

RESEARCH

Open Access



# Disaggregating proportional multistate lifetables by population heterogeneity to estimate intervention impacts on inequalities

Patrick Andersen<sup>1</sup>, Anja Mizdrak<sup>2</sup>, Nick Wilson<sup>2</sup>, Anna Davies<sup>2</sup>, Laxman Bablani<sup>1</sup> and Tony Blakely<sup>1\*</sup>

## Abstract

**Background:** Simulation models can be used to quantify the projected health impact of interventions. Quantifying heterogeneity in these impacts, for example by socioeconomic status, is important to understand impacts on health inequalities. We aim to disaggregate one type of Markov macro-simulation model, the proportional multistate lifetable, ensuring that under business-as-usual (BAU) the sum of deaths across disaggregated strata in each time step returns the same as the initial non-disaggregated model. We then demonstrate the application by deprivation quintiles for New Zealand (NZ), for: hypothetical interventions (50% lower all-cause mortality, 50% lower coronary heart disease mortality) and a dietary intervention to substitute 59% of sodium with potassium chloride in the food supply.

**Methods:** We developed a disaggregation algorithm that iteratively rescales mortality, incidence and case-fatality rates by time-step of the model to ensure correct total population counts were retained at each step. To demonstrate the algorithm on deprivation quintiles in NZ, we used the following inputs: overall (non-disaggregated) all-cause mortality & morbidity rates, coronary heart disease incidence & case fatality rates; stroke incidence & case fatality rates. We also obtained rate ratios by deprivation for these same measures. Given all-cause and cause-specific mortality rates by deprivation quintile, we derived values for the incidence, case fatality and mortality rates for each quintile, ensuring rate ratios across quintiles and the total population mortality and morbidity rates were returned when averaged across groups. The three interventions were then run on top of these scaled BAU scenarios.

**Results:** The algorithm exactly disaggregated populations by strata in BAU. The intervention scenario life years and health adjusted life years (HALYs) gained differed slightly when summed over the deprivation quintile compared to the aggregated model, due to the stratified model (appropriately) allowing for differential background mortality rates by strata. Modest differences in health gains (HALYs) resulted from rescaling of sub-population mortality and incidence rates to ensure consistency with the aggregate population.

**Conclusion:** Policy makers ideally need to know the effect of population interventions estimated both overall, and by socioeconomic and other strata. We demonstrate a method and provide code to do this routinely within proportional multistate lifetable simulation models and similar Markov models.

**Keywords:** Simulation modeling, Heterogeneity, Markov model, Proportional multistate lifetable, Inequalities, Sodium, Salt, Cardiovascular disease

\*Correspondence: [ablakely@unimelb.edu.au](mailto:ablakely@unimelb.edu.au)

<sup>1</sup> Population Interventions Unit, Centre for Epidemiology and Biostatistics, The University of Melbourne, 207-221 Bouverie St., Parkville, VIC 3010, Australia

Full list of author information is available at the end of the article

## Background

Quantifying the health impacts of population interventions by social strata is necessary for the design of effective policies to reduce health inequalities. These



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

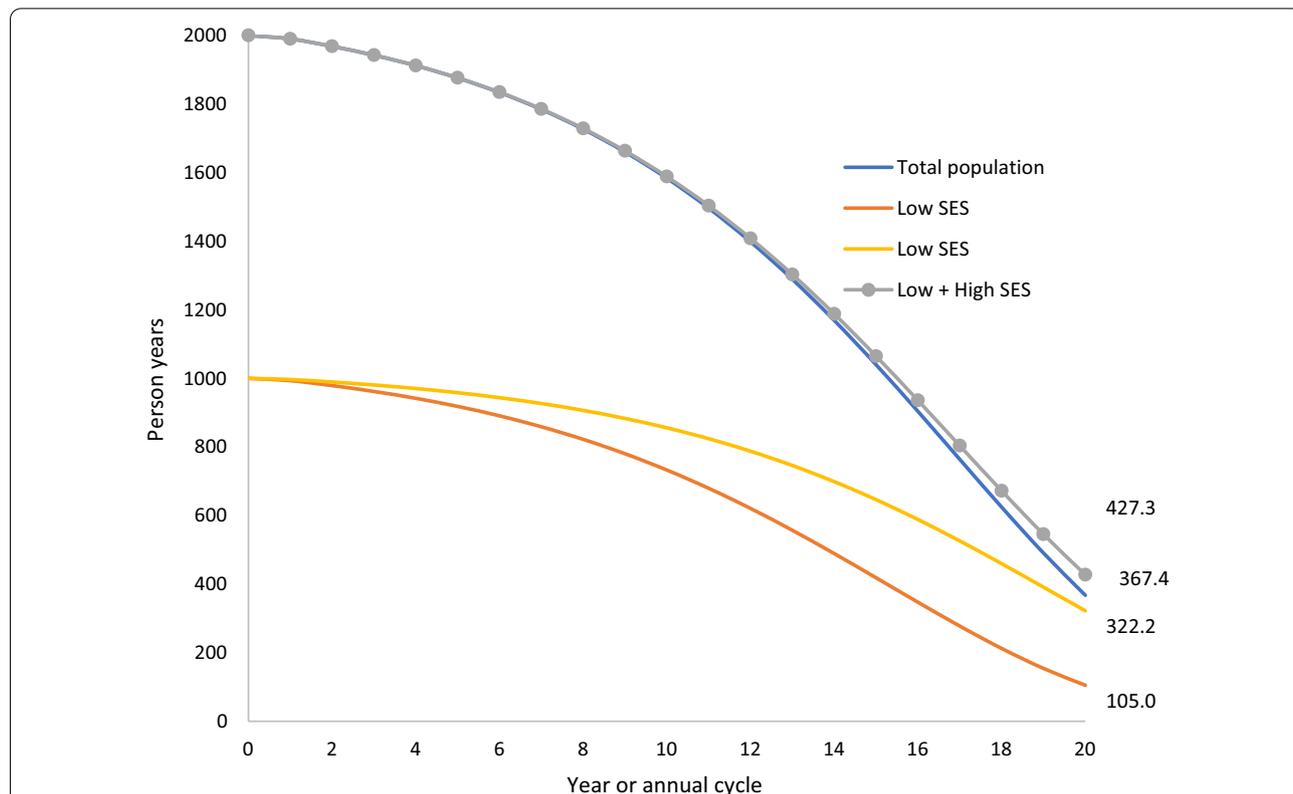
analyses are often based on simulation models, which attempt to quantify the future health (and cost) impacts of interventions. For example: what is the health gain of various tobacco [1], diet [2–5] and other preventive interventions by socioeconomic groups? And, how will the intervention change health inequalities?

Simulating interventions by socioeconomic group is difficult. The heterogeneity of epidemiological parameters (e.g. incidence, mortality, morbidity, case fatality, etc.) across socioeconomic strata and determining their effects on simulation outputs is challenging; both in specifying the relevant inputs correctly, and coherence in the modeling (e.g. ensuring that deaths and other outputs such as prevalence sum across strata to return that in the aggregated population).

For an illustrative example, consider a lifetable simulation for a population of 2000 members. Suppose that the mortality risk of the population is 1 per 100 and increases by 20% per year for 20 years to account for aging. After the first year, there will be an expected 1980 people alive, and by the 20th year, the expected

population will decrease to 367.4. The person-years lived by annual cycle are shown in Fig. 1.

Now suppose that our population is comprised of two distinct groups, e.g., high and low socioeconomic status (SES) groups, each with 1000 people. Suppose also that one group has twice the mortality risk of the other. Setting the mortality risks for the groups to be 4/3 and 2/3 deaths per 100 respectively yields the same expected number of living people after the first year as were obtained previously (i.e., 1980). By applying the annual 20% increase of mortality risk to each group and simulating to the end of 20th year, we generate the number of person years as shown in Fig. 1. In the 20th annual cycle there are 322.2 and 105.0 expected person years lived in the high and low mortality group respectively. This sums to 427.3, which is greater than the 367.4 obtained when the population was modeled as a whole. That is, the mortality parameters are mis-specified in the stratified model, because differing mortality rates by strata change the proportional distribution of people across strata away from 50:50 over time. An intervention scenario imposed on top of these mis-specified lifetables likewise



**Fig. 1** Demonstration of the principle issues addressed in our study: Persons years lived by annual cycle for a population of initial size  $n = 2000$ , with mortality rate 1 per 100 person years in the first year, and the high and low SES groups (each  $N = 1000$ ) that make up this total population. The mortality rate in the low SES group is twice that in the high SES group at baseline (i.e. 0.667 and 1.333 per 100 person years, respectively). The mortality rate increase by 20% per annum in the total population, and each SES group

incorrectly estimate the health gains for each stratum and the differences in health gain by stratum (i.e. the inequality impact).

The purpose of our study was to develop methods to disaggregate populations in Markov macro-simulation models, particularly proportional multistate lifetable models (PMSLT) [6, 7], retaining fidelity with the aggregate population in terms of numbers of deaths and other event counts. In so doing, we are implicitly assuming that aggregate epidemiological data (and forecasts for modeling interventions into the future) are more accurate than disaggregated data. For example, we assume that forecasts of future mortality and incidence rates by sex and age are more accurate (and more readily available) than such estimates additionally stratified by socioeconomic status. We are also adhering to a ‘burden of disease approach’ that places and emphasis on ensuring ‘everything adds up’ – at least in the business-as-usual (BAU) epidemiological model. We searched the literature for previous publications on this topic but could find none (see “[Appendix 1: Abstracts from literature search](#)”).

A few extensions of multistate models accounting for heterogeneity have been observed in the literature. Such extensions have highlighted, in a disease-free context with independent populations, that mortality is underestimated if heterogeneity is not accounted for [8, 9]. Further, log-linear, hazard, or Bayesian regression [10] methods can be used to calculate differences in multistate life table parameter values by an explicitly defined heterogeneous parameter of interest (say, SES or education) for further simulation. Other related works in epidemiology and demography demonstrate methods for obtaining the mortality of diseased and non-diseased cohorts from the overall mortality rate in continuous time PMSLT models [11, 12].

Our work differs from the previous literature in that we use a proportional version of multistate lifetables where rates only vary at discrete time steps. We also treat the aggregate population as the ‘working truth’, with our task being to obtain a unique set of valid rates for the strata in order to quantify the impacts on health inequalities.

First, we outline the mathematics of disaggregating populations, such that separately simulating the transitions of each sub-stratum under BAU through its state transition Markov chain yields death counts and person-years lived that are identical with the aggregate population. Second, we demonstrate intervention scenarios applied to both the aggregate and deprivation-stratified models. We deliberately chose interventions that do not have differing effect sizes by deprivation quintile. The three modeled scenarios are two hypothetical interventions (50% lower all-cause mortality, 50% lower coronary heart disease mortality) and a dietary intervention to

substitute 59% of sodium chloride with potassium chloride in the food supply in New Zealand.

We apply this disaggregation method to the disease and all-cause components of a PMSLT model. An intervention model, though, first calculates the intervention effect size as a change in risk factor distribution (now and into the future) to then generate a population impact fraction (PIF; or difference in disease incidence rates between BAU and intervention scenarios). We assume the modellers estimate the PIFs by strata of variable we are disaggregating the lifetables for, and therefore we do not present methods for disaggregation of PIFs.

## Methods

We consider two separate Markov processes which describe the mortality and disease lifetables used in PMSLT modeling [6, 7].

### The mortality lifetable

The *mortality lifetable* of a population  $P$  measures metrics such as the number of deaths, life years and life expectancy, capturing snapshots at discrete timesteps (e.g. years). Let  $A_t$  and  $D_t$  denote the number of people alive and dead at the end of the  $t$ -th timestep. We assume  $A_0 = N$  and  $D_0 = 0$ , where  $N$  is the total number of people that are observed initially, and for each  $t$  we have a mortality rate  $m_t$ . Then:

$$\begin{aligned} A_t &= A_{t-1}e^{-m_t}, \\ D_t &= D_{t-1} + A_{t-1}(1 - e^{-m_t}) \\ \Rightarrow D_t &= N - A_t \end{aligned}$$

In the *mortality lifetable disaggregation problem*,  $P$  consists of  $n$  underlying sub-populations  $P_1, \dots, P_n$ , where we assume that members of  $P$  remain in their respective sub-populations for life. The initial population for each  $P_1, \dots, P_n$  are known, and the initial mortality rates are known (either given, or solvable using mortality rate ratios between strata). However, whilst  $m_t$ , the total population mortality rate by future time step, is given, the stratum-specific mortality rates over time are not given. This situation is not uncommon, e.g. we may know the starting disaggregation of a population by socioeconomic strata, and the rate ratios for mortality or disease incidence comparing strata, but not the exact rates by strata over time. The goal is to use the aggregate populations mortality rates, the rate ratios comparing strata, and starting distribution of each sub-population ( $P_k$ ) to solve the mortality lifetables for each  $P_1, \dots, P_n$ . The sub-populations lifetables must be *consistent* with that of the aggregate population: at each time timestep  $t$ , the sum of the people in the alive (dead) compartment for each

sub-population is equal to the number of people in the alive (dead) compartment for the aggregate population. i.e.,

$$A_t = \sum_{k=1}^n A_{k,t}$$

$$D_t = \sum_{k=1}^n D_{k,t}$$

where  $A_{k,t}$  and  $D_{k,t}$  denote the number of people in the alive and dead compartments respectively for sub-population  $k$  at time  $t$ . Thus, the disaggregation problem reduces to finding mortality rates  $m_{k,t}$  for each sub-population  $k$  at timestep  $t > 0$  such that:

$$A_{t-1}e^{-m_t} = \sum_{k=1}^n A_{k,t-1}e^{-m_{k,t}}$$

If we let  $P_1$  be the reference sub-population, then the mortality rate ratios  $r$  for timestep  $t$  are given as scalars  $r_{1,t}^{\text{mort}}, \dots, r_{n,t}^{\text{mort}}$ , where  $r_{1,t}^{\text{mort}} = 1$ , such that  $m_{k,t} = r_{k,t}^{\text{mort}}m_{1,t}$  for each sub-population  $k$ . By substituting each  $r_{k,t}^{\text{mort}}m_{1,t}$  into the consistency equations, we obtain a set of equations with a unique solution. By solving these, we are then able to obtain a unique set of sub-population mortality rates for the mortality lifetable disaggregation problem. A method for solving for these rates, and the proof of the uniqueness of the solution is in “Appendix 2: Disaggregation details”.

### The mortality/morbidity lifetable

We now extend the mortality lifetable to a *mortality/morbidity lifetable* that also includes HALYs which incorporate the effects of morbidity. Let  $L_t$  denote population life years at  $t$ , where

$$L_t = \frac{A_t + A_{t-1}}{2}$$

Our HALY unit is a rescaling of the life-year using disability rates. Let  $w_t$  denote the prevalent years of life with disability (i.e. ‘YLDs’) from a burden of disease study at time  $t$ , divided by the population in that strata. Then the formula for HALYs at time  $t$ , denoted  $L_t^*$ , is:

$$L_t^* = L_t(1 - w_t)$$

For the *mortality/morbidity lifetable disaggregation problem*, an extension of the disaggregation problem in the previous

section, we are given a mortality/morbidity lifetable for a population  $P$  which includes all parameters from the mortality lifetable, along with morbidity weights  $w_t$  and HALYs  $L_t^*$ . As before,  $P$  consists of  $n$  sub-populations

$P_1, \dots, P_n$ , with their individual population counts, mortality rates, and YLDs given for the first time step. The goal is to use the aggregate lifetable to determine the mortality rates and morbidity rates ( $w_t$ ) rates for each sub-population  $P_k$ , and hence obtain the mortality/morbidity lifetables for each  $P_1, \dots, P_n$ . We have to ensure alive population counts for sub-populations and the aggregate population agree and also total HALYs for the sub-populations agree with the aggregate population HALYs. That is:

$$L_t^* = \sum_{k=1}^n L_{k,t}^*$$

where  $L_{k,t}^*$  denotes the HALYs for sub-population  $k$  at time  $t$ . To satisfy the above equations, we must solve values  $w_{k,t}$  for each  $k$  and  $t$  such that:

$$L_t(1 - w_t) = \sum_{k=1}^n L_{k,t}(1 - w_{k,t})$$

where  $L_{k,t}$  denotes life years for sub-population  $k$  at  $t$ .

We can assume that the mortality lifetable disaggregation problem has been solved as a subproblem, since it can be independently solved using the method in “The mortality lifetable” section. Then, we have alive population values, such that  $A_t = \sum_{k=1}^n A_{k,t}$ , which implies that  $L_t = \sum_{k=1}^n L_{k,t}$ . Hence, we can simplify the HALY constraints to:

$$L_t w_t = \sum_{k=1}^n L_{k,t} w_{k,t}$$

To solve the problem, we assume *morbidity (morb)* for each  $t$  (although in all likelihood ratios vary by age and sex, but are assumed constant over  $t$ ), which are  $r_{1,t}^{\text{morb}}, \dots, r_{n,t}^{\text{morb}}$ , and  $r_{1,t}^{\text{morb}} = 1$ , such that  $w_{k,t} = r_{k,t}^{\text{morb}}w_{1,t}$  for each sub-population  $k$ . After substituting each  $r_{k,t}^{\text{morb}}w_{1,t}$  into the HALY constraints and solving, we obtain:

$$w_{k,t} = r_{k,t}^{\text{morb}} \frac{L_t w_t}{\sum_{j=1}^n r_{j,t}^{\text{morb}} L_{j,t}}$$

Thus, we are able to use morbidity ratios to uniquely disaggregate the mortality/morbidity lifetable such that the HALYs in the sub-population lifetables are consistent with the aggregate population lifetable.

### The disease lifetable

A PMSLT, described in detail in [6, 7], works through changes in disease incidence or case fatality rates, where

each disease is assumed independent of other diseases. Similar to the all-cause mortality and morbidity lifetable (above), the issue here is in ensuring that each disease-specific subsidiary lifetable also returns the numbers and rates or the total population before it is disaggregated by heterogeneity (eg. SES).

We now consider an alternative type of lifetable: the *disease lifetable*. This lifetable consists of three compartments: a healthy compartment  $S$ , diseased compartment  $C$ , and dead compartment  $D$ . At each timestep, members of the population in  $S$  transition to  $C$  according to the *incidence rate*, and from  $C$  to  $D$  according to the *fatality rate*. For some diseases, members can transition from  $C$  to  $S$  as per the *remission rate*, however, for simplicity, we do not consider this possibility for now.

Let  $S_t, C_t$  and  $D_t$  denote the number of people in compartment  $S, C$  and  $D$  respectively at the end of the  $t$ -th step. We will assume initially that  $D_0 = 0$  and  $S_0 + C_0 = N$ , where  $N$  is the total number of people initially observed. Let  $i_t$  and  $f_t$  denote the incidence and fatality rates respectively at  $t$ . The equations for  $S_t, C_t$  and  $D_t$  are given by the system:

$$\begin{aligned} S_t &= S_{t-1}e^{-i_t}, \\ C_t &= C_{t-1}e^{-f_t} + S_{t-1}(1 - e^{-i_t}), \\ D_t &= N - S_t - C_t \end{aligned}$$

These equations are premised on a simplifying assumption that members of the population cannot die from the disease in the same timestep in which they contract the disease. This assumption can, in practice, be mitigated through choosing an appropriately small timestep.

For the *disease lifetable disaggregation problem*, we are given the disease lifetable for an aggregate population  $P$  complete with incidence rates  $i_t$ , fatality rates  $f_t$  and population counts  $S_t, C_t$  and  $D_t$  at each timestep  $t$ . We assume that  $P$  consists of  $n$  separate underlying sub-populations  $P_1, \dots, P_n$ , each with their own population counts, incidence rates and fatality rates. Let  $S_{k,t}, C_{k,t}$  and  $D_{k,t}$  denote the number of people in the healthy, diseased and dead compartments respectively for sub-population  $k$  at time  $t$ . We additionally assume for each sub-population we are given the initial disease prevalence, hence we can obtain  $S_{k,0}$  and  $C_{k,0}$ . The objective of the problem is to determine both the incidence rates and fatality rates of each sub-population  $P_k$  and hence obtain the disease lifetables for each  $P_1, \dots, P_n$ . The criteria for consistency for this disaggregation at each time timestep  $t$  are given by:

$$\begin{aligned} S_t &= \sum_{k=1}^n S_{k,t} \\ C_t &= \sum_{k=1}^n C_{k,t} \end{aligned}$$

i.e., we must choose incidence rates  $i_{k,t}$  and fatality rates  $f_{k,t}$  for each timestep  $t$  and sub-population  $k$  such that:

$$S_{t-1}e^{-i_t} = \sum_{k=1}^n S_{k,t-1}e^{-i_{k,t}}$$

and

$$\begin{aligned} C_{t-1}e^{-f_t} + S_{t-1}(1 - e^{-i_t}) \\ = \sum_{k=1}^n [C_{k,t-1}e^{-f_{k,t}} + S_{k,t-1}(1 - e^{-i_{k,t}})] \end{aligned}$$

We once again assume that we are given rate ratios for the sub-population rates at each timestep  $t$ , specifically *incidence rate ratios*  $r_{1,t}^i, \dots, r_{n,t}^i$  such that  $i_{k,t} = r_k^i w_{1,t}$  for each  $k$ , and *fatality rate ratios*  $r_{1,t}^f, \dots, r_{n,t}^f$  such that  $f_{k,t} = r_k^f w_{1,t}$  for each  $k$ .

We can apply the method used in the mortality problem to obtain unique incidence rates  $i_{k,t}$  that satisfy the constraints for the healthy population. After obtaining the sub-population incidence rates, the consistency constraint for the diseased population simplifies to:

$$\begin{aligned} C_{t-1}e^{-f_t} + S_{t-1} - S_t &= \sum_{k=1}^n [C_{k,t-1}e^{-f_{k,t}} + S_{k,t-1} - S_{k,t}] \\ \Rightarrow C_{t-1}e^{-f_t} + \left( S_{t-1} - \sum_{k=1}^n S_{k,t-1} \right) &- \left( S_t - \sum_{k=1}^n S_{k,t} \right) \\ = \sum_{k=1}^n C_{k,t-1}e^{-f_{k,t}} \\ \Rightarrow C_{t-1}e^{-f_t} &= \sum_{k=1}^n C_{k,t-1}e^{-f_{k,t}} \end{aligned}$$

Thus, by using two consecutive applications of the methods described in “[Appendix 1: Abstracts from literature search](#)”, first for the healthy compartment and incidence rates and then for the diseased compartment and fatality rates, we can use the rate ratios to obtain a consistent disaggregation of the disease lifetable.

The prototype code for the above methods is provided in a GitHub repository [13].

### Case studies of interventions by deprivation strata

Our case studies are applied to the NZ population, which we disaggregate by small area deprivation (a census index at the geographic unit of about 100 individuals).

#### Intervention 1: 50% reduction in ACMR

Our first intervention is a hypothetical 50% reduction of the all-cause mortality rate (ACMR) for Māori females (Māori being the Indigenous population of NZ and suffering elevated levels of deprivation, and a determinant of SES hence we routinely stratify by sex, age and ethnicity prior to then stratifying by SES). This intervention acts directly upon the mortality/morbidity lifetable by modifying the ACMR and so does not require modeling of any disease lifetables. Additional file 1: Table S1 shows the aggregate ACMR and morbidity values for Māori females at each age and Additional file 1: Table S2 gives the aggregate population counts for each 5-year age group. We apply an annual percentage change in mortality (APC) of -2.5% to the ACMR values for each year after 2011 until 2026 [14].

To uniquely disaggregate the main lifetable for Māori females we use the rate ratios for five categories of deprivation obtained from routine health data for ACMR and morbidity values (Additional file 1: Tables S3 and S4 respectively). The initial proportions for the five strata are set at 20% each.

To apply the intervention, we multiply each business-as-usual (BAU) ACMR rate by 0.5 for both the aggregate and stratified populations.

#### Intervention 2: 50% reduction in CHD incidence

In our second intervention, we reduce the incidence rate of coronary heart disease (CHD) for Māori females by 50%. This intervention requires the modeling of a CHD disease lifetable as well as the main mortality/morbidity lifetable. The incidence rate, fatality rate, morbidity ratio and initial prevalence values for the aggregate population are given in Additional file 1: Table S5. We assume an APC of -2% for both incidence and fatality rates for each year after 2011 until 2026 (giving a 4% per annum reduction in mortality).

To disaggregate the CHD lifetable, we use the rate ratios for five categories of deprivation from routine health data for CHD incidence and fatality rates given in Additional file 1: Tables S6 and S7 respectively. We also use *prevalence ratios* for CHD given in Additional file 1: Table S8 to obtain initial prevalence values for the sub-populations.

To apply the intervention, we multiply the CHD incidence rate in BAU by 0.5 for both the aggregate and

stratified populations. This change in incidence flows through into changes in CHD mortality and prevalence, and then to changes in all-cause mortality/morbidity in the main lifetable (as described in “[The mortality lifetable](#)” section).

#### Intervention 3: 59% reduction in NZ sodium consumption

In our final example, we apply a real-world dietary intervention. As described in Nghiem et al. [15], we assume that 59% of the dietary sodium in processed foods and table salt is replaced by potassium and magnesium salts. This is estimated to reduce daily sodium intake by 51.5% for the NZ population. We assume that the effect size is the same across sub-populations (consistent with similar sodium intakes across the sub-populations considered here).

This intervention reduces the incidence rates of CHD and stroke (due to a reduction of systolic blood pressure). As such, our modeling involves both CHD and stroke disease lifetables. The CHD lifetables are obtained as per section 2.5. The stroke lifetables are similarly obtained using the aggregate values from Additional file 1: Table S9 and rate ratios for incidence, fatality and prevalence from Additional file 1: Tables S10, S11 and S12 respectively. We again assume an APC = -2% for the incidence and fatality rates of stroke for each year after 2011 until 2026.

The PSMLT modeling technique we use to apply this intervention to the disease lifetables uses *potential impact fractions (PIFs)* to scale the BAU disease incidence rates. Our PIFs for each disease  $d \in \{\text{CHD, stroke}\}$  [7] are obtained using the RR shift method described in Barendregt and Veerman [16]:

$$\text{PIF}_d = \frac{\sum_{c \in \mathcal{C}} p_c \text{RR}_c - \sum_{c \in \mathcal{C}} p_c \text{RR}_c^*}{\sum_{c \in \mathcal{C}} p_c \text{RR}_c}$$

where  $\mathcal{C}$  is the set of risk exposure categories  $c$ ,  $p_c$  is the population fraction in category  $c$ , and  $\text{RR}_c$  and  $\text{RR}_c^*$  are the relative risks of  $c$  before and after the intervention respectively. The proportions of the female Māori population in sodium risk categories are obtained from our analysis of national nutrition survey data [17] and given in Additional file 1: Table S13. To determine  $\text{RR}_c$  and  $\text{RR}_c^*$ , we apply the relative risks per 1 g daily sodium increase, from Blakely et al. [5] (to the mean sodium intakes for each category) where the mean intakes for  $\text{RR}_c^*$  are obtained from a 51.5% reduction of the BAU mean intakes. Supporting tables are presented in Additional file 1: Tables S14 and S15.

**Table 1** Business as usual 60–64 year-old cohort (centered on age 62) of Māori females; comparing aggregated and deprivation heterogeneity disaggregated proportional multistate lifetables (PMSLT) models in term of all-cause mortality rates and health adjusted life years (HALYs) outputst

Cycle (age)	HALYs											
	All-cause mortality rate						PMSLT disaggregated by deprivation heterogeneity					
	Aggregate PMSLT	PMSLT heterogeneity	Inequality measures between most and least deprived			Total population PMSLT	Least deprived	Most deprived	Sum across five quintiles	Ratio of HALYs	Difference in HALYs	
0 (62)	0.0144	0.0111	0.0176	0.0144	1.5812	0.0065	7739	1578	1518	7739	0.9621	-59.8
1 (63)	0.0156	0.0121	0.0191	0.0156	1.5812	0.0070	7624	1559	1490	7624	0.9557	-69.2
2 (64)	0.0168	0.0130	0.0206	0.0168	1.5812	0.0076	7502	1540	1461	7502	0.9487	-79.0
3 (65)	0.0181	0.0147	0.0215	0.0181	1.4616	0.0068	7056	1448	1375	7056	0.9494	-73.3
...												
20 (82)	0.0551	0.0522	0.0582	0.0551	1.1159	0.0060	3557	774	652	3557	0.8424	-121.9
...												
Sum to 82	Na	Na	Na	Na	Na	Na	121,092	25,380	23,087	121,092	0.9097	-2293

## Results

Table 1 shows the PMSLT outputs in BAU for 60–64 year old Māori females (centered on age 62) alive in 2011, for the next 20 years until aged 82. Regarding ACMRs, there is no difference for the total population modeled in aggregate compared to the weighted average across deprivation heterogeneity strata—as there should be given the method described above. Similarly, the total HALYs by cycle and summed to age 80, are identical between the aggregate and disaggregated PMSLT. Also shown in Table 1 are the ACMR for the least and most deprived quintile (with a rate ratio of 1.5812 at age 62 decreasing to 1.1159 by age 82). Whilst the population distribution is 20% in each quintile of deprivation, the HALYs lived by the least deprived are greater than for the most deprived at all ages, and more so with increasing age such that summed from age 62 to 82 the least deprived have 121.9 (18.7%) more accrued HALYs than the most deprived—due to both higher morbidity and higher ACMR.

Table 2 shows the three intervention scenarios. For the (extreme) scenario of 50% reductions in ACMRs at all ages, the rate ratios comparing the most and least deprived Māori females are unchanged (as per specification), and the rate differences are halved. By the age of 82, there is a difference in the aggregated population mortality rate (0.0275) and that averaged (weighted) across quintiles (0.0276) at the third meaningful digit. Similarly, summed to age 82 the total HALYs incremental to BAU differ by 30 (0.2%) for the aggregate (13,233) compared to heterogeneity (13,204) models—due to the non-linear association of mortality rates with mortality risks, with mortality rates varying by strata.

Also shown in Table 2 are the measures of interest to assessing health inequality impacts of the interventions. For an intervention reducing CHD incidence by 50% and a ‘real world’ scenario of sodium substitution reducing both CHR and stroke incidence, there are reductions to the null in the ACMR rate ratio and rate differences—due to the higher rates of these diseases in deprived populations that make these interventions inequality reducing. Similarly, there are greater HALY gains for the deprived population in both relative and absolute terms (last two columns of Table 2; specifically, there is both a greater absolute incremental gain in HALYs for the most deprived (last column) and a greater relative gain (ratios in second to last column greater than 1)).

## Discussion

We developed algorithms for mathematically disaggregating a population into strata such that the lifetables for the sub-populations under BAU are consistent with the lifetable of the aggregate population. To the best of

our knowledge, this method is the first of its kind to perform this disaggregation with mathematical guarantees of consistency for the sub-populations (see discussion in “Background” section). These guarantees allow the lifetables for the sub-populations to be treated as the “working truth” and thus they can be used in simulations to estimate HALY gains by strata without any loss of fidelity.

We find a modest difference in the sum of HALYs gained across strata compared to ‘simply’ applying the intervention to the aggregate population—a difference that arises due to the non-linear association of rates with risks (i.e.  $\text{risk} = 1 - \exp[\text{rate}]$ ), and where the intervention effect modeled separately across strata then summed is more accurate than modeled simply in aggregate (as heterogeneity is allowed for).

There are two main assumptions that are required for the disaggregated outputs to be reliable. First, we assume that the lifetable for the aggregate population has been correctly parameterized and that the outputs of the BAU simulations for the aggregate lifetables are accurate for the population they model. This may be a strong assumption in practice; lifetable parameters are often obtained through projections and approximations and may involve a significant degree of uncertainty. However, since the sub-populations are mathematically consistent with the aggregate population, the amount of error in the estimated sub-population is only as large as that of the aggregate.

The second assumption is that the rate ratios comparing strata (e.g. disease incidence rate ratios), and the initial proportions in each sub-population accurately reflect reality. This implies that one must have accurate a priori knowledge of the relative differences between sub-populations.

The disaggregation algorithm is not computationally expensive to implement for typical applications (i.e., a manageable number of sub-populations), and should assist researchers aiming to quantify intervention effects by population heterogeneity.

Future research may try to integrate the system of differential equations as outlined by Barendgreth et al. [18] for aggregate data, allowing for cohort members to both enter and leave states in the same cycle. However, this will make the system considerably more complex (we cannot guarantee there is a unique and obtainable solution, requiring optimization methods), which may make it more difficult to implement algorithms efficiently compared to the simpler method we provide in this paper.

## Conclusion

We provide a method for disaggregating population by heterogeneity in a PMSLT or markov model, whereby over future time steps total deaths and HALYs across all

**Table 2** Intervention impacts for 60–64 year-old cohort of Māori females, comparing aggregated and deprivation heterogeneity disaggregated proportional multistate lifetables (PMSLT) models in terms of all-cause mortality rates and health adjusted life years (HALYs) outputs

Cycle (age)	HALYs incremental (incr) to BAU															
	All-cause mortality rate					Inequality measures between most and least deprived (% shift to null c.f. BAU [Table 1])			PMSLT disaggregated by deprivation heterogeneity			PMSLT				
	Total	Least deprived	Most deprived	Average across five quintiles	Rate ratio	Rate difference	Total	Least deprived	Most deprived	Sum across quintiles	Inequality most: least deprived	Aggregate PMSLT	PMSLT disaggregated by deprivation heterogeneity	Sum across quintiles	Ratio of incr HALYs gained	Difference in HALYs gained
<i>Intervention A: 50% reduction in ACMR rate</i>																
0 (62)	0.0072	0.0056	0.0088	0.0072	1.581 (0%)	0.0032 (50%)	28	4	7	28	1.52	2.3	28	1.52	2.3	
1 (63)	0.0078	0.0060	0.0095	0.0078	1.581 (0%)	0.0035 (50%)	85	13	20	85	1.51	6.9	85	1.51	6.9	
2 (64)	0.0084	0.0065	0.0103	0.0084	1.581 (0%)	0.0038 (50%)	145	23	35	145	1.51	11.7	145	1.51	11.7	
3 (65)	0.0091	0.0074	0.0108	0.0091	1.462 (0%)	0.0034 (50%)	200	32	48	199	1.49	15.7	199	1.49	15.7	
...																
20 (82)	0.0275	0.0261	0.0291	0.0276	1.116 (0%)	0.0030 (50%)	1178	220	249	1175	1.13	29.0	1175	1.13	29.0	
...																
Sum to 82							13,233	2325	2939	13,204	1.26	614	13,204	1.26	614	
<i>Intervention B: 50% reduction in CHD incidence rate</i>																
0 (62)	0.0144	0.0111	0.0176	0.0144	1.581 (0.0%)	0.0065 (0.0%)	3.0	0.5	0.7	3.0	1.62	0.28	3.0	1.62	0.28	
1 (63)	0.0155	0.0120	0.0190	0.0155	1.579 (0.4%)	0.0070 (0.9%)	9.3	1.4	2.3	9.3	1.62	0.87	9.3	1.62	0.87	
2 (64)	0.0166	0.0129	0.0204	0.0166	1.577 (0.7%)	0.0075 (1.7%)	16.5	2.5	4.0	16.4	1.61	1.52	16.4	1.61	1.52	
3 (65)	0.0178	0.0145	0.0211	0.0177	1.455 (1.5%)	0.0066 (2.9%)	21.7	3.3	5.2	21.7	1.59	1.92	21.7	1.59	1.92	
...																
20 (82)	0.0504	0.0482	0.0529	0.0504	1.099 (14.9%)	0.0048 (21.4%)	169	30	36	168	1.20	6.11	168	1.20	6.11	
...																
Sum to 82							1697	283	378	1688	1.34	94.8	1688	1.34	94.8	
<i>Intervention C: 59% sodium chloride substitution with potassium chloride</i>																
0 (62)	0.0144	0.0111	0.0176	0.0144	1.581 (0.0%)	0.0065 (0.0%)	0.51	0.08	0.12	0.50	1.63	0.05	0.50	1.63	0.05	
1 (63)	0.0156	0.0121	0.0191	0.0156	1.581 (0.1%)	0.0070 (0.2%)	1.58	0.24	0.39	1.57	1.63	0.15	1.57	1.63	0.15	
2 (64)	0.0168	0.0130	0.0206	0.0168	1.580 (0.1%)	0.0076 (0.3%)	2.77	0.41	0.67	2.77	1.62	0.26	2.77	1.62	0.26	
3 (65)	0.0180	0.0147	0.0215	0.0180	1.460 (0.3%)	0.0068 (0.5%)	3.85	0.58	0.92	3.84	1.60	0.35	3.84	1.60	0.35	
...																
20 (82)	0.0545	0.0517	0.0576	0.0545	1.114 (1.5%)	0.0059 (2.4%)	21.9	4.0	4.6	21.7	1.17	0.67	21.7	1.17	0.67	
...																
Sum to 82							229.9	38.5	51.0	228.7	1.32	12.51	228.7	1.32	12.51	

heterogeneity strata sum to those in the parent whole population model. This method is intended for use in modelling of population interventions by (say) sex and age, but also by strata such as SES and disease risk, for policy making intended to reduce inequalities in health or focus on targeted populations to maximise cost effectiveness.

### Appendix 1: Abstracts from literature search

We conducted a literature review to determine if methods to solve the problem of consistently solving for sub-populations with simulation models had previously been proposed. We used the PubMed search engine with the Boolean string “(population) AND (disaggregation OR heterogeneity) AND (life table OR Markov process OR Markov chain) AND (rate OR transition probability) AND (technique OR algorithm)”, which was automatically expanded on by the engine to include related and MeSH terms. This returned 82 results, to which we added 5 articles that were found through previous searches. After screening titles and abstracts for relevancy, the list of articles was reduced to 14 (see below), of which we determined that none involved the same problem as investigated in this paper.

1. Title: The burden of chronic respiratory diseases and their heterogeneity across the states of India: the Global Burden of Disease Study 1990–2016.

Abstract: **BACKGROUND:** India has 18% of the global population and an increasing burden of chronic respiratory diseases. However, a systematic understanding of the distribution of chronic respiratory diseases and their trends over time is not readily available for all of the states of India. Our aim was to report the trends in the burden of chronic respiratory diseases and the heterogeneity in their distribution in all states of India between 1990 and 2016. **METHODS:** Using all accessible data from multiple sources, we estimated the prevalence of major chronic respiratory diseases and the deaths and disability-adjusted life-years (DALYs) caused by them for every state of India from 1990 to 2016 as part of the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) 2016. We assessed heterogeneity in the burden of chronic obstructive pulmonary disease (COPD) and asthma across the states of India. The states were categorised into four groups based on their epidemiological transition level (ETL). ETL was defined as the ratio of DALYs from communicable diseases to those from non-communicable diseases and inju-

ries combined, with a low ratio denoting high ETL and vice versa. We also assessed the contribution of risk factors to DALYs due to COPD. We compared the burden of chronic respiratory diseases in India against the global average in GBD 2016. We calculated 95% uncertainty intervals (UIs) for the point estimates. **FINDINGS:** The contribution of chronic respiratory diseases to the total DALYs in India increased from 4.5% (95% UI 4.0–4.9) in 1990 to 6.4% (5.8–7.0) in 2016. Of the total global DALYs due to chronic respiratory diseases in 2016, 32.0% occurred in India. COPD and asthma were responsible for 75.6% and 20.0% of the chronic respiratory disease DALYs, respectively, in India in 2016. The number of cases of COPD in India increased from 28.1 million (27.0–29.2) in 1990 to 55.3 million (53.1–57.6) in 2016, an increase in prevalence from 3.3% (3.1–3.4) to 4.2% (4.0–4.4). The age-standardised COPD prevalence and DALY rates in 2016 were highest in the less developed low ETL state group. There were 37.9 million (35.7–40.2) cases of asthma in India in 2016, with similar prevalence in the four ETL state groups, but the highest DALY rate was in the low ETL state group. The highest DALY rates for both COPD and asthma in 2016 were in the low ETL states of Rajasthan and Uttar Pradesh. The DALYs per case of COPD and asthma were 1.7 and 2.4 times higher in India than the global average in 2016, respectively; most states had higher rates compared with other locations worldwide at similar levels of Socio-demographic Index. Of the DALYs due to COPD in India in 2016, 53.7% (43.1–65.0) were attributable to air pollution, 25.4% (19.5–31.7) to tobacco use, and 16.5% (14.1–19.2) to occupational risks, making these the leading risk factors for COPD. **INTERPRETATION:** India has a disproportionately high burden of chronic respiratory diseases. The increasing contribution of these diseases to the overall disease burden across India and the high rate of health loss from them, especially in the less developed low ETL states, highlights the need for focused policy interventions to address this significant cause of disease burden in India. **FUNDING:** Bill & Melinda Gates Foundation; and Indian Council of Medical Research, Department of Health Research, Ministry of Health and Family Welfare, Government of India

Reference: [19]

2. Title: Interacting effects of unobserved heterogeneity and individual stochasticity in the life history of the southern fulmar.

**Abstract:** Individuals are heterogeneous in many ways. Some of these differences are incorporated as individual states (e.g. age, size, breeding status) in population models. However, substantial amounts of heterogeneity may remain unaccounted for, due to unmeasurable genetic, maternal or environmental factors. Such unobserved heterogeneity (UH) affects the behaviour of heterogeneous cohorts via intra-cohort selection and contributes to inter-individual variance in demographic outcomes such as longevity and lifetime reproduction. Variance is also produced by individual stochasticity, due to random events in the life cycle of wild organisms, yet no study thus far has attempted to decompose the variance in demographic outcomes into contributions from UH and individual stochasticity for an animal population in the wild. We developed a stage-classified matrix population model for the southern fulmar breeding on Ile des Pétrels, Antarctica. We applied multievent, multistate mark-recapture methods to estimate a finite mixture model accounting for UH in all vital rates and Markov chain methods to calculate demographic outcomes. Finally, we partitioned the variance in demographic outcomes into contributions from UH and individual stochasticity. We identify three UH groups, differing substantially in longevity, lifetime reproductive output, age at first reproduction and in the proportion of the life spent in each reproductive state. -14% of individuals at fledging have a delayed but high probability of recruitment and extended reproductive life span. -67% of individuals are less likely to reach adulthood, recruit late and skip breeding often but have the highest adult survival rate. -19% of individuals recruit early and attempt to breed often. They are likely to raise their offspring successfully, but experience a relatively short life span. Unobserved heterogeneity only explains a small fraction of the variances in longevity (5.9%), age at first reproduction (3.7%) and lifetime reproduction (22%). UH can affect the entire life cycle, including survival, development and reproductive rates, with consequences over the lifetime of individuals and impacts on cohort dynamics. The respective role of UH vs. individual stochasticity varies greatly among demographic outcomes. We discuss the implication of our finding for the gradient of life-history strategies observed among species and argue that individual differences should be accounted for in demographic studies of wild populations.

Reference: [20]

3. Title: The impact of individual-level heterogeneity on estimated infectious disease burden: a simulation study.

**Abstract:** **BACKGROUND:** Disease burden is not evenly distributed within a population; this uneven distribution can be due to individual heterogeneity in progression rates between disease stages. Composite measures of disease burden that are based on disease progression models, such as the disability-adjusted life year (DALY), are widely used to quantify the current and future burden of infectious diseases. Our goal was to investigate to what extent ignoring the presence of heterogeneity could bias DALY computation.

**METHODS:** Simulations using individual-based models for hypothetical infectious diseases with short and long natural histories were run assuming either "population-averaged" progression probabilities between disease stages, or progression probabilities that were influenced by an a priori defined individual-level frailty (i.e., heterogeneity in disease risk) distribution, and DALYs were calculated.

**RESULTS:** Under the assumption of heterogeneity in transition rates and increasing frailty with age, the short natural history disease model predicted 14% fewer DALYs compared with the homogenous population assumption. Simulations of a long natural history disease indicated that assuming homogeneity in transition rates when heterogeneity was present could overestimate total DALYs, in the present case by 4% (95% quantile interval: 1-8%).

**CONCLUSIONS:** The consequences of ignoring population heterogeneity should be considered when defining transition parameters for natural history models and when interpreting the resulting disease burden estimates.

Reference: [21]

4. Title: Estimation of heterogeneity in malaria transmission by stochastic modelling of apparent deviations from mass action kinetics.

**Abstract: BACKGROUND:** Quantifying heterogeneity in malaria transmission is a prerequisite for accurate predictive mathematical models, but the variance in field measurements of exposure overestimates true micro-heterogeneity because it is inflated to an uncertain extent by sampling variation. Descriptions of field data also suggest that the rate of *Plasmodium falciparum* infection is not proportional to the intensity of challenge by infectious vectors. This appears to violate the principle of mass action that is implied by malaria biology. Micro-heterogeneity may be the reason for this anomaly. It is proposed that the level of micro-heterogeneity can be estimated from statistical models that estimate the amount of variation in transmission most compatible with a mass-action model for the relationship of infection to exposure.

**METHODS:** The relationship between the entomological inoculation rate (EIR) for *falciparum* malaria and infection risk was reanalysed using published data for cohorts of children in Saradidi (western Kenya). Infection risk was treated as binomially distributed, and measurement-error (Poisson and negative binomial) models were considered for the EIR. Models were fitted using Bayesian Markov chain Monte Carlo algorithms and model fit compared for models that assume either mass-action kinetics, facilitation, competition or saturation of the infection process with increasing EIR.

**RESULTS:** The proportion of inocula that resulted in infection in Saradidi was inversely related to the measured intensity of challenge. Models of facilitation showed, therefore, a poor fit to the data. When sampling error in the EIR was neglected, either competition or saturation needed to be incorporated in the model in order to give a good fit. Negative binomial models for the error in exposure could achieve a comparable fit while incorporating the more parsimonious and biologically plausible mass action assumption. Models that assume negative binomial micro-heterogeneity predict lower incidence of infection at a given average exposure than do those assuming exposure to be uniform. The negative binomial model moreover provides an estimate of the variance of the within-cohort distribution of the EIR and hence of within cohort heterogeneity in exposure.

**CONCLUSION:** Apparent deviations from mass action kinetics in parasite transmission can arise

from spatial and temporal heterogeneity in the inoculation rate, and from imprecision in its measurement. For parasites like *P. falciparum*, where there is no plausible biological rationale for deviations from mass action, this provides a strategy for estimating true levels of heterogeneity, since if mass-action is assumed, the within-population variance in exposure becomes identifiable in cohort studies relating infection to transmission intensity. Statistical analyses relating infection to exposure thus provide a valid general approach for estimating heterogeneity in transmission but only when they incorporate mass action kinetics and shrinkage estimates of exposure. Such analyses make it possible to include realistic levels of heterogeneity in dynamic models that predict the impact of control measures on transmission intensity.

Reference: [22]

5. Title: A class of latent Markov models for capture-recapture data allowing for time, heterogeneity, and behavior effects.

**Abstract:** We propose an extension of the latent class model for the analysis of capture-recapture data which allows us to take into account the effect of a capture on the behavior of a subject with respect to future captures. The approach is based on the assumption that the variable indexing the latent class of a subject follows a Markov chain with transition probabilities depending on the previous capture history. Several constraints are allowed on these transition probabilities and on the parameters of the conditional distribution of the capture configuration given the latent process. We also allow for the presence of discrete explanatory variables, which may affect the parameters of the latent process. To estimate the resulting models, we rely on the conditional maximum likelihood approach and for this aim we outline an EM algorithm. We also give some simple rules for point and interval estimation of the population size. The approach is illustrated by applying it to two data sets concerning small mammal populations.

Reference: [23]

6. Title: Analysis of the relationship between socioeconomic factors and stomach cancer incidence in Slovenia.

**Abstract:** An unequal population distribution of well-known major risk factors explains much of the variation in the incidence of stomach cancer worldwide. The aim of this study was to determine whether geographical variation of the stomach cancer incidence rate between Slovenia's municipalities during years 1995-2001 could partially be explained by variations in the socioeconomic status as an indirect stomach cancer risk factor. A composite measure of each region's socioeconomic status, labelled as deprivation index, was created from basic socioeconomic characteristics of each municipality using factor analysis. Municipalities' standardized incidence ratios for all stomach cancers and non-cardia stomach cancer were calculated. A fully Bayesian spatial model with a conditionally autoregressive prior was applied using Markov chain Monte Carlo techniques and WinBUGS software. Spatially smoothed maps of stomach cancer incidence rates by 192 Slovenian municipalities show a clear west-to-east gradient. This pattern resembles the geographical variation of socioeconomic indices, but these indices are not significant predictors of stomach cancer incidence. Geographical variation of stomach cancer incidence in Slovenia could be partially explained by the heterogeneous socioeconomic characteristics of its municipalities. It is possible that the socioeconomic status indices used in our study were not enough powerful predictors of stomach cancer risk. Some further methodological research is needed to explain why this association was not statistically evident with the current modeling approach.

Reference: [24]

7. Title: Calculation of disease dynamics in a population of households.

**Abstract:** Early mathematical representations of infectious disease dynamics assumed a single, large, homogeneously mixing population. Over the past decade there has been growing interest in models consisting of multiple smaller subpopulations (households, workplaces, schools, communities), with the natural assumption of strong homogeneous mixing within each subpopulation, and weaker transmission between subpopulations. Here we consider a model of SIRS (susceptible-infectious-recovered-susceptible) infection dynamics in a very large (assumed infinite) population of households, with the simplifying assumption that each household is of the same size

(although all methods may be extended to a population with a heterogeneous distribution of household sizes). For this households model we present efficient methods for studying several quantities of epidemiological interest: (1) the threshold for invasion; (2) the early growth rate; (3) the household offspring distribution; (4) the endemic prevalence of infection; and (5) the transient dynamics of the process. We utilize these methods to explore a wide region of parameter space appropriate for human infectious diseases. We then extend these results to consider the effects of more realistic gamma-distributed infectious periods. We discuss how all these results differ from standard homogeneous-mixing models and assess the implications for the invasion, transmission and persistence of infection. The computational efficiency of the methodology presented here will hopefully aid in the parameterisation of structured models and in the evaluation of appropriate responses for future disease outbreaks.

Reference: [25]

8. Title: Estimating heterogeneous transmission with multiple infectives using MCMC methods.

**Abstract:** We developed a general procedure for estimating the transmission probability adjusting for covariates when susceptibles are exposed to several infectives concurrently and taking correlation within transmission units into account. The procedure is motivated by a study estimating efficacy of pertussis vaccination based on the secondary attack rate in a rural sub-Saharan community (Niakhar, Senegal) and illustrated with simulations. The procedure is also appropriate to estimate the pairwise transmission probability in transmission studies of live vaccine virus in a collection of transmission units, such as day-care centres or retirement centres. Previously, analyses either excluded transmission units with multiple infectives or ignored co-infectives. Excluding transmission units with multiple infectives is statistically less efficient and ignoring co-infectives can lead to biased estimation. Modelling is carried out by regressing the latent pairwise transmission probability from each infective to a susceptible on covariates and specifying a transmission linkage function linking the latent pairwise transmission probability to the overall transmission probability. Parameters are estimated using Markov chain Monte Carlo methods.

Reference: [26]

9. Title: Mortality of a heterogeneous cohort; description and implications.

Abstract: A recent model for heterogeneous mortality by Vaupel et al. is shown to be based on incorrect definitions. An alternative formulation is presented. The results indicate that current methods for computing the survivorship and life expectation functions underestimate the true values. A method is given for determining the possible magnitude of this underestimation. The method is illustrated by a numerical example using U.S. data.

Reference: [27]

10. Title: A New Approach to Estimating Life Tables with Covariates and Constructing Interval Estimates of Life Table Quantities.

Abstract: Extant approaches to constructing life tables generally rely on the use of population data, and differences between groups defined by discrete characteristics are examined by disaggregating the data before estimation. When sample data are used, few researchers have attempted to include covariates directly in the process of estimation, and fewer still have attempted to construct interval estimates for state expectancies when covariates are used. In this paper, we present a Bayesian approach that is useful for producing interval estimates for single-decrement, multiple-decrement, and multistate life tables. The method involves (1) estimating a hazard or survival model using Bayesian Markov chain Monte Carlo (MCMC) methods to produce a sample from the posterior distribution for the parameters of the model; (2) generating distributions of transition probabilities for selected values of covariates using the sample of model parameters; (3) using these distributions of transition probabilities as inputs for life table construction; and (4) summarizing the distribution of life table quantities. We illustrate the method on data simulated from the Berkeley Mortality Database, data from the National Health and Nutrition Examination Survey (and follow-ups), and data from the National Long Term Care Survey, and we show how the results can be used for hypothesis testing.

Reference: [28]

11. Title: Disaggregation of Statistical Livestock Data Using the Entropy Approach.

Abstract: A process of agricultural data disaggregation is developed to address the lack of updated disaggregated data concerning main livestock categories at subregional and county level in the Alentejo Region, southern Portugal. The model developed considers that the number of livestock units is a function of the agricultural and forest occupation, and data concerning the existing agricultural and forest occupation, as well as the conversion of livestock numbers into normal heads, are needed in order to find this relation. The weight of each livestock class is estimated using a dynamic process based on a generalized maximum entropy model and on a crossentropy minimization model, which comprises two stages. The model was applied to the county of Castelo de Vide and their results were validated in cross reference to real data from different sources.

Reference: [29]

12. Title: Life table methods for heterogeneous populations: distributions describing the heterogeneity.

Abstract: Taking account of heterogeneity between the individuals in population based mortality studies is important. A systematic way of describing heterogeneity is by an unobserved quantity called frailty, entering the hazard multiplicatively. Until now most studies have used a gamma distributed frailty, which is mathematically convenient; for example, the distribution among survivors is also gamma. This paper shows that several other distributions have equally simple properties, the main example being the inverse Gaussian distribution. Consequences of the different distributions are examined; the inverse Gaussian makes the population homogeneous with time, whereas for the gamma the relative heterogeneity is constant.

Reference: [30]

13. Title: Obtaining Multistate Life Table Distributions for Highly Refined Subpopulations From Cross-Sectional Data: A Bayesian Extension of Sullivan's Method.

Abstract: Multistate life table methods are often used to estimate the proportion of remaining life that individuals can expect to spend in various states, such as healthy and unhealthy states. Sullivan's

method is commonly used when panels containing data on transitions are unavailable and true multi-state tables cannot be generated. Sullivan’s method requires only cross-sectional mortality data and cross-sectional data indicating prevalence in states of interest. Such data often come from sample surveys, which are widely available. Although the data requirements for Sullivan’s method are minimal, the method is limited in its ability to produce estimates for subpopulations because of limited disaggregation of data in cross-sectional mortality files and small cell sizes in aggregated survey data. In this article, we develop, test, and demonstrate a method that adapts Sullivan’s approach to allow the inclusion of covariates in producing interval estimates of state expectancies for any desired subpopulation that can be specified in the cross-sectional prevalence data. The method involves a three-step process: (1) using Gibbs sampling to sample parameters from a bivariate regression model; (2) using ecological inference for producing transition probability matrices from the Gibbs samples; (3) using standard multistate calculations to convert the transition probability matrices into multistate life tables.

Reference: [10]

14. Title: On the heterogeneity of human populations as reflected by mortality dynamics.

**Abstract:** The heterogeneity of populations is used to explain the variability of mortality rates across the lifespan and their deviations from an exponential growth at young and very old ages. A mathematical model that combines the heterogeneity with the assumption that the mortality of each constituent subpopulation increases exponentially with age, has been shown to successfully reproduce the entire mortality pattern across the lifespan and its evolution over time. In this work we aim to show that the heterogeneity is not only a convenient consideration for fitting mortality data but is indeed the actual structure of the population as reflected by the mortality dynamics over age and time. In particular, we show that the model of heterogeneous population fits mortality data better than other commonly used mortality models. This was demonstrated using cohort data taken for the entire lifespan as well as for only old ages. Also, we show that the model can reproduce seemingly contradicting observations in late-life mortality dynamics. Finally, we show that the homogenisation of a

population, observed by fitting the model to actual data of consecutive periods, can be associated with the evolution of allele frequencies if the heterogeneity is assumed to reflect the genetic variations within the population.

Reference: [31]

### Appendix 2: Disaggregation details

Given mortality rate ratios  $r_{1,t}^{\text{mort}}, \dots, r_{n,t}^{\text{mort}}$  for timestep  $t$  with  $m_{k,t} = r_{k,t}^{\text{mort}} m_{1,t}$  for each sub-population  $k$ , our goal is to find sub-population mortality rates  $m_{k,t}$  such that:

$$A_{t-1}e^{-m_t} = \sum_{k=1}^n A_{k,t-1}e^{-m_{k,t}}$$

for each timestep  $t$ . By substituting  $m_{k,t} = r_{k,t}^{\text{mort}} m_{1,t}$  into the constraint, we obtain the equation:

$$A_{t-1}e^{-m_t} = \sum_{k=1}^n A_{k,t-1}e^{-r_{k,t}^{\text{mort}} m_{1,t}}$$

Consider the constraint for the first timestep, where  $t = 1$ .

$$A_0e^{-m_1} = \sum_{k=1}^n A_{k,0}e^{-r_{k,1}^{\text{mort}} m_{1,1}}$$

Let  $x := e^{-m_{1,1}}$ . If we assume that  $m_{1,1} \geq 0$ , then  $0 < x \leq 1$ . By substituting  $x$  into the equation, we obtain:

$$\sum_{k=1}^n A_{k,0}x^{r_{k,1}^{\text{mort}}} - A_0e^{-m_1} = 0$$

Let  $f := \sum_{k=1}^n A_{k,0}x^{r_{k,1}^{\text{mort}}} - A_0e^{-m_1} = 0$ . Assuming that  $A_{k,0} \geq 0$  for each  $k$ , and  $A_0 = \sum_{k=1}^n A_{k,0}$ , it can be seen that  $f(0) < 0$  and  $f(1) \geq 0$ . Since  $f$  is monotonically increasing in the interval  $(0, 1]$ , this implies that  $f$  has a unique root for  $x \in (0, 1]$ . Let  $x'$  be the unique root for  $f$  in the interval  $(0, 1]$ . Then  $m'_{1,1} := -\log(x')$  is the unique rate in the interval  $[0, \infty)$  that satisfies the constraint. By using the rate ratios, we can obtain mortality rates for each sub-population at time  $t = 1$  in terms of  $m'_{1,1}$  such that the sub-population quantities for the alive compartment are consistent with the aggregate population at  $t = 1$ . In order to obtain a value for  $x'$ , one can use one of several root finding methods to obtain a root for  $f(x)$  in the interval  $(0, 1]$ , such as Brent’s method [32].

Once the sub-population mortality rates are known for  $t = 1$ , we can use the recursive formulae for the

alive and dead populations to obtain explicit values for the population quantities  $A_{k,1}$  and  $D_{k,1}$ . The above process can then be repeated for  $t = 2$  with the constraint

$$A_1 e^{-m_2} = \sum_{k=1}^n A_{k,1} e^{-r_{k,2}^{\text{mort}} m_{1,2}}$$

to obtain the unique rate  $m_{1,2}'$  which satisfies the constraint. By continuing to iterate this process, we can compute the sub-population mortality rates and population quantities up to any time period.

We summarize this procedure as the following algorithm.

---

Mortality Lifetable Disaggregation Algorithm

**Input:**

- $m_t, A_t$  for each timestep  $t$
- $A_{k,0}$  for each sub-population  $k$
- $r_{k,t}^{\text{mort}}$  for each sub-population  $k$  and timestep  $t$

**Output:**

- Values for  $m_{k,t}$  and  $A_{k,t}$  for each sub-population  $k$  and timestep  $t$  that are consistent with  $m_t, A_t$  at each timestep  $t$

for each timestep  $t$  in order from  $t = 1$

$$\text{Let } f := \sum_{k=1}^n A_{k,t-1} x^{r_{k,t}^{\text{mort}}} - A_{t-1} e^{-m_t} = 0$$

Find the unique root  $x'$  for  $f$  in the interval  $(0, 1]$

for  $k = 1, \dots, n$

$$m_{k,t} := -r_{k,t}^{\text{mort}} \log(x')$$

$$A_{k,t} := A_{k,t-1} e^{-m_{k,t}}$$


---

**Abbreviations**

ACMR: All-cause mortality rate; BAU: Business-as-usual; CHD: Coronary heart disease; HALY: Health adjusted life year; NZ: New Zealand; PIF: Population impact fraction; PMSLT: Proportional multistate lifetable; SES: Socioeconomic position; YLD: Years of life lived with disability.

**Supplementary Information**

The online version contains supplementary material available at <https://doi.org/10.1186/s12963-022-00282-7>.

**Additional file 1.** Supplementary Material.

**Acknowledgements**

Not applicable.

**Authors' contributions**

PA led the analyses and write-up. AM contributed to analyses and generated the input data by deprivation, and contributed to write-up. NW contribute to conceptualization, interpretation and write-up. AD contributed to analyses and generated the input data by deprivation, and contributed to write-up. LB contribute to analyses and write-up. TB initiated the paper, led the conceptualization, and finalized the drafting. All authors read and approved the final manuscript.

**Funding**

Nil.

**Availability of data and materials**

All data used in the calculations and examples in this paper are provided in supplementary tables.

**Declarations**

**Ethics approval**

Not applicable.

**Consent to publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>Population Interventions Unit, Centre for Epidemiology and Biostatistics, The University of Melbourne, 207-221 Bouverie St., Parkville, VIC 3010, Australia.

<sup>2</sup>Burden of Disease Epidemiology, Equity and Cost-Effectiveness Programme (BODE3), Department of Public Health, University of Otago, PO Box 7343, Wellington, Wellington South 6242, New Zealand.

Received: 24 February 2021 Accepted: 2 January 2022

Published online: 15 January 2022

**References**

1. Blakely T, Cobiac LJ, Cleghorn CL, Pearson AL, van der Deen FS, Kvizhidze G, et al. Health, health inequality, and cost impacts of annual increases in tobacco tax: multistate life table modeling in New Zealand. *PLoS Med.* 2015;12(7):e1001856.
2. Scarborough P, Allender S, Clarke D, Wickramasinghe K, Rayner M. Modelling the health impact of environmentally sustainable dietary scenarios in the UK. *Eur J Clin Nutr.* 2012;66(6):710–5.
3. Lal A, Mantilla-Herrera AM, Veerman L, Backholer K, Sacks G, Moodie M, et al. Modelled health benefits of a sugar-sweetened beverage tax across different socioeconomic groups in Australia: a cost-effectiveness and equity analysis. *PLoS Med.* 2017;14(6):e1002326.
4. Ananthapavan J, Sacks G, Brown V, Moodie M, Nguyen P, Barendregt J, et al. Assessing cost-effectiveness of obesity prevention policies in Australia 2018 (ACE-obesity policy). Melbourne: Deakin University; 2018.
5. Blakely T, Cleghorn C, Mizdrak A, Waterlander W, Nghiem N, Swinburn B, et al. The effect of food taxes and subsidies on population health and health costs: a modelling study. *Lancet Public Health.* 2020;5(7):e404–13.
6. Barendregt JJ, Van Oortmarsen GJ, Van Hout BA, Van Den Bosch JM, Bonneux L. Coping with multiple morbidity in a life table. *Math Popul Stud.* 1998;7(1):29–49 (109).
7. Blakely T, Moss R, Collins J, Mizdrak A, Singh A, Carvalho N, et al. Proportional multistate lifetable modelling of preventive interventions: concepts, code and worked examples. *Int J Epidemiol.* 2020;49:1624–36.
8. Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography.* 1979;16(3):439–54.
9. Rogers A. Heterogeneity and selection in multistate population analysis. *Demography.* 1992;29(1):31–8.
10. Lynch SM, Brown JS. Obtaining multistate life table distributions for highly refined subpopulations from cross-sectional data: a bayesian extension of Sullivan's method. *Demography.* 2010;47(4):1053–77.
11. Barendregt JJ, Baan CA, Bonneux L. An indirect estimate of the incidence of non-insulin-dependent diabetes mellitus. *Epidemiology.* 2000;11(3):274–9.
12. Hoogenveen RT, Boshuizen HC, Engelfriet PM, van Baal PHM. You only die once: accounting for multi-attributable mortality risks in multi-disease models for health-economic analyses. *Med Decis Mak.* 2017;37(4):403–14.
13. population-interventions/het\_module\_prototype [Internet]. Population Interventions; 2020 [cited 2020 Nov 5]. Available from: [https://github.com/population-interventions/het\\_module\\_prototype](https://github.com/population-interventions/het_module_prototype)
14. Woodward A, Blakely T. The healthy country? a history of life and death in New Zealand. Auckland: University of Auckland Press; 2014.

15. Nghiem N, Blakely T, Cobiac LJ, Cleghorn CL, Wilson N. The health gains and cost savings of dietary salt reduction interventions, with equity and age distributional aspects. *BMC Public Health*. 2016;16(1):423.
16. Barendregt JJ, Veerman JL. Categorical versus continuous risk factors and the calculation of potential impact fractions. *J Epidemiol Commun Health*. 2010;64(3):209–12.
17. Parnell W, Wilson N, Thomson C, Mackay S, Stefanogiannis N. A focus on nutrition: key findings of the 2008/09 New Zealand Adult Nutrition Survey. Wellington: Ministry of Health; 2011.
18. Barendregt JJ, van Oortmarsen GJ, Vos T, Murray CJ. A generic model for the assessment of disease epidemiology: the computational basis of DisMod II. *Popul Health Metr*. 2003;14(1):4.
19. India State-Level Disease Burden Initiative CRD Collaborators. The burden of chronic respiratory diseases and their heterogeneity across the states of India: the Global Burden of Disease Study 1990–2016. *Lancet Glob Health*. 2018;6(12):e1363–74.
20. Jenouvrier S, Aubry LM, Barbraud C, Weimerskirch H, Caswell H. Interacting effects of unobserved heterogeneity and individual stochasticity in the life history of the southern fulmar. *J Anim Ecol*. 2018;87(1):212–22.
21. McDonald SA, Devleeschauwer B, Wallinga J. The impact of individual-level heterogeneity on estimated infectious disease burden: a simulation study. *Popul Health Metr*. 2016;08(14):47.
22. Smith TA. Estimation of heterogeneity in malaria transmission by stochastic modelling of apparent deviations from mass action kinetics. *Malar J*. 2008;11(7):12.
23. Bartolucci F, Pennoni F. A class of latent Markov models for capture-recapture data allowing for time, heterogeneity, and behavior effects. *Biometrics*. 2007;63(2):568–78.
24. Zadnik V, Reich BJ. Analysis of the relationship between socio-economic factors and stomach cancer incidence in Slovenia. *Neoplasma*. 2006;53(2):103–10.
25. Ross JV, House T, Keeling MJ. Calculation of disease dynamics in a population of households. *PLoS ONE*. 2010;5(3):e9666.
26. Chu H, Préziosi M-P, Halloran ME. Estimating heterogeneous transmission with multiple infectives using MCMC methods. *Stat Med*. 2004;23(1):35–49.
27. Nour E. Mortality of a heterogeneous cohort; description and implications. *Biom J*. 1984;26(8):931–40.
28. Lynch SM, Brown JS. A new approach to estimating life tables with covariates and constructing interval estimates of life table quantities. *Sociol Methodol*. 2005;35:189–237.
29. Xavier A, Costa Freitas MdB, Fragoso R. Disaggregation of statistical livestock data using the entropy approach. *Adv Oper Res*. 2014;214:1–9.
30. Hougaard P. Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika*. 1984;71(1):75–83.
31. Avraam D, Arnold S, Vasieva O, Vasiev B. On the heterogeneity of human populations as reflected by mortality dynamics. *Aging (Albany NY)*. 2016;8(11):3045–64.
32. Brent RP. An algorithm with guaranteed convergence for finding a zero of a function. *Comput J*. 1971;14(4):422–5.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

