be documented in technical supporting information to evaluations, building up a body of knowledge.

4.3 Criteria for inclusion and exclusion of costs

As previously stated in Section 1.2.5, the perspective is primarily that of the health system, including costs borne by the health sector and patients/families for both publicly funded (Vote:Health) and privately funded healthcare. The perspective may be broadened in specific cases where excluding other costs would substantially misrepresent the value of the intervention (see Section 1.2.5). The perspective we adopt is close to that used by PHARMAC,[61] and the ACE-Prevention (Australia) programme.[27]

The specific costs that are included and excluded are covered in more detail in the BODE^3 Technical Report on Direct Costing of Interventions,[19] and should be referred to when conducting costing for a BODE^3 analysis.

All set up and on-going running costs (over and above current practice) of the intervention will be included in direct costing from the point in time of *a decision being made to implement the intervention by Government.*

Health sector costs may include (as appropriate), but are not limited to, those relating to: hospitalisation, outpatient visits, surgical and other preventative or therapeutic interventions, emergency department, specialist visits, GP visits, nursing, allied health professionals, pharmaceuticals, diagnostic testing, laboratory tests, palliative care, and residential home care. For preventive programmes, costs may additionally include legislative costs, media costs, programme administration and running costs, training and recruitment costs etc. Capital costs and overheads are included.

As well as costs borne by the government health sector, costs to patients for private healthcare and out-of-pocket payments for visits to health professionals, pharmaceuticals and other miscellaneous expenses will be included (where practicable) for both intervention costs and cost offsets. The way in which the direct intervention costs for private treatment are incorporated will vary depending on the relevance to the intervention. Where the intervention is performed predominantly in the public sector and a patient would have incurred costs to the public sector if they had not chosen to have private treatment, the public health sector cost will be used. The public health sector cost better represents the true cost of treatment as it does not include the additional profit and likely higher capital and staff costs of private treatment. However, the proportion of patients that undertake private treatment may be important for scaling cost-offsets. The HealthTracker databases that will be used to calculate health system and population costs contain primarily those costs borne by the Vote:Health (government) budget; reporting of information by private providers is encouraged but not mandatory. Thus, total costs will be underestimated. This will be addressed by scaling-up costs to account for the proportion of privately treated patients whose costs have not been included in HealthTracker. Eighty percent of the New Zealand health system is funded publicly. The remainder is funded privately by patients (15%) or health insurance (5%).[78]

Direct costs of patient travel (e.g., petrol and vehicle running costs) but not patient time to travel are in scope.

Productivity costs (C4 costs) are outside the scope of our health system perspective. Furthermore, there is controversy over ‘double-counting’ and the best way of valuing any ‘lost contribution’. However, if there is a compelling case for their inclusion in specific analyses they may be included in sensitivity analysis. Similarly, unpaid caregiver costs for time spent caring for the patient are considered out of scope.

Even when a health system perspective is taken, the revenue gained from a tax intervention (e.g., higher taxes on alcohol, tobacco or unhealthy food) will not be included in the costs because it is a transfer payment. Likewise, income support payments such as Sickness or Invalids' benefits are not included, as they are 'transfer payments'. Finally, we do not intend to include so-called ‘dead-weight costs’ of any tax increases required to fund an intervention.

4.4 Intervention Costs

Costs of interventions will be costed directly. The default position is to use activity costing, with event pathways and patient flowcharts. Macro- and micro-costing will also be used as appropriate, depending on the individual intervention. For instance, macro-costing is appropriate for discrete events such as doctor’s visits and hospital stays. Micro-costing may be used when sufficient detail is available, such as individual patient data from a clinical trial.

Following standard costing methodologies[1] (including the approach of ACE-Prevention (Australia) [27]), pathways will be constructed to describe the major components of the

intervention, and the activities that occur within each component. Through this process, the resources consumed are identified, and quantity consumed is multiplied by the relevant unit cost. To avoid bias, the process for identifying costs should be identical for the intervention and comparator.

Well-defined steps for both the intervention and comparator will be constructed; further detail is supplied in a supporting Technical Report on Direct Costing of Interventions.[19]

4.5 Health System and Population Costs; Cost Offsets from Disease Averted/Incurred

As stated in Section 1.2.5, BODE^3 will include both related and unrelated health system costs in costs averted/incurred (i.e., S1 costs as per Figure 1, page 8).

Unrelated costs include costs from future admissions and treatment of diseases not targeted by the intervention, but arising due to prolongation of life. This has been a contentious issue, although there seems to be a growing theoretical consensus to include both related and unrelated costs.[79 80] Moreover, it is often empirically difficult to determine what is related and unrelated. For example, presumably only some of future heart disease treatment costs can be attributed to a smoking intervention, as heart disease is due to multiple risk factors.

Van Baal et al (2011) provide a useful framework that can be applied to BODE^3.[80] They define types of cost as follows:

- oc = direct costs of the intervention (C1 costs as in Figure 1)
- ac = average health care costs, by sex and year of age
- dc = average health care costs in last year of life, by sex and year of age
- sc = survivor average health care costs, by sex and year of age (i.e., for all but last year of life, and having excluded costs incurred in last year of life)

The most commonly used ICER in preventive interventions for intervention y compared to intervention x is:

$$ICER1 = \frac{[oc(y) + ac(y)] - [oc(x) + ac(x)]}{HALYs\ gained\ (y) - HALYs\ gained\ (x)}$$

An improvement on ICER1 that allows for costs in last year of life is:

$$ICER2 = \frac{[oc(y) + sc(y) + dc(y)] - [oc(x) + sc(x) + dc(x)]}{HALYs\ gained\ (y) - DALYs\ gained\ (x)}$$

This framework is being used within BODE^3.

4.6 Methodological Considerations

Details of costs methodologies are presented in the BODE^3 Technical Report on Direct Costing of Interventions,[19] A brief overview is provided here.

Average costs will be used for stand-alone, mutually exclusive programmes. Marginal costs will be used as appropriate for scaling up or down of interventions (when there is no substantive change in fixed costs), and for interventions that occur in series.

The reference year for cost values is 2011, in line with the burden of disease data being for the 2011 population. Costs before or after 2011 will need to be CPI adjusted to 2011 real costs. A discounting rate of 3% p.a. will be the default for both costs and benefits (see Section 1.3.8).

Goods and Services Tax (GST) is a 'transfer payment' and will therefore be excluded from all costs used in the cost-effectiveness analysis.

As a principle, incremental changes in overhead costs resulting from the intervention will be included. The default position for BODE^3 when a unit cost does not already include overheads is the application of a 50% overhead, consistent with PHARMAC.[72]

PART II: ABC-CBA METHOD AND PROTOCOL

ABC-CBA will be a platform of data and parameters on cancer incidence and survival, projected into the future, by cancer sites, sex, age, ethnicity and deprivation (and sometimes stage). This platform of data will be captured in the form of Markov and other (e.g. DES) models, with (for example) survival rates transformed into time-dependent transition probabilities of death from the cancer in question. Costs and disability weights are laid over the Markov model states, so that each state (usually of a month's duration) has an attributed cost to the health sector and a disability (or more generally utility) weight assigned. This platform will then be used for evaluations of proposed interventions, where the interventions are parameterised in terms of cost of the intervention, and changes in incidence, survival, disability weight, stage distribution and occurrence of sequelae. Initial ABC-CBA model development occurred as part of a Burden of Cancer study [81], conducted by Blakely, Costilla and Tobias during 2009-10 at the Ministry of Health. This study estimated the burden of cancers, using DALYs and for cancers *incident* in the year 2006 – not the burden of cancer cross-sectionally present in 2006. This focus on incident cancers was deliberate, both due to the good cancer registry data (for incidence and thence survival) and to allow an easier migration to economic decision modelling (i.e., ABC-CBA as described in this protocol). The burden of cancer estimates used an external model lifetable; for the economic decision modelling in ABC-CBA we will be using each population's own lifetable (e.g., Māori males) so that we evaluate that population's potential HALY change due to an intervention.

There are both similarities and differences between the ABC-CBA and NZACE-Prevention models. The similarities dominate, and include: the same 'Key Principles' and assumptions as outlined in Section 1 of this Protocol; same datasets and methods for costing; default Markov models, with strong reliance on lifetables;and an emphasis on a common epidemiological model structure (and inherent cost-offset model) that is utilised by all intervention modelling. The differences with NZACE-Prevention largely arise in the core models, and include:

- ABC-CBA analyses will model incident cancers in 2011 (and subsequent years for prevention and screening models; Figure 6) and their future stream of mortality (and hence YLLs) and morbidity (and hence HALYs gained). Some NZACE analyses (namely those using multistate lifetables) include prevalent disease as well.
- ABC-CBA makes much use of cancer registry data (incidence and survival estimates), and the University of Otago's CancerTrends study (for estimates of how both cancer incidence and survival varies by ethnicity and deprivation, and over time from 1981 to 2004). NZACE-Prevention makes much use of input parameters arising from a national revision of the BDS being conducted by the Ministry of Health in parallel to BODE^3.

The overall conceptual approach of the ABC-CBA models is outlined in Figure 6, sub-figures a through d:

a. Baseline model. The first task is to determine the baseline or 'business as usual' number and/or rates of:

- incident cases (by cancer site, sex, age, ethnicity, deprivation, and occasionally stage or other category of disease severity at presentation ('sub-type')) in 2011
- incident cases by calendar year after 2011 (arising from the closed cohort of the New Zealand population alive in 2011)
- survival and disease progression among the 2011 incident cases, and all other future calendar year cohorts of incident cases.

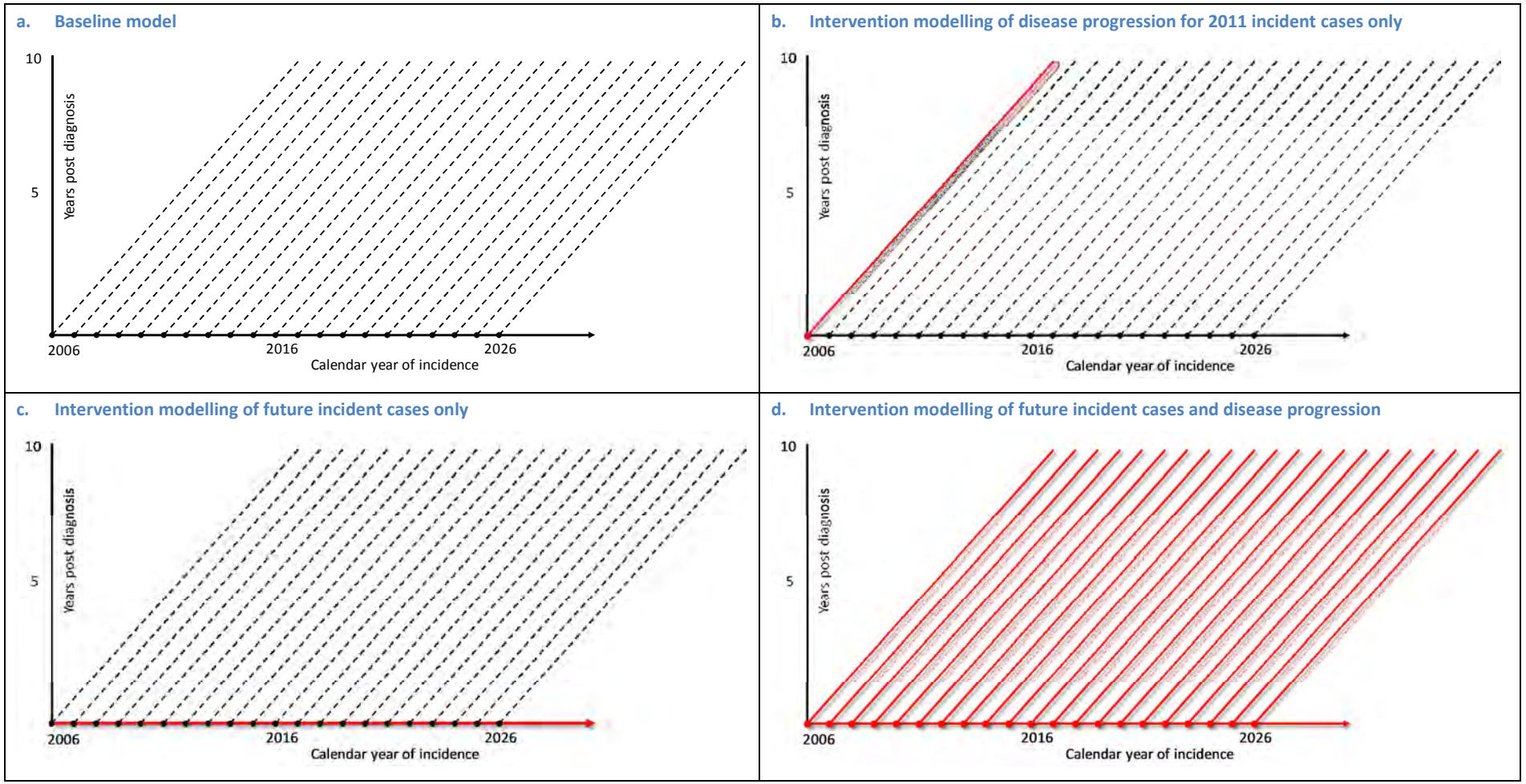
b. Treatment intervention models, that consider just 2011 incident cancer cases. Cancer interventions are modelled by changes in survival, disability weights and sequelae after diagnosis.
c. Prevention intervention models, that allow changes in future cancer incidence (including distribution or rates by stage or sub-type), and thus can model cancer prevention and screening and early diagnosis interventions. (It is assumed that there is no impact on 'business as usual' disease progression post diagnosis).
d. Both Treatment and Prevention models that merge both b. and c. That is, cancer control interventions that impact of both incidence and disease progression. (An example might a screening programme accompanied by improved treatment services).

This conceptual approach will be repeatedly used in this Part II of the Protocol.

The majority of this Part II of the protocol will assume that Markov modelling is used – both in the creation of the baseline models (Figure 6a), and intervention modelling (b, c, and d in Figure 6). However, we may use micro-simulation or discrete event simulation (DES) in some instances for modelling interventions (i.e., Figure 6b, and perhaps d), and some sections are included below on possible approaches.

It is useful to think of 'a' in Figure 6 (baseline models) as a large databank of input parameters for subsequent 'real' modelling of the interventions (models b, c and d). This applies regardless of whether Markov or other modelling of the interventions is used. This separation is also useful when one considers the enormous heterogeneity in the baseline models; we will often only model through a few 'types of people' in intervention modelling (e.g., 50-59 year old with regional stage cancer). That is, the baseline model will often provide a databank of parameters, from which only some are selected to go forward to intervention modelling.

Figure 6: Conceptual cohort approach to modelling in ABC-CBA



The remainder of Part II of this Protocol provides more detailed information on the assumptions, logic and development of the models necessary to achieve the modelling objectives of ABC-CBA.

- Section 5 describes the data inputs to the baseline or business as usual model. This includes both epidemiological inputs, and cost off-set inputs.
- Section 6 describes the approach to modelling interventions, most notably the parameters that need to be varied to model a cancer control intervention.

This protocol does not include details of the computer programs and code. Rather, this protocol serves as the higher-level specification.

5 Structure and data inputs for baseline models

5.1 Cancer sites

Table 7 shows the cancer site groupings and ICD10 codes used. In the majority of cases cancer site definitions were the same across the sources that were used to build the core ABC-CBA model. Some exceptions, and the ABC-CBA definitions, are:

- a narrower definition of brain, myeloma and ovarian cancer, to be consistent with survival analyses
- colon and rectal cancers were combined
- non-melanoma skin cancers were excluded.

Likewise, ethnic differences in incidence data were sourced from the CancerTrends study, which used the groupings shown in Table 7. When inconsistencies in site groupings occurred, the ethnic variations for the closest matching analysis in CancerTrends were used, or analyses rerun with different groupings if possible.

Existing Ministry of Health survival estimates were not available for gallbladder, bone and connective tissue cancers, 'other adult' cancers and childhood cancers. For these four sites, specific data was taken from the Cancer Registry and mortality files to allow excess mortality rate modelling; otherwise, the data sets already in existence for previous relative survival analyses [82] were used for modified excess mortality rate modelling.

Table 7: Cancer groupings used by studies relevant to core ABC-CBA model

Cancer	ABC-CBA	Ministry of Health cancer trends and projections [31 32]		CancerTrends (UOW,)[30]	NZHIS Survival [82]
	ICD10	ICD9	ICD10	ICD10	ICD10
All childhood	C00–C96 (< 15)	140–208 (<1 5)	C00–C96 (< 15)	C00–C97 (< 15) [g]	
All adult		140–208 (≥ 15)		C00–C97 D45 (≥ 15) g	
Bladder	C67	188	C67	C67	C67
Bone and connective	C40–41	170–171	C40–41	C40–4141	
Brain	C71	191	C70–72	C71	C71
Breast (female)	C50	174	C50	C50	C50
Cervix	C53	180	C53	C53	C53
Uterus	C54–55	182	C54–55	C54-55	C54–55
Colon	C18–21	153–154	C18	C18–20 [h]	C18–21
Rectum, sigmoid, anus[i]			C19–21		
Gallbladder	C23–24	156	C23–24	C23–24	
Hodgkin's disease	C81	201	C81	C81	C81
Kidney and other urinary	C64–66, C68	189	C64–66, C68	C64	C64–66, C68
Larynx	C32	161	C32	C30–32	C01–14,C32
Lip, mouth	C00-14	140–149	C00–14	C00–14	
Pharynx					
Leukaemia	C91–95	204–208	C91–95	C91-95	C91–95
Liver	C22	155	C22	C22	C22
Lung, trachea, bronchus	C33-34	162	C33–34	C33–34	C33–34
Melanoma	C43	172	C43	C43	C43
Myeloma	C90	203	C88, C90	C90	C90
NHL	C82–85, C96	200, 202	C82–85, C96	C82–85 [j]	C82–85, C96
Oesophagus	C15	150	C15	C15	C15
Ovary	C56	183	C56, C57.0–57.4	C56	C56
Pancreas	C25	157	C25	C25	C25
Prostate	C61	185	C61	C61	C61
Stomach	C16	151	C16	C16	C16
Testis	C62	186	C62	C62	C62
Thyroid	C73	193	C73	C73	C73

[g] CancerTrends input is the first diagnosed malignant cancer after each census date – which is much the same as any cancer.

[h] C21 (anal cancers) were not included in the HRC-funded CancerTrends grouping.

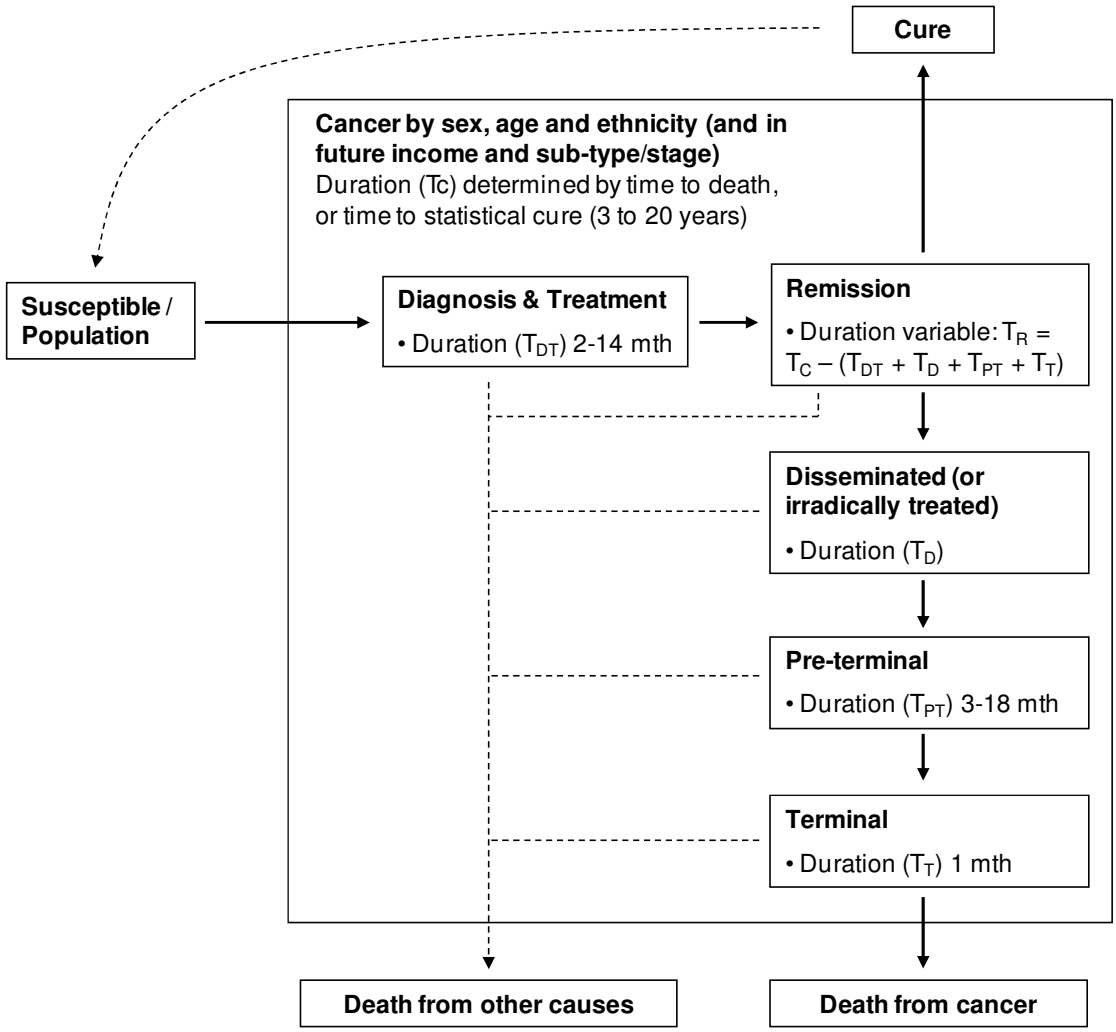
[i] ICD code for anus is C21.

[j] Regression analyses for a combined C82-85 and C96 group have been conducted on CancerTrends data, and are used in ABC-CBA.

5.2 Cancer disease model structure, cure rates and disability weights

For each cancer, a model of disease progression is needed to estimate the health adjusted years of life (HALYs) (and to allow for subsequent modelling of interventions that impact on morbidity rather than mortality). The general ABC-CBA disease model is shown below in Figure 7. Sex-, age-, ethnicity-, deprivation- (and occasionally stage/sub-type-specific) inputs for incidence and survival are specified for each cancer, using a common model structure (states, duration in each state, disability weight for each state, and sequelae). In Markov models the cycle length will be one month for treatment models, and perhaps one year (with averaging of some of the monthly parameters below) for preventive models. Also, for initial ABC-CBA modelling we set T_D (duration in disseminated state) to zero for all cancers, and incorporate this state with the pre-terminal state. The disseminated state is conceptually identified both to be consistent with other burden of disease models and to allow for flexibility in future scenario modelling.

Figure 7: General ABC-CBA cancer disease model



T_C = total cancer duration; T_{DT} = time in diagnosis and treatment state; T_R = time in remission state; T_D = time in disseminated state; T_{PT} = time in pre-terminal state; T_T = time in terminal state.

Note that cancer subjects can move to the 'death from other causes' state from any disease state within the model, based on the population mortality risks by sex, age, ethnicity and deprivation from life tables. There is also a parallel chain of states for those people who have permanent sequelae (e.g., leg. amputated due to bone cancer), but this is not shown in Figure 1.

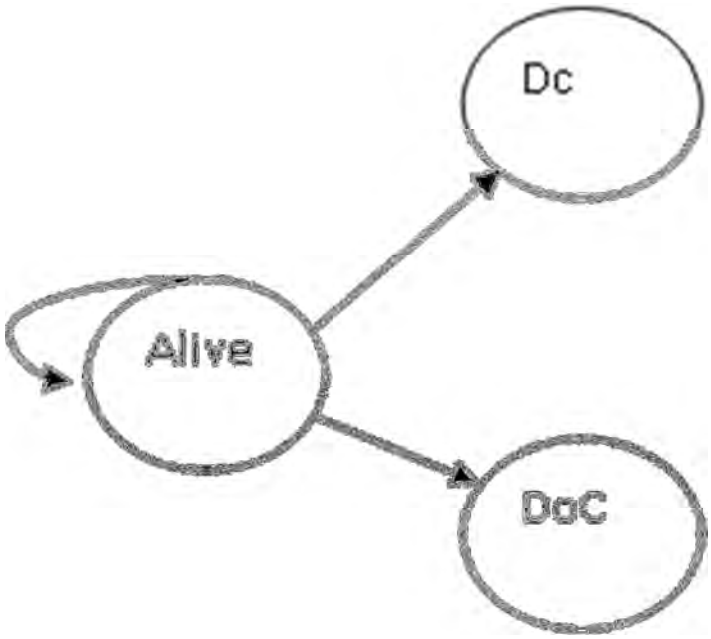
The maximum durations in each state (excluding remission) are given in Table 9 (page 66). They are maximums in that if a person dies relatively soon after diagnosis, they will not live enough months post diagnosis to traverse all possible states. Accordingly, the duration in the terminal state takes priority over that in the pre-terminal state, which in turn takes priority over that in diagnosis and treatment. The duration in the remission state is simply the residual of the duration in all other states, subtracted from the total cancer duration time (T_C). Table 8 gives examples of the time in different states for four scenarios. For example, a person dying of cancer after 6 months.

Table 8: Time spent in each state of the disease model (months), for different scenarios for a cancer with given maximal durations in each state as shown

Scenarios	**Total cancer duration: T_C**	**Time in state (months)**			
		Terminal: T_T (max = 1 mth)	**Pre-terminal: T_{PT} (max = 11 mth)**	**Diagnosis and treatment: T_{DT} (max = 6 mth)**	**Remission: T_R (duration = residual)**
Person dying of cancer at the end of year 3	36	1	11	6	18
Person dying of cancer at the end of month 6	6	1	5	–	–
Survivor	48 (cure time)	–	–	6	42
Person dying of a cause other than cancer at the end of year 3	36	–	–	6	30

Note that this has implications for how HALYs are calculated in ABC-CBA; they are tallied up 'backwards' once people have died from the cancer, died from other causes, or survived. That is, we do not actually model transition probabilities between these disease states (e.g., pre-terminal to terminal), but rather model transition probabilities to 'death from cancer' or 'death from other causes' at any stage from diagnosis to the cure time. For those transitions to 'death from cancer', we then back-calculate time and DWs in the various disease states. Likewise for 'death from other causes'.

Figure 8: Default ABC-CBA cancer model



5.2.1 Extending the cancer disease models to include stage/sub-type

The cancer disease models need to be re-specified and recalculated by stage (e.g., localised, regional spread, or metastasised; SEER or TNM classification) or subtype (e.g., size of tumour, if this is more relevant than stage) for colorectal, cervical and breast cancers, at least initially. These stage-specific models will be developed as required. However, the structure, maximum time in each state and DWs in each state are likely to remain unchanged. But because for advanced disease the excess mortality rate will be much higher (or survival much less), much of the time in Remission and Diagnosis and Treatment states will not occur due to displacement by Terminal and pre-Terminal states. Exceptions include Leukaemia, where type-specific (e.g. AML, CML, ALL and CLL) disease models were undertaken in the Australian BDS, but averaged below. Details on cancer-specific models can be found at: www.aihw.gov.au/bod-yld-by-disease/.

5.2.2 Duration and disability weights, by state, for each cancer model

Table 9 below lists the duration and OLD disability weights (DWs; i.e. those not revised to the GBD 2010 DWs) for each state and each cancer site included in the ABC-CBA baseline model, and as used in the *Burden of Cancer* report[81] (i.e., cancer not disaggregated by stage or sub-type). These durations and DWs in Table 9 are taken or adapted from the Australian BDS 2003 study [44 83], which in turn largely used the Dutch disability weights.[42 44 83] Adaptations made for the *Burden of Cancer* report are described elsewhere [81] (www.otago.ac.nz/wellington/research/bode3/abc-cba/index.html, pages 11-17). Most notably, they include some weighted averages of the DWs across stage or subtype distributions, to allow analyses aggregated by stage (or sub-type).

Table 9: OLD Disability weights (DW) and duration time (T, in years) for the disease model states used in the *Burden of Cancer* report ‡

Cancer site	Statistical cure time (years)	Diagnosis and treatment		Remission		Pre-terminal (including disseminated cancer)		Terminal	
		T_{DT}	DW †	T_R	DW †	T_{PT}	DW †	T_T	DW †
All childhood	5	0.67	0.66	Residual	0.20	0.50	0.75	0.08	0.93
Bladder	10	0.17	0.27	Residual	0.18	0.92	0.64	0.08	0.93
Bone and connective	10	0.50	0.41	Residual	0.30	0.92	0.75	0.08	0.93
Brain	5 (< 55 years); 10 (≥ 55 years)	0.25	0.68	Residual	0.18	0.67*	0.75	0.08	0.93
Breast (female)	20	0.33	0.29	Residual	0.26	0.92	0.79	0.08	0.93
Cervix	5	0.25	0.43	Residual	0.20	0.42	0.75	0.08	0.93
Colorectal	8	0.75	0.43	Residual	0.25	0.25	0.83	0.08	0.93
Gallbladder	7	0.17	0.43	Residual	0.20	0.92	0.73	0.08	0.93
Hodgkin's disease	10	0.33	0.66	Residual	0.19	0.42	0.75	0.08	0.93
Kidney and other urinary	10	0.17	0.27	Residual	0.18	0.92	0.64	0.08	0.93
Larynx	10	0.25	0.56	Residual	0.37	0.67	0.90	0.08	0.93
Leukaemia, < 45 years	10	1.17	0.55	Residual	0.19	0.25	0.75	0.08	0.93
Leukaemia, ≥ 45 years	10	0.50	0.55	Residual	0.19	0.25	0.75	0.08	0.93
Lip, mouth, pharynx	10	0.25	0.56	Residual	0.37	0.67	0.90	0.08	0.93
Liver	7	0.17	0.43	Residual	0.20	0.92	0.73	0.08	0.93
Lung, trachea, bronchus	6	0.42	0.70	Residual	0.47	0.42	0.83	0.08	0.93
Melanoma	6	0.17	0.22	Residual	0.19	0.25	0.81	0.08	0.93
Myeloma	20	0.75	0.19	Residual	0.19	0.42	0.75	0.08	0.93
NHL	20	0.33	0.66	Residual	0.19	0.42	0.75	0.08	0.93
Oesophagus	6	0.17	0.56	Residual	0.37	0.92	0.90	0.08	0.93
Ovary	10	0.25	0.43	Residual	0.20	0.42	0.75	0.08	0.93
Pancreas	5	0.17	0.43	Residual	0.20	0.92	0.73	0.08	0.93
Pleura, thymus, heart	5	0.25	0.35	Residual	0.30	0.67 *	0.75	0.08	0.93
Prostate	20	0.17	0.27	Residual	0.20	1.50	0.64	0.08	0.93
Stomach	6	0.50	0.53	Residual	0.38	0.92	0.73	0.08	0.93
Testis	3	0.25	0.27	Residual	0.18	0.75	0.64	0.08	0.93
Thyroid	5	0.17	0.27	Residual	0.18	0.75	0.64	0.08	0.93
Uterus	6	0.25	0.43	Residual	0.20	0.42	0.75	0.08	0.93
Other adult cancer**	10	0.35	0.44	Residual	0.24	0.66	0.75	0.08	0.93

* The Australian Burden of Disease study Excel spreadsheets state one-year duration on the flow diagram and 0.67 years in text notes. We have elected to follow the text notes.

** The duration and DWs for 'other adult cancer' are simply averages of the specified adult cancer sites.

† Note that as disease models are developed by stage or sub-type the 'fixed' DWs will be those at this most disaggregated level. The 'average' DW for all stages or sub-types combined may no longer exactly equal that reported in this Table. For example, if stage distribution in the future varies, then the weighted (by stage distribution) average of stage-specific DWs (fixed) will vary.

Regarding cancer states, only six health states had an estimated DW in the GBD 2010[41]. They are shown below in Table 10, including the descriptors that were provided in the surveys. Of note:

- No DW is provided for an equivalent of a remission state.
- No differentiation is provided between cancer types (e.g. oesophageal cancer is a far more debilitating disease in diagnosis and primary treatment than – say – myeloma).
- The DWs are generally lower – by about a third – than the old DWs as shown in Table 9 above.

Table 10: New GBD 2010 cancer disability weights (source[41])

Cancer State	Descriptor	Estimate	95% uncertainty interval
Cancer, diagnosis and primary therapy state [≈ DT]	*Has pain, nausea, fatigue, weight loss and high anxiety*	0.294	(0.199 to 0.411)
Cancer, metastatic [≈ PT]	*Has severe pain, extreme fatigue, weight loss and high anxiety*	0.484	(0.330 to 0.643)
Mastectomy	*Had one of her breasts removed and sometimes has pain or swelling in the arms*	0.038	(0.022 to 0.059)
Stoma	*Has a pouch attached to an opening in the belly to collect and empty stools*	0.086	(0.055 to 0.131)
Terminal phase, with medication (for cancers, end-stage kidney/liver disease) [≈ T]	*Has lost a lot of weight and regularly uses strong medication to avoid constant pain. The person has no appetite, feels nauseous, and needs to spend most of the day in bed.*	0.508	(0.348 to 0.670)
Terminal phase, without medication (for cancers, end-stage kidney/liver disease) [≈ T]	*Has lost a lot of weight and has constant pain. The person has no appetite, feels nauseous, and needs to spend most of the day in bed.*	0.519	(0.356 to 0.683)

For the purposes of ABC-CBA and BODE^3, we wished to retain the variation in DWs by cancer type following the Australian BDS framework as described above. Below we describe a process for generating DWs by cancer type and disease model state. It should be noted that the Ministry of Health's New Zealand BDS 2006 has undertaken a similar disaggregation, with the main difference being the focus on *prevalent* cancer as opposed to *prospective* or *incident* cancer in BODE^3. Otherwise, the Ministry's and BODE^3's approaches are very similar.

We now describe the process to estimate more specific DWs for cancer, as shown in Table 11.

We assumed that the relative difference in DWs (by DT, R, and PT disease states) in the Australian BDS (and as shown above in Table 9) still apply. The task then was to 'scale' down the DWs as shown in Table 9 to give the 'new' average DW as shown in Table 10. The average across all cancer types and states is given by the proportion of cancer cases diagnosed in 2006 entering each disease state, weighted again by the length of time in that disease state. For example, about 1.5% of all cancers in 2006 were myeloma. However, the DT disease state for myeloma (9 months) is longer than most cancers, meaning the myeloma's contribution to the average DT DW across all cancers is a 'weight' of about 3%. (The GBD DWs are still based on the prevalent concept of disease – not incident). The task therefore was to find the 'scale' factor that when multiplied into this 'weight' and 'old DW' for each cancer type, and then summed across all cancers gave the 'new average DW' as shown in Table 10. For the DT state, this scaling factor was 0.67; thus all new DWs for the DT states in Table 11 below are two thirds of the old DWs shown in Table 9 above.

A similar weighting process was undertaken for PT states, assuming that the distribution of 2006 diagnosed cancer cases reaching the PT state is simply that for those destined to die of their cancer. These numbers – allowing for background mortality risk – were intermediate outputs in the Burden of Cancer Report (Appendix B) [81], and were combined with cancer-specific times in PT states to generate the weights. Accordingly, the scaling factor was 0.65, such that the new PT DWs in Table 11 were scaled by 0.65 relative to the old DWs in Table 9.

The old DWs for the T state were 0.93 across all cancers – whereas the PT old DWs varied considerably by cancer. If we had used the above PT scaling method for T DWs, there would have been many instances where the T DW for a given cancer was less than the PT DW. (This is because the new overall cancer PT and T DWs were similar at 0.484 and 0.508). We decided having T DWs less than the PT DW for many cancers was incoherent. Therefore, we instead set the T DWs for each cancer equal to 0.508/0.484 (the ratio of average new T to PT DWs) multiplied by the new PT DW.

The GBD 2010 set of DWs excludes the remission state. Therefore, we estimated the R new DW for each cancer using the cancer-specific ratio of old DWs for DT to R, multiplied into the new DT DW. For example, the old bladder cancer DT and R DWs (Table 9) were 0.27 and 0.18, and new bladder cancer DT DW is 0.181. Therefore, the new R DW for bladder cancer is $0.18/0.27 \times 0.181 = 0.121$.

The new GBD 2010 DWs have been estimated with 95% uncertainty intervals (Table 10), using a logistic scale. We assumed the standard deviation (upper minus lower limit [on logistic scale] divided by 3.92) was directly applicable to the DT, PT and T cancer-specific DWs. For the R DWs, we used the standard deviation on the logistic scale for the DT DW. The resultant 95% uncertainty intervals are also shown in Table 11 below. Thinking ahead to their use in uncertainty analyses in BODE^3, four comments are justified here:

1. We have added model structure uncertainty through our range of assumptions above. Therefore, there is a case for further widening the uncertainty compared to using the standard deviations from Salomon et al. Conversely, though, one could argue that the DWs from Salomon et al are for such a wide and loosely specified set of cancers that uncertainty in calibrated specific cancers might be less. As a compromise, we have elected to assume the standard deviations from Salomon et al are directly applicable without scaling up. However, this could be probed in later sensitivity analyses if the baseline DWs drive considerable uncertainty in the model outputs (e.g. costs, HALYs and cost per HALY).
2. The 95% intervals presented in Table 11 are simply derived by converting the central estimate (which is a *mean*) to the logistic scale, and adding and subtracting 1.96 × standard deviation. For low DWs the average of this specified distribution on the logistic scale when transformed back to the DW or proportion scale is not exactly equal to the mean estimate given in Table 11. For example, the new DW for brain cancer R state is 0.121. That is -1.983 on the ln[odds] or logistic scale, with a standard deviation of 0.263 on the logistic scale. The average of 10,000 DWs calculated for random draws from this logistic normal distribution was 0.1236, 2% greater than the starting mean estimate of 0.1210. This is because of the slight right hand skew in the DWs for this parameterization when transforming back to DW or proportion scale. However, even for this worst case scenario (the lowest DW in Table 11), the error is negligible – especially in light of the structural assumptions made above.
3. The slight error noted above, one can specify these DWs on the logistic scale in economic decision models. However, it is also equally possible to fit a beta distribution.
4. The method and comments above apply specifically to the new cancer DWs, but are also applicable to non-cancer DWs in NZACE-Prevention – and any derivation of new DWs for NZACE-Prevention.

Table 11: NEW Disability weights (mean DW; approximate 95% uncertainty range †) and duration time (T, in years) for the cancer disease model states used in BODE^3, combining 2010 BGD disability weights and Australian BDS cancer disease models (see text for details)

Cancer site	Statistical cure time (years)	Diagnosis and treatment		Remission		Pre-terminal (including disseminated cancer)		Terminal	
		T_{DT}	DW	T_R	DW	T_{PT}	DW	T_T	DW
Bladder	10	0.17	0.181 (0.116 to 0.270)	Residual	0.121 (0.076 to 0.187)	0.92	0.416 (0.271 to 0.576)	0.08	0.422 (0.273 to 0.588)
Bone and connective	10	0.5	0.275 (0.184 to 0.388)	Residual	0.201 (0.130 to 0.297)	0.92	0.487 (0.332 to 0.645)	0.08	0.495 (0.334 to 0.656)
Brain	5 or 10	0.25	0.455 (0.333 to 0.584)	Residual	0.121 (0.076 to 0.187)	0.67	0.487 (0.332 to 0.645)	0.08	0.495 (0.334 to 0.656)
Breast (female)	20	0.33	0.194 (0.126 to 0.288)	Residual	0.174 (0.112 to 0.261)	0.92	0.513 (0.355 to 0.668)	0.08	0.521 (0.358 to 0.680)
Cervix	5	0.25	0.288 (0.194 to 0.404)	Residual	0.134 (0.085 to 0.206)	0.42	0.487 (0.332 to 0.645)	0.08	0.495 (0.334 to 0.656)
Colorectal	8	0.75	0.288 (0.194 to 0.404)	Residual	0.167 (0.107 to 0.252)	0.25	0.539 (0.379 to 0.691)	0.08	0.548 (0.383 to 0.703)
Gallbladder	7	0.17	0.288 (0.194 to 0.404)	Residual	0.134 (0.085 to 0.206)	0.92	0.474 (0.320 to 0.633)	0.08	0.482 (0.323 to 0.644)
Hodgkin's disease	10	0.33	0.442 (0.321 to 0.570)	Residual	0.127 (0.080 to 0.196)	0.42	0.487 (0.332 to 0.645)	0.08	0.495 (0.334 to 0.656)
Kidney and other urinary	10	0.17	0.181 (0.116 to 0.270)	Residual	0.121 (0.076 to 0.187)	0.92	0.416 (0.271 to 0.576)	0.08	0.422 (0.273 to 0.588)
Larynx	10	0.25	0.375 (0.264 to 0.502)	Residual	0.248 (0.164 to 0.356)	0.67	0.584 (0.424 to 0.729)	0.08	0.594 (0.429 to 0.740)
Leukaemia	10	0.75	0.368 (0.258 to 0.494)	Residual	0.127 (0.080 to 0.196)	0.25	0.487 (0.332 to 0.645)	0.08	0.495 (0.334 to 0.656)
Lip, mouth, pharynx	10	0.25	0.375 (0.264 to 0.502)	Residual	0.248 (0.164 to 0.356)	0.67	0.584 (0.424 to 0.729)	0.08	0.594 (0.429 to 0.740)
Liver	7	0.17	0.288 (0.194 to 0.404)	Residual	0.134 (0.085 to 0.206)	0.92	0.474 (0.320 to 0.633)	0.08	0.482 (0.323 to 0.644)

Cancer site	Statistical cure time (years)	Diagnosis and treatment		Remission		Pre-terminal (including disseminated cancer)		Terminal	
		T_{DT}	DW	T_R	DW	T_{PT}	DW	T_T	DW
Lung, trachea, bronchus	6	0.42	0.469 (0.345 to 0.597)	Residual	0.315 (0.215 to 0.435)	0.42	0.539 (0.379 to 0.691)	0.08	0.548 (0.383 to 0.703)
Melanoma	6	0.17	0.147 (0.093 to 0.225)	Residual	0.127 (0.080 to 0.196)	0.25	0.526 (0.367 to 0.680)	0.08	0.535 (0.371 to 0.691)
Myeloma	20	0.75	0.127 (0.080 to 0.196)	Residual	0.127 (0.080 to 0.196)	0.42	0.487 (0.332 to 0.645)	0.08	0.495 (0.334 to 0.656)
NHL	20	0.33	0.442 (0.321 to 0.570)	Residual	0.127 (0.080 to 0.196)	0.42	0.487 (0.332 to 0.645)	0.08	0.495 (0.334 to 0.656)
Oesophagus	6	0.17	0.375 (0.264 to 0.502)	Residual	0.248 (0.164 to 0.356)	0.92	0.584 (0.424 to 0.729)	0.08	0.594 (0.429 to 0.740)
Ovary	10	0.25	0.288 (0.194 to 0.404)	Residual	0.134 (0.085 to 0.206)	0.42	0.487 (0.332 to 0.645)	0.08	0.495 (0.334 to 0.656)
Pancreas	5	0.17	0.288 (0.194 to 0.404)	Residual	0.134 (0.085 to 0.206)	0.92	0.474 (0.320 to 0.633)	0.08	0.482 (0.323 to 0.644)
Prostate	20	0.17	0.181 (0.116 to 0.270)	Residual	0.134 (0.085 to 0.206)	1.5	0.416 (0.271 to 0.576)	0.08	0.422 (0.273 to 0.588)
Stomach	6	0.5	0.355 (0.247 to 0.480)	Residual	0.255 (0.169 to 0.364)	0.92	0.474 (0.320 to 0.633)	0.08	0.482 (0.323 to 0.644)
Testis	3	0.25	0.181 (0.116 to 0.270)	Residual	0.121 (0.076 to 0.187)	0.75	0.416 (0.271 to 0.576)	0.08	0.422 (0.273 to 0.588)
Thyroid	5	0.17	0.181 (0.116 to 0.270)	Residual	0.121 (0.076 to 0.187)	0.75	0.416 (0.271 to 0.576)	0.08	0.422 (0.273 to 0.588)
Uterus	6	0.25	0.288 (0.194 to 0.404)	Residual	0.134 (0.085 to 0.206)	0.42	0.487 (0.332 to 0.645)	0.08	0.495 (0.334 to 0.656)
Other adult cancer**	10	0.35	0.295 (0.200 to 0.412)	Residual	0.161 (0.103 to 0.243)	0.66	0.487 (0.332 to 0.645)	0.08	0.495 (0.334 to 0.656)
All childhood	5	0.67	0.442 (0.321 to 0.570)	Residual	0.134 (0.085 to 0.206)	0.5	0.487 (0.332 to 0.645)	0.08	0.495 (0.334 to 0.656)

* The Australian Burden of Disease study Excel spreadsheets state one-year duration on the flow diagram and 0.67 years in text notes. We have elected to follow the text notes.

** The duration and DWs for 'other adult cancer' are simply averages of the specified adult cancer sites.

† Uncertainty intervals use the estimated standard deviation on the logit scale from Salomon et al (2012) [41] about the logit of the central estimate. Note that the central estimate (as given in this table) may not exactly be the average on the logit scale given the standard deviation due to scale transformations; such error only become notable for DWs less than 0.15, and can usually be ignored.

Another issue is that of the DW in the remission state for cancers with a long statistical cure time. For example, breast cancer deaths can occur many years after diagnosis, and even at 20 years post-diagnosis the relative survival is still lower than that expected based on population mortality rates. For those women who survive to the nominated statistical cure time (i.e., 20 years), and even those who relapse or die of breast cancer (say) 10 to 20 years after diagnosis, it seems inappropriate to assume that the loss of quality of life and hence the DW is constant throughout the (up to) 20 years. Rather, it seems more realistic to assume that the DW reduces with each subsequent year of disease-free survival. (Other burden of disease studies circumvent this issue by assuming five years' duration for all YLD calculations, even if the statistical cure time is in excess of five years. However, we wanted a state-based model for future scenario modelling of interventions that may occur any time until 'cure' was pronounced). In the *Burden of Cancer* report we assumed that the remission disability weight reduces by 20% per annum from the first year onwards. We will continue to use this approach but we may need to consult more generally on this default assumption, and also undertake sensitivity analyses for this 20% (or whatever is decided) per annum reduction in the DW.

'New' GBD 2010 disability weights for the stoma and mastectomy sequelae are shown in Table 10 above; DWs for other sequelae are available in the 2012 Lancet paper.[41]

5.2.3 Cure times

Statistical cure times are also presented in Table 9. Determining the time of statistical cure is critical for a state-based model like ABC-CBA. (An alternative is to assume asymptotic reduction in excess mortality due to cancer, and use parametric functions (e.g., Weibull curves) to estimate survivorship. For flexibility of modelling we have elected to use a state-based model that requires specifying statistical cure times. However, if we do model some interventions with DES, then a parametric specification of survival would be desirable – see later sections of this Protocol). The methods used to determine statistical cure times are specified in detail in the *Burden of Cancer* report [81] (available at www.otago.ac.nz/wellington/research/bode3/abc-cba/index.html, pages 17 to 22). Briefly, the process required inspection of relative survival curves from New Zealand and Sweden, and determining the time after diagnosis at which the excess mortality from the cancer under question was trivial.

When stage or sub-type disease models are developed, consideration will need to be given to the possibility of cure time varying by stage or sub-type.

5.3 Data inputs to baseline models

Section 5.2 above describes the baseline disease model structure (and starting point at least for actual disease modelling). This section provides an overview of some of the baseline data inputs, i.e., as shown in Figure 6a. This will be done by first subtracting the excess mortality rate for the cancer in question (Section 5.3.3 below, and using 'business as usual' cancer

excess mortality rate rather than intervention simulated rate), then the using exponential link formula:

$$1 - s = 1 - \exp(-r_t T)$$

Importantly, detail on the derivation and specification of key baseline data is provided elsewhere in supporting Technical Reports, namely:

- Cancer incidence rates: BODE^3 Technical Report No 5 (www.otago.ac.nz/wellington/otago025053.pdf)[12]
- Cancer survival and excess mortality rates: BODE^3 Technical Report No 10[17]
- Lifetables for background mortality rates: BODE^3 Technical Report No 4[11] *(also applies to NZACE-Prevention)*
- Cancer incidence rates: BODE^3 Technical Report No 5 (www.otago.ac.nz/wellington/otago025053.pdf)
- Cancer survival and excess mortality rates: BODE^3 Technical Report No 10
- Lifetables for background mortality rates: BODE^3 Technical Report No 4 *(also applies to NZACE-Prevention).*

5.3.1 Cancer incidence

The base year is 2011, requiring estimated 2011 incidence rates by cancer site, sex, age, ethnicity and deprivation (and for some cancers stage or sub-type). This is a demanding task, but one for which we have reasonably good data in New Zealand.

Our general approach is to merge three sources of data:

- Ministry of Health estimated projections by sex and age
- CancerTrends estimated differences in incidence by ethnicity and deprivation
- Stage or sub-type distribution from Cancer Registry, international studies, or expert opinion.

5.3.1.1 Ministry of Health projections for 2011 by sex and age

The Ministry of Health has undertaken analysis on trends and projections of cancer incidence in New Zealand.[31 32] We use these estimates, rather than actual rates, for two reasons:

- They are smoothed, and therefore not prone to random variation in 2011. (For example, just by chance there may have been more or less than expected cases in certain strata in 2011).
- We also need to include future cancer incidence rates in the ABC-CBA model, which will come from these Ministry of Health projections. Therefore, it is consistent to also use the Ministry of Health 'projections' for 2011.

These incidence projection models are only by sex and age, due to a lack of ethnic and socioeconomic data back to 1950 – the data that was used to drive the projection models.

5.3.1.2 Incorporating ethnic and deprivation variation in 2011 incidence estimates, and beyond

In the previous Burden of Cancer report, we allowed for ethnic variation in cancer incidence rates by mathematically solving what the disaggregated 2006-2011 projected rates would be by ethnicity, using *published* ethnic incidence rate ratios from CancerTrends.[30] We will use a similar approach to generate cancer incidence rates by both ethnicity and deprivation for ABC-CBA, but apply a more generalised framework (utilising new regression analyses on CancerTrends data) that will also be applicable to 2011 and beyond projections (see below and [17]).

The goal of the regression modelling is to produce rate ratios for the joint distribution of ethnicity (Māori, non-Māori only) and deprivation (deciles 1-3, 4-7, and 8-10), by sex and age group. These rate ratios can then be combined with the above sex by age incidence projections, and the (projected) population distribution by sex and age, to generate cancer incidence rates by any combination of sex by age by ethnicity by deprivation (and by year for 2007 and beyond). Parsimony is important - it is the baseline model 'only' we are building (not the intervention effects per se). We followed a similar approach to that used for NZCMS analyses to build sub-population lifetables.[40]

Further detail can be found in Cancer incidence rates: BODE^3 Technical Report No 5 (www.otago.ac.nz/wellington/otago025053.pdf).[12]

5.3.2 Expected background mortality

See Section 1.3.4 (page 20) and BODE^3 Technical Report 4 [11] for the generation of background mortality rates. For the purposes of the Markov models, the background mortality rates need converting to transition probabilities.
Suppose $m_1(i)$ and $m_2(i)$ are death rates by cycle (assumed 1 month) from cancer and from other causes respectively. Then $m_1(i) + m_2(i) = m(i)$ is the total death rate. Consequently total death probability is equal to:

$$p(i) = 1 - \exp(-m(i))$$

Then, the probability of dying of cancer, $p_{12}(i)$ is equal to:

$$p_{12}(i) = \frac{m_1(i)}{m(i)}\, p(i)$$

and the probability of dying of other causes, $p_{13}(i)$ is equal to:

$$p_{13}(i) = \frac{m_2(i)}{m(i)}\, p(i)$$

The probability that an individual will stay alive, consequently, is equal to:

$$p_{11}(i) = 1 - p_{12}(i) - p_{13}(i)$$

5.3.3 Relative survival and excess mortality rate modelling

Background or expected mortality rates provide the transition probabilities for 'other causes of death'. Relative survival and excess mortality rate modelling provides the transition probabilities and time to event for 'deaths from cancer'. Detail on the derivation of excess mortality rates for use in ABC-CBA are available in BODE^3 Technical Report No 10.[17]

6 Modelling interventions

Detail is provided in BODE^3 Technical Report 17.[16] Modelling will either be Markov (macro-simulation or micro-simulation) or discrete event simulation. Simultaneous consideration of heterogeneity, parameter uncertainty and stochastic uncertainty will evaluation specific, using the frameworks outlined by Koercamp et al (2011).[45] Each evaluation is 'unique'; models will adhere to the principles in this Protocol, but necessarily have variations.

In some respects ABC-CBA models (at least the *baseline* variants) are self-calibrated, in that they are constructed with national incidence, survival and mortality data.

The process of model validation for *intervention* simulation models will follow standard process described elsewhere.[84]

PART III: NZACE-PREVENTION METHOD

7 NZACE: Overview of ACE-Prevention model

The ACE-Prevention (Australia) model is well established with its own protocol[71], peer-reviewed published methodological approach [8], numerous outputs with modelling-specific descriptions of methods [27 66 69 70 85 86], and existing Excel spreadsheet models and programs.

The epidemiological parts of ACE-Prevention model essentially works as follows:

- An intervention is conceptualised in terms of the population affected and some (counterfactual) change in the population distribution of a risk factor. This will often involve working risk factor data from surveys, and 'shifting' it to a new distribution based on a putative intervention.
- Just as DALYs are attributable to risk factors in comparative risk assessment (CRA), the counterfactual distribution of the risk factor is modelled through epidemiological relative risk functions by disease or injury to generate a change in the DALY distribution of the population. This change will usually be through the change in incidence of diseases, captured by multiple disease states in each sex by age group in lifetables. (Hence the term multistate lifetable). For example, a salt reduction intervention will require having the population disaggregated at each single year of life into those with and without IHD and stroke (based on incidence and case fatality rates that apply over the 'life' of the synthetic cohort in the life-table. Thus, a salt intervention will reduce the IHD and stroke incidence rates (but not case fatality), which will flow on to reduced IHD and stroke mortality (and hence averting YLLs) and reduced prevalence of IHD and stroke (and hence averting YLDs).
- It is also possible to alter case fatality rates for a treatment intervention, and hence avert YLLs (but probably incur some YLDs due to increasing the prevalence pool).

Working with this main epidemiological structure, direct costs of the intervention are estimated, and cost-offsets by disease state attached to the prevalent numbers in the life-table.

7.1 General comments

ACE-Prevention (Australia) is fairly conservative in some of its assumptions. For example, some ACE-Prevention (Australia) interventions conservatively assumed no benefit to younger people (e.g., the salt reduction interventions [87] do not assume any benefit (ever) for those initially aged <30 years). This is a reasonable simplifying step – but may generate quite a gap in the DALY averted benefit. One of the overall principles of BODE^3 will be to input best estimates, not conservative estimates, wherever possible and where the impact of so doing is more than negligible.

For simplicity, disability weights (and associated DALYs) are not attributed to risk factor states, e.g., having high blood pressure/hypertension, high cholesterol, etc. This is despite considerable evidence that getting such diagnoses do often contribute to morbidity (associated with "disease labelling" but also directly e.g., headaches from hypertension).

7.2 Particular assumptions implicit in ACE-Prevention (Australia) modelling

ACE-Prevention (Australia) models follow the principles and assumptions outlined in Section 1 of this Protocol. Here we just list some particular aspects of the existing or preceding ACE-Prevention (Australia) models that illustrate some of these assumptions, focusing on those that we may vary slightly (or conduct sensitivity analyses on) in NZACE-Prevention.

7.2.1 Age weights and 'fair innings'

The original DALYs used in the GBD study incorporate age weights, reflecting an assumption of maximal social value of life at around 20.[88] This assumption proved highly controversial[89] [90], and age weights were subsequently dropped in future iterations of the GBD and country BDS. They have also not been included in economic evaluations using DALYs.

However, just as ethnicity is an equity concern, so too is age.[89] [90] Williams advances a fair innings argument that society values health gains for the young more than the old.[48] [91] One option in NZACE-Prevention will be to trial using the original DALY age weights. That said, the nature of preventive interventions having time lags, the lifetable modelling that already assigns more potential health gain to younger people, and discounting, all mean that age weights may not make much relative difference to the final cost-effectiveness ranking.

7.2.2 Time lags to health benefits

It is possible to set time lags from intervention establishment to actual change in disease incidence. The ACE-Prevention (Australia) salt analyses allowed for some time lag between intervention (i.e., year 2003) and changes in stroke and IHD incidence as a sensitivity analysis. The physical activity analyses allowed for no time lags (e.g., colorectal cancer incidence changed from year 2003). The assumption of 'running at steady state' can be taken to mean that no time-lags to disease prevention apply (i.e., has been running at steady state for decades).

In NZACE-Prevention we will model both with and without time lags, the former being the cost-effectiveness for at intervention actually starting in 2006, and the latter the cost-effectiveness for an intervention in steady state in the future. Similarly we will also consider phase-in periods for some interventions e.g., mandated step-wise reductions in national-level of tobacco for sale and food-grade salt for sale.

8 NZACE: Modification to New Zealand setting

8.1 Modification of epidemiological parameters

The model structure for NZACE-Prevention will be the same as that for ACE-Prevention (Australia) – but with the addition of additional strata for ethnicity and deprivation.

Many input parameters to the baseline model will be changed, for example:

- Population socio-demographics will be those for New Zealand in 2006, and subsequent projections.
- Epidemiological parameters will be sourced from the New Zealand BDS
- Prevalent YLDs will be sourced from the New Zealand BDS.

The table below includes a list of the types of parameters that are required for the ACE modelling, and comments on where they will be sourced from in NZACE-Prevention. *The italicised comments given are particular to stroke and IHD prevention by way of example.* A major initial body of work for NZACE-Prevention will be assembling all the input variables below for all disease and injury states (aligned with ICD codes used in the NZ BDS).

Table 12: Input variables (and their source) for NZACE-Prevention modelling

Input variable	**Source**
1. Incidence of disease or injury – baseline 2006	Usually hospitalisation data analysed from incidence perspective. Likely to involve calibration with duration, prevalence and mortality data using DISMOD. Should be routine parameter available from the NZ BDS. *The incidence of IHD and stroke at each year of age (and for each sex by ethnicity by deprivation combination) from New Zealand NMDS hospitalisation data and/or ARCOS.*
2. Case fatality of disease or injury –baseline 2006	Usually hospitalisation data analysed from mortality perspective. Likely to involve calibration with incidence, duration, prevalence and mortality data using DISMOD. Should be routine parameter available from the NZ BDS. *The case fatality from IHD and stroke at each year of age (and for each sex by ethnicity by deprivation combination) from New Zealand NMDS hospitalisation data and/or ARCOS, linked to NMDS mortality data.*
(3. Prevalence – baseline 2006)	(Not a primary input into ACE models, as actually an output or consequence of the incidence and case-fatality data. However, it will be important in calibration. Should be routine parameter available from the NZ BDS).

4. Risk factor distribution – baseline 2006.	Will be sourced from survey data. Should be routine parameter available from the NZ BDS. *Risk factor distributions for BP, cholesterol, smoking, glucose, BMI, etc…, by sex, age, ethnicity and deprivation.*
5. Counterfactual risk factor distribution.	Will be a set of link-models that shift the risk factor distribution based on a plausible intervention effect. *For example, the shift in blood pressure distribution from a salt reduction intervention.*
6. Counterfactual case fatality rate.	Will be a set of link-models that reduce case fatality based on a plausible treatment effect. *For example, the shift in IHD mortality from stenting. (Note many pharmaceutical treatments will actually work through changing risk factor distributions).*
7. Cost-offsets	Health sector costs attributed to prevalent states in the multi-state life table. Will be calculated from HealthTracker data.
8. Disability weights	From the GBD study, attributed to prevalent states in the multi-state life table.

8.2 Modification to costing

As described above for BODE^3 generally, and ABC-CBA specifically, direct costing will be conducted specifically for BODE^3 simulated interventions, and cost off-sets will (hopefully) be determined from HealthTracker data with AIHW data as a possible prior or default option.

8.3 Intervention selection

A three stage process was used to select interventions to model in New Zealand. First, salt reduction strategies were selected to test out implementation processes in the New Zealand context (since the ACE-Prevention (Australia) salt model was well developed and relatively simple). Second, priority risk factors were selected, and included tobacco, cholesterol and hypertension in the first wave of analyses.[9] Third, a process of selecting exact interventions from these risk factor domains was undertaken.[10] (See Section 2.2, page 43).

8.4 Other modifications

Various minor differences will apply to some of the methods NZACE-Prevention compared to ACE-Prevention (Australia) in the domains of:

- What is included in the "health perspective" with the routine exclusion of industry costs (though these may be included in scenarios that move towards a societal perspective).

- The routine inclusion of health costs relating to extra life lived as a result of the intervention/s.
- Possibly a slightly different approach to set-up costs (see Section 1.2.9).
- A greater focus on discussing aspects of health inequalities (particularly for Māori vs non-Māori, but also for gender and age).

Further details on these issues will be dealt with in topic specific technical reports on the BODE^3 website, e.g., on the methods used in the salt reduction modelling work.

References

1. Drummond M, Sculpher M, Torrance G, O'brien B, Stoddart G. *Methods for the Economic Evaluation of Health Care Programmes (Third Edition)*. London: Oxford University Press, 2005.
2. Lash T, Fox M, Fink A. *Applying Quantitative Bias Analysis to Epidemiological Data*. New York: Springer, 2009.
3. Rothman K, Greenland S, Lash T. *Modern Epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 2008.
4. Briggs A, Sculpher M, Claxton K. *Decision modelling for health economic evaluation*. New York: Oxford University Press, 2006.
5. Gold M, Siegel J, Russell L, Weinstein M, editors. *Cost-effectiveness in Health and Medicine*. New York: Oxford University Press, 1996.
6. Tan-Torres Edejer T, Baltussen R, Dadam T, Hutubessy R, Acharya A, Evans D, Murray CJL. Making Choices in Health: WHO guide to cost-effectiveness analysis. Geneva: World Health Organization, 2003.
7. Vos T, Carter R, Et Al. Assessing Cost-Effectiveness in the Prevention of Non-Communicable Disease (ACE–Prevention) Project 2005–09. Economic Evaluation Protocol 2007.
8. Carter R, Vos T, Moodie M, Haby M, Magnus A, Mihalopoulis C. Priority setting in health: origins, description and application of the Australian Assing Cost-Effectiveness initiative. *Expert Rev Pharmacoeconomics Outcomes Res* 2008;8:593 - 617.
9. Wilson N, Blakely T, Foster R, Hadorn D, Vos T. What are the Priority Health Risk Factors for Researching Preventive Interventions as Part of NZACE-Prevention? . *Burden of Disease Epidemiology Equity and Cost Effectiveness Programme Technical Report No.1.* Wellington: Department of Public Health, University of otago, Wellington, 2010.
10. Wilson N, Nzace-Prevention Team. Possible NZACE-Prevention Interventions for Stakeholders to Critique. *Burden of Disease Epidemiology Equity and Cost Effectiveness Programme Technical Report No.2.* Wellington: Department of Public Health, University of otago, Wellington, 2010.
11. Kvizhinadze G, Blakely T. Projected NZ Life Tables. *Burden of Disease Epidemiology, Equity and Cost-Effectiveness Programme - Technical Report No.4*. Wellington: Department of Public Health, University of Otago, Wellington 2011.
12. Costilla R, Atkinson J, Blakely T. Incorporating Ethnic and Deprivation Variation to Cancer Incidence Estimates over 2006-2026 for ABC-CBA. *Burden of Disease Epidemiology, Equity and Cost-Effectiveness Programme - Technical Report No.5*. Wellington: Department of Public Health, University of Otago, Wellington 2011.
13. Foster R, Preval N, Blakely T, Wilson N, O'dea D. Costing of Pharmaceuticals in New Zealand for Health Economic Studies: Backgrounder and Protocol for Costing. *Burden of Disease Epidemiology, Equity and Cost-Effectiveness Programme - Technical Report No.6*. Wellington: Department of Public Health, University of Otago, Wellington 2011.
14. Foster R. Literature Search and Data Synthesis Methods for Estimating Inputs for Health Economic Modelling. *Burden of Disease Epidemiology, Equity and Cost-Effectiveness*

Programme - Technical Report No.7. Wellington: Department of Public Health, University of Otago, Wellington 2011.

15. Nghiem N, Wilson N, Blakely T. Price Elasticities for Health Economic Modelling of Food Pricing Interventions in Australia and New Zealand. *Burden of Disease Epidemiology, Equity and Cost-Effectiveness Programme - Technical Report No.9*. Wellington: Department of Public Health, University of Otago, Wellington 2011.
16. Kvizhinadze G, Blakely T. Modelling Options for ABC-CBA *Burden of Disease Epidemiology, Equity and Cost-Effectiveness Programme - Technical Report No.17*. Wellington: Department of Public Health, University of Otago, Wellington 2012.
17. Blakely T, Costilla R, Soeberg M. Cancer excess mortality rates over 2006-2026 for ABC-CBA. *Burden of Disease Epidemiology, Equity and Cost-Effectiveness Programme - Technical Report No.10*. Wellington: Department of Public Health, University of Otago, Wellington 2012.
18. Foster R. Determination of effect size for modelling in BODE^3: A worked example of reducing dietary saturated fat intake on cardiovascular events. *Burden of Disease Epidemiology, Equity and Cost-Effectiveness Programme - Technical Report No.8*. Wellington: Department of Public Health, University of Otago, Wellington 2012.
19. Foster R, T B, Wilson N, O'dea D. Protocol for Direct Costing of Health Sector Interventions for Economic Modelling (Including Event Pathways). . *Burden of Disease Epidemiology, Equity and Cost-Effectiveness Programme, Technical Report No. 12.* . Wellington: University of Otago, Wellington, 2012.
20. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling Good Research Practices—Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Value in Health* 2012;15(6):796-803.
21. Roberts M, Russell LB, Paltiel AD, Chambers M, Mcewan P, Krahn M. Conceptualizing a Model: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. *Value in Health* 2012;15(6):804-11.
22. Siebert U, Alagoz O, Bayoumi AM, Jahn B, Owens DK, Cohen DJ, Kuntz KM. State-Transition Modeling: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-3. *Value in Health* 2012;15(6):812-20.
23. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Möller J. Modeling using Discrete Event Simulation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. *Value in Health* 2012;15(6):821-27.
24. Pitman R, Fisman D, Zaric GS, Postma M, Kretzschmar M, Edmunds J, Brisson M. Dynamic Transmission Modeling: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-5. *Value in Health* 2012;15(6):828-34.
25. Briggs AH, Weinstein MC, Fenwick EaL, Karnon J, Sculpher MJ, Paltiel AD. Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. *Value in Health* 2012;15(6):835-42.
26. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, Mcdonald KM, Wong JB. Model Transparency and Validation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Value in Health* 2012;15(6):843-50.
27. Vos T, Carter R, Barendregt J, Mihalopoulis C, Veerman L, Magnus A, . . . For the Ace-Prevention Team. Assessing Cost-Effectiveness in the Prevention (Ace-Prevention):

Final Report. Brisbane and Melbourne: University of Queensland and Deakin University, 2010.

28. Murray C, Evans D, Acharya A, Baltussen R. Development of WHO guidelines on generalized cost-effectiveness analysis. *Health Economics* 2000;9:235-51.
29. Blakely T, Foster R, Wilson N, And Bode^3 Team. Burden of Disease Epidemiology, Equity and Cost-Effectiveness (BODE3) Study Protocol. Version 1.0. *Burden of Disease Epidemiology, Equity and Cost-Effectiveness Programme - Technical Report No. 3*. Wellington: Department of Public Health, University of Otago, Wellington, 2011.
30. Blakely T, Shaw C, Atkinson J, Tobias M, Bastiampillai N, Sloane K, . . . Cunningham R. CancerTrends: Trends in cancer incidence by ethnic and socioeconomic group, New Zealand 1981-2004. Wellington: University of Otago, and Ministry of Health (www.otago.ac.nz/wellington/research/hirp/otago020023.html), 2010
31. Ministry of Health. Cancer in New Zealand. Trends and Projections. Wellington: Ministry of Health, 2002.
32. Ministry of Health. Cancer Incidence Projections: 1999–2003 update. Wellington: Ministry of Health, 2008.
33. Karnon J, Brennan A, Akehurst R. A Critique and Impact Analysis of Decision Modeling Assumptions. *Medical Decision Making* 2007;27(4):491-99.
34. Kuntz KM, Fenwick E, Briggs A. Appropriate Cohorts for Cost-Effectiveness Analysis: To Mix or Not to Mix? *Medical Decision Making* 2010;30(4):424-25.
35. Dewilde S, Anderson R. The Cost-Effectiveness of Screening Programs Using Single and Multiple Birth Cohort Simulations: A Comparison Using a Model of Cervical Cancer. *Medical Decision Making* 2004;24(5):486-92.
36. Hoyle M, Anderson R. Whose Costs and Benefits? Why Economic Evaluations Should Simulate Both Prevalent and All Future Incident Patient Cohorts. *Medical Decision Making* 2010;30(4):426-37.
37. Kvizhinadze G, Blakely T. Projected New Zealand Lifetables. *Burden of Disease Epidemiology, Equity and Cost-Effectiveness Programme (BODE³).*

Technical Report: Number 4. Wellington: Department of Public Health, University of Otago, 2011.

38. Statistics New Zealand. A History of Survival in New Zealand: Cohort life tables 1876–2004. Wellington: Statistics New Zealand, 2007.
39. Blakely T, Carter K, Wilson N, Edwards R, Woodward A, Thomson G, Sarfati D. If nobody smoked tobacco in New Zealand from 2020 onwards, what effect would this have on ethnic inequalities in life expectancy? *New Zealand Medical Journal* 2010;123(1320):26-36.
40. Carter KN, Blakely T, Soeberg M. Trends in survival and life expectancy by ethnicity, income and smoking in New Zealand: 1980s to 2000s. *New Zealand Medical Journal* 2010;123(1320):13-24.
41. Salomon JA, Vos T, Hogan DR, Gagnon M, Naghavi M, Mokdad A, . . . Murray CJL. Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. *The Lancet* 2012;380(9859):2129-43.

42. Stouthard M, Essink-Bot M-L, Bonsel GJ, Barendregt J, Kramer P, Van De Water H, . . . Van Der Maas P. *Disability Weights for Diseases in The Netherlands*. Rotterdam, 1997.
43. Stouthard MEA, Essink-Bot M-L, Bonsel GJ, On Behalf of the Dutch Disability Weights Group. Disability weights for diseases. *The European Journal of Public Health* 2000;10(1):24-30.
44. Begg S, Vos T, Barker B, Stevenson C, Stanley L, Lopez A. The burden of disease and injury in Australia 2003. Canberra: AIHW, 2007.
45. Koerkamp BG, Stijnen T, Weinstein MC, Hunink MGM. The Combined Analysis of Uncertainty and Patient Heterogeneity in Medical Decision Models. *Medical Decision Making* 2011;31(4):650-61.
46. Greenland S. Bayesian perspectives for epidemiologic research: III. Bias analysis via missing-data methods. *International Journal of Epidemiology* 2009:dyp278.
47. Cookson R, Drummond M, Weatherly H. Explicit incorporation of equity considerations into economic evaluation of public health interventions. *Health Economics, Policy, & Law* 2009;4(Pt 2):231-45.
48. Williams AH. Intergenerational equity: an exploration of the 'fair innings' argument. *Health Economics* 1997;10:583-5.
49. Culyer AJ, Wagstaff A. Equity and equality in health and health care. *Journal of Health Economics* 1993;12(4):431-57.
50. Williams AH, Cookson RA. Equity-efficieny trade-offs in health technology assessment. *International Journal of Technology Assessment in Health Care* 2006;22(1):1-9.
51. Stolk EA, Pickee SJ, Ament AHJA, Busschbach JJV. Equity in health care prioritisation: An empirical inquiry into social value. *Health Policy* 2005;74(3):343-55.
52. Bleichrodt H, Diecidue E, Quiggin J. Equity weights in the allocation of health care: the rank-dependent QALY model. *Journal of Health Economics* 2004;23(1):157-71.
53. Bleichrodt H, Doctor J, Stolk E. A nonparametric elicitation of the equity-efficiency trade-off in cost-utility analysis. *Journal of Health Economics* 2005;24(4):655-78.
54. Paul D, Rebecca S, Aki T, Alan W. QALY maximisation and people's preferences: a methodological review of the literature. *Health Economics* 2005;14(2):197-208.
55. Stolk E, Van Donselaar G, Brouwer W, Busschbach J. Reconciliation of economic concerns and health policy: illustration of an equity adjustment procedure using proportional shortfall. *Pharmacoeconomics* 2004;22:1097-107.
56. Schwappach D. Resource allocation, social values and the QALY: a review of the debate and empirical evidence. *Health Expect* 2002;5:210-22.
57. Hauck K, Smith P, Goddar M. The Economics of Priority Setting for Health Care: A Literature Review. Washington: The World Bank, 2004.
58. Williams A. Intergenerational equity: an exploration of the 'fair innings' argument. *Health Econ* 1997;6:117 - 32.
59. Ministry of Health. He Korowai Oranga. Wellington: Ministry of Health, 2001.
60. Ministry of Health. The Best Use of Available Resources: An approach to prioritisation. Wellington: Ministry of Health, 2005.
61. Pharmac. Prescription for Pharmocoeconmic Analysis: Methods for cost utility analysis. Wellington: Pharmac, 2007.

62. National Health Committee (National Advisory Committee on Health and Disability). New Technology Assessment in New Zealand: a discussion document. Wellington: National Advisory Committee on Health and Disability., 2002.
63. Wilson N, Blakely T, Foster R, Hadorn D, Vos T. What are the Priority Risk Factors for Researching Preventive Interventions in New Zealand? *Bull World Health Organ* 2012;90:88-96.
64. World Health Organization. The Global Burden of Disease: 2004 Update. Geneva: World Health Organization, 2008.
65. World Health Organization. Global health risks: mortality and burden of disease attributable to selected major risks. Geneva: World Health Organization, 2009.
66. Cobiac L, Vos T, Doran C, Wallace A. Cost-effectiveness of interventions to prevent alcohol-related disease and injury in Australia. *Addiction* 2009;104(10):1646-55.
67. Byrnes J, Cobiac L, Doran C, Vos T, Shakeshaft A. Cost-effectiveness of volumetric alcohol taxation in Australia. *mja* 2010;192(8):439-43.
68. Shih S, Carter R, Sinclair C, Mihalopoulous C, Vos T. Economic evaluation of skin cancer prevention in Australia. *Prev Med* 2009;49:449-53.
69. Bertram M, Lim S, Barendregt J, Vos T. Assessing the cost-effectiveness of drug and lifestyle intervention following opportunistic screening for pre-diabetes in primary care. *Diabetologia* 2010;53(5):875-81.
70. Cobiac LJ, Vos T, Barendregt JJ. Cost-Effectiveness of Interventions to Promote Physical Activity: A Modelling Study. *PLoS Med* 2009;6(7):e1000110.
71. Vos T, Carter R, Doran C, Anderson I, Lopez A, Wilson A. Assessing Cost-Effectiveness in the Prevention of Non-Communicable Disease (ACE–Prevention) Project 2005–09: Economic Evaluation Protocol (As per September 2007). Brisbane: Centre for Burden of Disease and Cost Effectiveness 2007.
72. Pharmac. Prescription for Pharmacoeconomic Analysis: Methods for Cost-Utility Analysis. Version 2.1. Wellington: PHARMAC Pharmaceutical Management Agency, 2012.
73. Nhmrc. NHMRC additional levels of evidence and grades for recommendations for developers of guidelines - Stage 2 Pilot. Canberra (ACT): National Health and Medical Research Council, 2009.
74. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, . . . Glanville J. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technology Assessment* 2004;8(36).
75. Leal J, Wordsworth S, Legood R, Blair E. Eliciting Expert Opinion for Economic Models: An Applied Example. *Value in Health* 2007;10(3):195-203.
76. Aspinall W. A route to more tractable expert advice. *Nature* 2010;463(7279):294-5.
77. Hsu C, Sandford BA. The Delphi technique: making sense of consensus. *Practical Assessment, Research and Evaluation*, 2007.
78. Ministry of Health. Hospital Events 2008/09 and 2009/10. Wellington: Ministry of Health, 2012.
79. Rappange DR, Van Baal PHM, Van Exel NJA, Feenstra TL, Rutten FFH, Brouwer WBF. Unrelated Medical Costs in Life-Years Gained: Should They be Included in Economic Evaluations of Healthcare Interventions? *Pharmacoeconomics* 2008;26(10):815.

80. Van Baal PHM, Feenstra TL, Polder JJ, Hoogenveen RT, Brouwer WBF. Economic evaluation and the postponement of health care costs. *Health Economics* 2011;20(4):432-45.
81. Blakely T, Costilla R, Tobias M. The Burden of Cancer: New Zealand 2006. Wellington: Ministry of Health, 2010.
82. New Zealand Health Information Services. Cancer Patient Survival Covering the Period 1994 to 2003. Wellington: Ministry of Health, 2006.
83. Begg SJ, Vos T, Barker B, Stanley L, Lopez AD. Burden of disease and injury in Australia in the new millennium: measuring health loss from diseases, injuries and risk factors. *Medical Journal of Australia* 2008;188(1):36-40.
84. Petrou S, Gray A. Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. *BMJ* 2011;342.
85. Byrnes JM, Cobiac LJ, Doran CM, Vos T, Shakeshaft AP. Cost-effectiveness of volumetric alcohol taxation in Australia. *Med J Aust* 2010;192 (8):439-43.
86. Cobiac LJ, Vos T, Veerman L. Cost-effectiveness of Weight Watchers and the Lighten Up to a Healthy Lifestyle program. *Aust N Z J Public Health* 2010;34:240-7.
87. Cobiac LJ, Vos T, Veerman JL. Cost-effectiveness of interventions to reduce dietary salt intake. *Heart* 2010;96(23):1920-25.
88. Murray J, Lopez A. *The Global Burden of Disease: A comprehensive assessment of the mortality from disease, injuries, and risk factors in 1990 and projected to 2020*. Boston: Harvard School of Public Health, 1996.
89. Arneson T, Nord E. The value of DALY life: problems with ethics and valditiy of disability adjusted life years. *BMJ* 1999;319:1423-5.
90. Murray C, Salomon J, Mather C. A critical examination of summary measures of poulations health. *Bulletin of the World Health Organization* 2000;78:981-93.
91. Vos T, Corry J, Haby MM, Carter R, Andrews G, Vos T, . . . Andrews G. Cost-effectiveness of cognitive-behavioural therapy and drug interventions for major depression. *Australian & New Zealand Journal of Psychiatry* 2005;39(8):683-92.