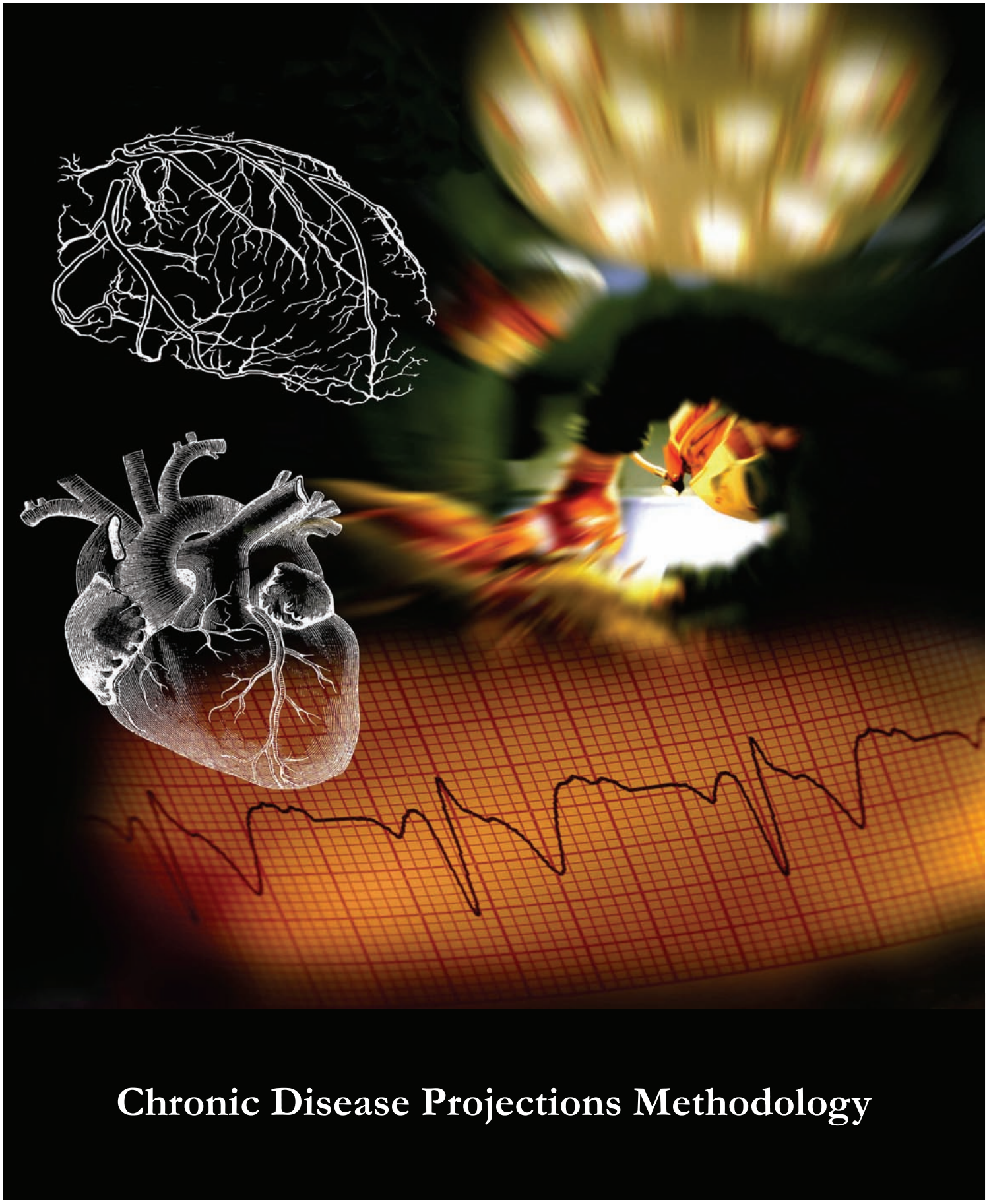


Chronic Disease Projections Methodology

Alberta

Surveillance and Environmental Health 2008

Acknowledgements

This work was completed by Shaun Malo.

The invaluable suggestions and feedback from the following individuals are gratefully acknowledged:

Larry Svenson, Alberta Health and Wellness
Dr. Donald Schopflocher, University of Alberta
Kimberly Simmonds, Alberta Health and Wellness
Douglas C. Dover, Alberta Health and Wellness
Yan Jin, Alberta Health and Wellness

For more information contact:

Surveillance and Environmental Health
Alberta Health and Wellness
24th Floor, Telus Plaza North Tower
10025 Jasper Avenue
P.O. Box 1360 Station Main
Edmonton, Alberta T5J 2N3
CANADA

Phone: 780-422-4771
Fax: 780-427-1470
Internet: www.health.alberta.ca

ISBN: 978-0-7785-6617-5

Contents

Introduction 1

Background – Disease prevalence projections 1

Methodology 3

1) Creating a case definition 3

2) Calculate, model, and project incidence risk 4

3) Calculate, model, and project incidence rates for new in-migrants 4

4) Calculate, model, and project death rates for those with the disease and those at risk 5

5) Calculate, model and project migration 6

6) Calculate, model and project fertility rates 6

7) Calculate, model and project remission rates 6

8) Applying the extended cohort component model 7

Appendix 1: Singular value decomposition (SVD) 9

Appendix 2: Methodology for modeling and projecting incidence risk 10

Appendix 3: Calculating survival rates 14

Appendix 4: Calculating migration 16

Appendix 5: Methodology for modeling and projecting migration 19

Appendix 6: Extended cohort component model 22

Glossary of notation 29

Introduction

The Surveillance and Environmental Health branch has developed a framework for projecting disease prevalence in Alberta. This report provides a detailed overview of the steps involved to develop disease prevalence projections for Alberta and its health regions.

The focus of this report is methodology, and does not include examples pertaining to any specific diseases. This report, intended to act as an "instruction manual", provides a clearly articulated methodological framework, with formulas, notation, and definitions, and can be a valuable resource to researchers involved in chronic disease projections.

The methodology section provides an overview of the eight steps to calculate prevalence projections. Calculation details are provided in appendices.

Background – Disease prevalence projections

Disease prevalence projections are a prediction of how many people in the future will have a certain attribute, in this case a disease. Disease prevalence projections are typically applied to chronic diseases.

The most obvious use for disease prevalence projections is to help make informed policy decisions. There are two main functions within a disease prevalence projections exercise, namely forecasting and scenario analysis. A forecast is based on a most likely scenario about the state of the future. The single forecast is then used to direct planning for various aspects of planning in the health system. When performing a scenario analysis, one is not so much interested in the absolute value of the projection, as in the effect changing factors have on the projections. For example, how does a small decrease in the incidence rate of a disease, or a small increase in the survival rate of people with a disease, affect how many people will have the disease in the future.

From a public health perspective, the incidence rate is typically the lever (i.e. prevention policy) that is considered for controlling how many people will have a disease in the future. If the effect a prevention policy has on incidence rates can be quantified, the policy can in turn be evaluated in terms of cost effectiveness by looking at how many fewer people will have the disease in the future.

Disease projections are typically generated using a variety of approaches ranging in complexity. Most commonly, the process involves projecting incidence rates (how quickly people will develop the disease) and death rates (the rate at which people with the disease will die). The methods for how these rates are modeled and projected also range in complexity.

Population projections created in a separate process are often applied to these projected rates, and projected new cases and deaths are added and subtracted from the current prevalence of a disease and simulated forward in time.

When using population projections as a basis for prevalence projections, a variety of issues should be considered including:

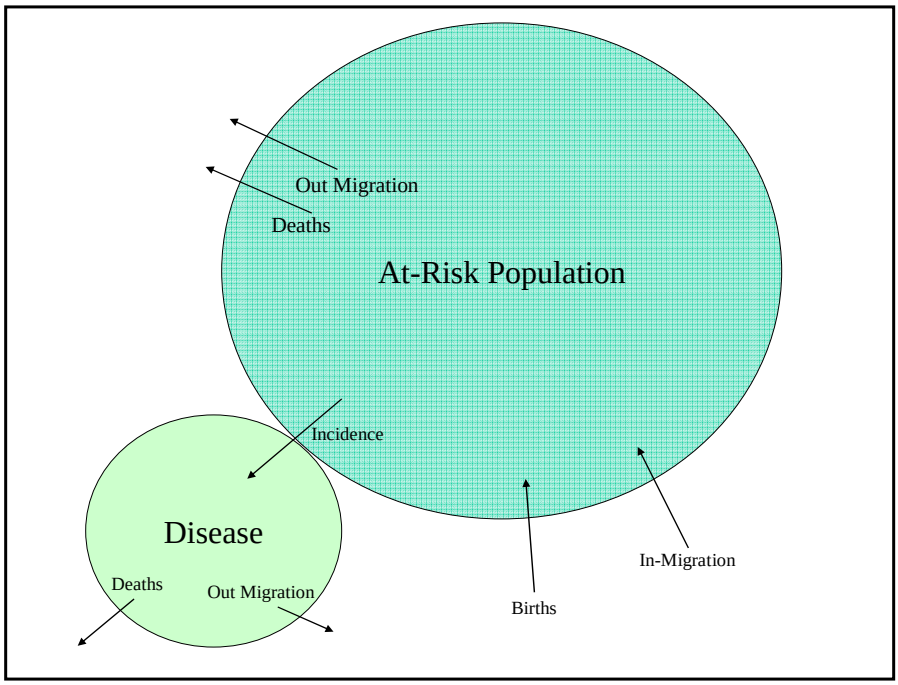
1) The projected incidence rates should use the projected at-risk populations in the denominators to determine new cases. Using population projections (either the value itself or the per cent change applied to existing at-risk population) as a proxy for the projected at-risk population is potentially inappropriate.

2) Scenario analysis of assessing how varying levels of future incidence affect the prevalence of the disease becomes illogical. Since death rates are typically different for people with and without a disease, changing the incidence will affect overall population.

The methodology proposed in this report assumes the at-risk population and the population with the disease are modeled simultaneously; which can be thought of as an enhanced population projections exercise where the disease status of an individual is measured in addition to age, sex, and geography. This deviates from commonly-used methods to develop disease prevalence projections.

Figure 1 shows how the model works for the single disease case. It can be seen that incidence becomes less influential on the overall population when i) the death rates of those with and without the disease are the same, and ii) the propensity of someone to out-migrate is the same for someone with or without the disease. Therefore the approach of using population projections from a separate process as the basis of calculating new cases becomes inappropriate when death rates and propensity to migrate externally are affected by the presence of the disease. Usually chronic conditions are associated with higher death rates, making the proposed methodology more appealing.

Figure 1: Flow Diagram; In and Out of Disease States



Source: Surveillance and Environmental Health, Alberta Health and Wellness

Methodology

The process to generate disease prevalence projections is complex. It is helpful to break the process into eight smaller steps in order to present the overall process in an understandable manner. This section describes in detail each of these eight steps, and is intended to be a blueprint for anyone wishing to duplicate prevalence projection work using this framework.

The eight steps are as follows:

1) Creating a case definition;
2) Calculate, model, and project incidence risk
3) Calculate, model, and project incidence rates for new in-migrants;
4) Calculate, model, and project death rates for those with the disease and those at risk;
5) Calculate, model, and project migration;
6) Calculate, model, and project fertility rates;
7) Calculate, model, and project remission rates;
8) Applying the 'extended' cohort component model;

1) Creating a case definition

The first step involves the establishment of a disease case definition.

Case definitions, which are the rules to determine who has the disease, can range in levels of complexity. For chronic diseases, they are often based on diagnosis information in administrative health data and can depend on various factors such as number of services, types of services, location of services, types of service providers, length of residency, drug information, etc.

In Alberta, the three most commonly used sources of administrative data for chronic disease case definitions are physician claims, hospitalization, and ambulatory care (of which emergency visits are usually of interest)

If a person meets all the criteria of a case definition, the next step is to determine when they developed the disease. This is usually called the incidence date , and is typically based on either i) the first date with a known relevant diagnosis, or ii) the date in which the individual meets the criteria of the case definition.

Another issue to be aware of in establishing case definitions is that, depending on the criteria used, a case could be "retroactively" changed when new administrative data becomes available. For example, in 2007 a person had one visit to the doctor in 2005 with a relevant diagnosis, but the criteria of the case definition require two visits, so the person is not counted as a case. In 2009, the person saw the doctor again and now meets the criteria. This person now has an incidence date for 2005 and is a prevalent case in each subsequent year. This concept of having case definition criteria that can result in retroactive changes to previously reported incidence and prevalence figures is important, especially when dealing with prevalence projections, where performance of past projections will need to be evaluated. The evaluation will become convoluted if prevalence figures can change several years after the projections were created.

In addition, the historical trends of incidence counts, for which the prevalence projections are based, would be inconsistent since the incidence counts for the earlier years use more complete information than later years.

It is therefore recommended, to avoid issues of retroactive changes to incidence and prevalence counts, to use a case definition where the incidence date is determined at the time the case definition is satisfied, and to not base the criteria in the case definition on the length of residency.

2) Calculate, model, and project incidence risk

Incidence risk is the probability that someone, who is at risk of developing the disease at the start of the year, will develop the disease within the upcoming year. Incidence risk measures what proportion of a population will develop the disease.

Incidence rate measures the rate (how quickly) at which people develop the disease.

The decision to use incidence risk arrives from the need to look separately at the incidence of:

- people currently at risk in Alberta as of mid-year, and
- people not in Alberta as of mid-year, or new migrants.

Incidence risk (denoted by $I_{t-1,t,x,r}$) is the probability that someone of age x living in region r, who is at risk of developing disease, as of June 30 of year t-1, will develop disease from July 1 of year t-1 to June 30 or year t.

Therefore, if $NEWCASE_{t-1,t,x,r}$ refers to the number of people of age x, who develop the disease between July 1, of year t-1 to June 30, of year t, in region r, who were present in the at-risk population ($P^{AR}_{t-1,x,r}$) as of June 30 of year t-1, then $I_{t-1,t,x,r} = \frac{NEWCASE_{t-1,t,x,r}}{P^{AR}_{t-1,x,r}}$.

The incidence risk is modeled and projected, for males and females separately using the singular value decomposition.

Appendix 1 - Defines the singular value decomposition (SVD).

Appendix 2 - Explains the process for modeling and projecting incidence risk.

3) Calculate, model, and project incidence rates for new in-migrants

The rate at which new in-migrants develop the disease needs to be looked at separately from those who are already in the population at the start of the year.

$NEWMIG_{t,t+1,x,r}$ can be denoted as the number of people who develop the disease (at x years of age) between July 1, of year t to June 30, of year t+1, in region r, who were not present in the at-risk population at the start of the year (June 30 of year t), but were present as of June 30 of year t+1.

The incidence rate for new migrants can then be computed as follows:

$$\text{INCNEW}_{t,t+1,x,r} = \frac{\text{NEWMIG}_{t,t+1,x,r}}{\frac{1}{2} \times \text{EXMG}_{t,t+1,x,r}},$$

where $\text{EXMG}_{t,t+1,x,r}$ is the number of people who move into region r, from outside of Alberta from July 1 of year t to June 30 of year t+1, at age x. The migration is divided by two, since it is assumed that an individual will on average arrive half way through the year.

Appendix 4 – Shows calculation details for the various types of migration.

The projected values of incidence rates for new migrants can be generated using the Lee-Carter method, as for the case of incidence risks. Since these rates are typically less stable from year to year than those for incidence risk (since dealing with smaller populations), it may be appropriate to average the provincial rates over all historical years and use these 'averaged' rates for each projection year and health region.

4) Calculate, model, and project death rates for those with the disease and those at risk

In the situation of projecting prevalence for one disease, two sets of death rates need to be calculated; one for those with the disease, and one for the at-risk population.

Age specific death rates are calculated for males and females separately. Age specific death rates for the at-risk population are calculated as:

$$\text{ASDR}^{\text{AR}}_{t,x,r} = \frac{\text{D}^{\text{AR}}_{t,x,r}}{\text{P}^{\text{AR}}_{t,x,r}},$$

where $\text{D}^{\text{AR}}_{t,x,r}$ is the number of people in year t, who are at risk of getting the disease, residing in health region r, dying at age x. $\text{P}^{\text{AR}}_{t,x,r}$ is the population at risk of developing the disease, x years of age, in health region r, as of June 30 of year t.

Age specific death rates for those with the disease are calculated as:

$$\text{ASDR}^{\text{DIS}}_{t,x,r} = \frac{\text{D}^{\text{DIS}}_{t,x,r}}{\text{P}^{\text{DIS}}_{t,x,r}},$$

where $\text{D}^{\text{DIS}}_{t,x,r}$ is the number of people in year t, who have the disease, residing in health region r, dying at age x. $\text{P}^{\text{DIS}}_{t,x,r}$ is the population (prevalence) with the disease, x years of age, in health region r, as of June 30 of year t.

The age-specific death rates are analyzed and projected using the Lee-Carter methodology, just as for incidence risks.

Appendix 2 - Explains the process for modeling and projecting incidence risk.

In order to be applicable for use in the projection of disease prevalence, the projected death rates are converted into survival rates (or sometimes referred to as survivorship ratios) using life tables, that is the probability that someone of a given age, x, at a given point in time (i.e. June 30 of year t-1) will survive to June 30 of year t.

Appendix 3 – Shows methodology of converting death rates to survival rates.

5) Calculate, model and project migration

Migration can be broken into inter-regional migration and external migration. Inter-regional migration refers to the movement of a person within the province from one health region to another during a year. External migration refers to movement of a person from Alberta to somewhere outside of Alberta, or vice-versa.

Inter-regional and external migration are both analyzed separately for gains and losses. The four migration measures (inter-regional migration gains and losses and external migration gains and losses) are modeled and projected using the singular value decomposition.

Appendix 4 – Shows calculation details for migration.

Appendix 5 - Explains the process for modeling and projecting migration.

External migration losses are calculated for the entire population. It is assumed that the proportion of the general population with the disease is applied to external migration losses. For example, if there are 1,000 people migrating outwards, and 20 per cent of the population has the disease, then it is assumed that 200 of the people migrating will have the disease. The feasibility of such an assumption may vary by disease.

6) Calculate, model and project fertility rates

Fertility rates affect the prevalence of a disease, if the disease of interest is one that is prevalent in young people or if projections are going significantly into the future. If fertility is relevant, the rates would be modeled in the same way as incidence, which is outlined in appendix 2.

Furthermore, if the disease is one which is reasonably prevalent in women in their childbearing years, it would need to be examined as to whether fertility rates need to be analyzed separately for women with or without the disease, since the presence of the disease may effect fertility.

7) Calculate, model and project remission rates

Remission will need to be modeled if the disease of interest is one where the disease can be absent for any period of time, whether temporary or permanent. It would be modeled in the same way as incidence, which is outlined in appendix 2.

8) Applying the extended cohort component model

In its most basic form, without any reference to age or geography, the general cohort component model can be depicted as $\text{Pop}_t = \text{Pop}_{t-1} + \text{Births}_{t-1,t} - \text{Deaths}_{t-1,t} + \text{NetMigration}_{t-1,t}$. That is the population in year t will equal the population in the previous year. Population is adjusted for births, deaths, and migration during the one year from t-1 to t.

This cohort component method is the primary approach used by demographers to calculate population projections. That is, the components of population change (mortality, fertility, and migration) are each projected separately, and then input into the cohort component model to project the population.

A model for generating disease prevalence projections is called, in this report, an extended cohort component model, where the components of population change (mortality, fertility, migration, and incidence) are each projected separately. The extended cohort component model, in its most basic form (one disease), with no reference to age or geography under the assumption that one can't be born with the disease, can then be re-written as equations 1 and 2 below.

$$1)\ P_t^{DIS} = P_{t-1}^{DIS} + (P_{t-1}^{AR} \times I_{t-1}) + \frac{1}{2}(EXMG_{t-1,t} \times INCNEW_{t-1,t}) - D_{t-1,t}^{DIS} - EXML_{t-1,t}^{DIS}$$

where,

- P_t^{DIS} is the population (prevalence) with the disease at June 30 of year t.
- P_{t-1}^{DIS} is the population (prevalence) with the disease at June 30 of year t-1.
- P_{t-1}^{AR} is the at-risk population at June 30 of year t.
- I_{t-1} is the probability that someone who is at risk of developing the disease on June 30 of year t-1 will develop the disease from July 1 of year t-1 to June 30 of year t.
- $EXMG_{t-1,t}$ is the number of people moving into Alberta from July 1 of year t-1 to June 30 of year t.
- $INCNEW_{t-1,t}$ is the rate at which new migrants develop the disease.
- $D_{t-1,t}^{DIS}$ is the number of people with the disease who die between July 1 of year t-1 to June 30 of year t.
- $EXML_{t-1,t}^{DIS}$ is the number with the disease of people who migrate out of Alberta from July 1 of year t-1 to June 30 of year t.

$$2)\ P_t^{AR} = P_{t-1}^{AR} + B_{t-1,t} - D_{t-1,t}^{AR} + (EXMG_{t-1,t} \times (1 - \frac{1}{2} INCNEW_{t-1,t})) - EXML_{t-1,t}^{AR}$$

where,

- P_t^{AR} is the at-risk population at June 30 of year t.

- P_{t-1}^{AR} is the at-risk population at June 30 of year t-1.
- $B_{t-1,t}$ is the number of births from July 1 of year t-1 to June 30 of year t.
- $D_{t-1,t}^{AR}$ is the number of people at risk of developing the disease who die from July 1 of year t-1 to June 30 of year t.
- $1 - INCNEW_{t-1,t}$, is the complement of the incidence rate for new migrants.
- $EXML_{t-1,t}^{AR}$ is the number of people who migrate out of Alberta from July 1 of year t-1 to June 30 of year t who are at-risk of developing the disease

3) $P_t = P_t^{DIS} + P_t^{AR}$, where P_t is the population of Alberta at June 30 of year t.

In order to project the number of people with the disease, populations with and without the disease need to be projected simultaneously, since it is from the at-risk population where most new cases arise.

These equations are valid under the assumptions that:

- people can't be born with the disease, but must develop it after birth
- there is no remission of the disease
- new migrants into Alberta can't have the disease, but must develop it after in-migrating.

Appendix 6 – Provides detailed equations for application of the extended cohort component model with reference to age and geography.

Appendix 1: Singular value decomposition (SVD)

The singular value decomposition decomposes a matrix into three matrices. For example if A is an m x n real matrix with m > n then A has the form

$A = U D V^T$

Where U is an m by n matrix, V is a square matrix; both having orthogonal columns (i.e. $U^TU = V^TV = I$. D is an n by n diagonal matrix. Typically the matrices are organized such that the values of the diagonal of D are decreasing.

If the values of the diagonal of D are decreasing, a model with significantly lower rank may be able to adequately reproduce the original matrix A. Suppose A is a p by k (p single year of ages by k years) matrix of mortality rates. Applying the SVD to the matrix A, results in matrix of age parameters (U = k by p), a matrix of singular values (D = k by k) and a matrix of time parameters (V = k by k). The first vectors of U and V, and the first element of the diagonal of D, will result in an estimate of the matrix A. If the first singular value is high in proportion to the total of all the singular values, the estimated matrix will explain a high proportion of the variation in A. Adding more vectors will improve the estimate of A, until eventually including all vectors will completely reproduce A.

This approach has great appeal because the coefficients of only a few time components may need to be projected to reproduce the projected values for A. The SAS software package can be used to apply the singular value decomposition to a matrix using the IML procedure.

Appendix 2: Methodology for modeling and projecting incidence risk

The methodology for modeling and projecting regional incidence risks by single year of age and sex, using the Singular Value Decomposition (see appendix 1), is summarized in the following steps.

For each sex:

1) Construct an array of incidence risks $INC_{i,j,k}$ where

i ranges across year of age from 1 to I,
j ranges across different regions, 1 to J,
k ranges across different calendar years 1 to K.

$INC_{i,j,k} = \frac{NEW_{i,j,k,k+1}}{P^{AR}_{i,j,k}}$, $P^{AR}_{i,j,k}$ is the at-risk population, aged i years in region j in year k, and $NEW_{i,j,k,k+1}$ is the number of people from the at-risk population, aged i years in region j who develop the disease in the forthcoming year (k to k+1).

2) Collapse the array of incidence risks across regions to generate a matrix of provincial incidence risks

$$INC_{i,.,k} = \frac{\sum_{j=1}^{J} NEW_{i,j,k,k+1}}{\sum_{j=1}^{J} P^{AR}_{i,j,k}}$$

3) Compute the log-centered matrix of incidence risks, according to the method of Lee and Carter

$$L_{i,.,k} = \ln(INC_{i,.,k}) - \frac{\sum_{k=1}^{K} \ln(INC_{i,.,k})}{K}$$

4) Apply the singular value decomposition to $L_{i,.,k}$, giving

$L_{i,.,k} = AD Y'$, where A is an I by K matrix of age components, D is a K by K matrix of singular values, and Y is a K by K matrix of time components.

5) Determine the number of components, N, needed to appropriately reproduce the matrix $L_{i,.,k}$

$$\hat{L}_{i,.,k} = \sum_{n=1}^{N} A_n D_n Y'_n$$

Where A_n is the n^{th} component (column) of A, D_n is the n^{th} element along the ordered diagonal of singular values of D, and Y_n is the n^{th} component of Y.

6) Adjust backwards to original units

$$\hat{INC}_{i,.,k} = \exp\{\hat{L}_{i,.,k} + \frac{\sum_{k=1}^{K} \ln(INC_{i,.,k})}{K}\}$$

7) Uncollapse the values of $\hat{INC}_{i,.,k}$ across J regions so

$C1_{i,j,k} = \hat{INC}_{i,.,k}$ for all values of j = 1 to J

8) Compute the array of residuals $RES1_{i,j,k} = INC_{i,j,k} - C1_{i,j,k}$

9) Collapse $RES1_{i,j,k}$ across K years

$$RES1_{i,j,.} = \frac{\sum_{k=1}^{K} RES1_{i,j,k}}{K}$$

10) Let $RES1^{*}_{i,j,.} = \text{Sqrt}(RES1_{i,j,.})$

11) Apply the singular value decomposition to $RES1^{*}_{i,j,.}$, giving

$RES1^{*}_{i,j,.} = ADR'$, where A is an I by J matrix of age components, D is a J by J matrix of singular values, and R is a J by J matrix of RHA components.

12) Determine the number of components, N, needed to appropriately reproduce the matrix $RES1^{*}_{i,j,.}$.

$$\hat{RES1}^{*}_{i,j,.} = \sum_{n} A_n D_{n'} R_n$$

Where A_n is the n^{th} component (column) of A, D_n is the n^{th} element along the ordered diagonal of singular values of D, and R_n is the n^{th} component of R.

13) Transform $\hat{RES1}^{*}_{i,j,.}$ back to original units, $\hat{RES1}_{i,j,.} = \text{sign} \times (\hat{RES1}^{*}_{i,j,.})^2$

Where sign=+1 if $\hat{RES1}^{*}_{i,j,.} > 0$ and sign=-1 if $\hat{RES1}^{*}_{i,j,.} < 0$

14) Uncollapse the values of $\hat{RES1}_{i,j,.}$ across k years so $C2_{i,j,k} = \hat{RES1}_{i,j,.}$ for all k=1 to K.

15) Compute the array of residuals $RES2_{i,j,k} = RES1_{i,j,k} - C2_{i,j,k}$.

16) Collapse $RES2_{i,j,k}$ across I ages by applying a weighted average,

If $W_{i,j,k} = \dfrac{P_{i,j,k}}{\sum_{i=1}^{I} P_{i,j,k}}$ is the weight at each age, then $RES2_{.,j,k} = \sum_{i=1}^{I} W_{ijk} \times RES2_{i,j,k}$

17) Apply the SVD to $RES2_{.,j,k}$ (apply SVD to the K by J matrix since J < K)

$RES2_{.,j,k} = YDR'$, where Y is an K by J matrix of time components, D is a J by J matrix of singular values, and R is a J by J matrix of region components.

18) Determine the number of components, N, needed to appropriately reproduce the matrix $RES2_{.,j,k}$

$$\hat{RES2}_{.,j,k} = \sum_{n=1}^{N} Y_n D_n R'_n$$

Where Y_n is the n^{th} component (column) of Y, D_n is the n^{th} element along the ordered diagonal of singular values of D, and R_n is the n^{th} component of R.

19) Uncollapse the values of $\hat{RES2}_{.,j,k}$ across I years of age
so $C3_{i,j,k} = \hat{RES2}_{.,j,k}$ for all i=1 to I.

20) The original matrix of incidence risks $INC_{i,j,k}$ is then estimated as

$\hat{INC}_{i,j,k} = C1_{i,j,k} + C2_{i,j,k} + C3_{i,j,k}$, where C1 describes how provincial level incidence over time changes with age; C2 describes regional differences of incidence against age, and C3 describes how incidence over time changes across regions.

21) The final step is to project the time components forward p years, throughout the projection period. $C1_{i,j,k}$ and $C3_{i,j,k}$ are generated for the k=K+1 to k=K+p future years using the

projected values of the time components, while $C2_{i,j,k}$ remains invariant throughout the projection period. The projected arrays of incidence risks are then calculated as

$$\hat{INC}_{i,j,k} = C1_{i,j,k} + C2_{i,j,k} + C3_{i,j,k} \text{ for all } k=K+1 \text{ to } K+p$$

Appendix 3: Calculating survival rates

Survival rates are calculated from life tables, which are tables that contain information about how a hypothetical birth cohort would die out. Period life tables are used, which use the death information from a fixed period of time (in this case one year). The information in the life table can be interpreted as what would happen to a cohort of people, if these mortality patterns were to continue throughout their lifetime.

The life table below was constructed using actual 2005 data for males in Alberta. The definition and calculation detail for each column are shown below. Unabridged life tables are used (i.e. one-year age groups), since projections are generated by one-year age groups.

Age	D_x	P_x	m_x	a_x	q_x	p_x	l_x	d_x	L_x	T_x	e^0_x	S_x
0	159	20,847	0.00763	0.1	0.00757	0.99243	100,000	757.5	99,318	7,754,879	77.55	0.99886
1	16	21,109	0.00076	0.5	0.00076	0.99924	99,243	75.2	99,205	7,655,561	76.56	0.99950
2	5	20,543	0.00024	0.5	0.00024	0.99976	99,167	24.1	99,155	7,556,356	75.56	0.99983
3	2	19,819	0.00010	0.5	0.00010	0.99990	99,143	10.0	99,138	7,457,201	74.57	0.99987
4	3	19,964	0.00015	0.5	0.00015	0.99985	99,133	14.9	99,126	7,358,062	73.58	0.99978
5	6	20,531	0.00029	0.5	0.00029	0.99971	99,118	29.0	99,104	7,258,937	72.59	0.99969
.												
85	279	2,649	0.105323	0.5	0.10005	0.89995	36,451	3,647.0	34,627	215,872	2.16	0.89520
86	231	1,983	0.11649	0.5	0.11008	0.88992	32,804	3,611.0	30,998	181,244	1.81	0.87565
87	252	1,669	0.150989	0.5	0.14039	0.85961	29,193	4,098.4	27,144	150,246	1.50	0.85493
88	230	1,413	0.162774	0.5	0.15052	0.84948	25,094	3,777.3	23,206	123,102	1.23	0.85478
89	174	1,165	0.149356	0.5	0.13898	0.86102	21,317	2,962.6	19,836	99,897	1.00	0.80144
90+	840	3,664	0.229258	-	1.00000	0.00000	18,355	18,354.5	80,061	80,061	0.80	-

1) D_x is the actual number of people dying at age x.

2) P_x is the 2005 mid-year population at age x.

3) m_x is the age-specific death rate for people of age x; $m_x = \frac{D_x}{P_x}$.

4) a_x is called the separation factor, and represents the average person-years lived at age x for people dying at age x. For those dying at age 0, it is assumed they will on average live a tenth of a year (a_0=0.1) and for all other ages it is set at 0.5.

5) q_x is the probability of dying between the ages x and x+1. More specifically, it is the proportion of people who are alive at exact age x (i.e. a birthday) who die before reaching exact age x+1. The exception is for the age interval 90+, where 100% of people will die before exiting the age interval, which is open ended.

For ages x= 0 to 89, $q_x = \frac{m_x}{1+(a_x \times m_x)}$; for age x=90+, q_x=1.

6) p_x is the probability of surviving from exact age x to exact age x+1; $p_x = 1 - q_x$.

7) l_x is the number of people left alive at age x; $l_0 = 100{,}000$ and $l_{x+1} = l_x \times p_x$.

8) d_x is the number of people dying at age x (out of the hypothetical cohort of 100,000 people); $d_x = l_x - l_{x+1}$.

9) L_x is the number of person years lived at age x; $L_x = l_{x+1} + (a_x \times d_x)$.

10) T_x is the total person years lived above age x. i.e. these 100,000 people would live a total 7,754,699 person years; $T_x = \sum_{i=x}^{\infty} L_i$.

11) e_x^0 is the expectation of life at age x; $e_x^0 = \frac{T_x}{l_x}$.

12) Survivorship ratios are usually used in population process modeling. They are calculated as $S_x = \frac{L_{x+1}}{L_x}$ for ages x = 0 to 88.The survivorship ratios (S_x) and the survival probabilities above (p_x) are exactly the same when a_x=1.

Since in a population process modeling exercise, a stock of 86 year olds that are being survived forward for one year could have people ranging from their 86th birthday to almost their 87th birthday, S_x is used instead of p_x. For example S_{86}= .876, would be the proportion of 86 year olds (i.e. those who are somewhere from exactly 86 years to 86 years + 364 days) who would be 87 years of age 1 year later.

The one year survivorship ratio for the oldest age group (89 and above) is calculated as $S_{89} = \frac{T_{90+}}{T_{89}}$.

In other words, 80,061/99,897 = 80.1% of people over the age of 89 and over would be expected to live one more year. This survivorship ratio would be applied to those who are 89 years of age, as well as those 90 and above.

Appendix 4: Calculating migration

Migration data is obtained from the Alberta Health Care Insurance Stakeholder Plan Registry. The key to compiling migration data is to merge population files from two successive years, as of June 30 in each year, by individual recipient identifier. Subsequently, each individual can be assigned into one of the following statuses:

1. No Change: Individual resides on the registry in each year with no change in health region from one year to the next.

2. Inter-Region Migration (IRM): Individual resides on the registry in each year, and has a change in health region from one year to the next.

3. NEW: Individual does not exist on the registry in the first year, but appears in the second year.

4. EXIT: Individual appears on the registry in the first year, but does not appear in the second year.

Suppose t-1 and t are two successive years of interest; that is, migration is being calculated from June 30, of year t-1 to June 30 of year t.

Let $POP_IRM_{t,x,r}$ equal the number of individuals with a status of IRM, aged x, and residing in Health Region r on June 30 of year t.

Let $POP_NEW_{t,x,r}$ equal the number of individuals with a status of NEW, aged x, and residing in Health Region r on June 30 of year t.

Let $POP_EXIT_{t-1,x,r}$ equal the number of individuals with a status of EXIT, aged x, and residing in Health Region r on of June 30 of year t-1.

For Internal Migration gains and losses, the calculation details are as follows for ages 1 to 89.

$IRMG_{t-1,t,x,r}$ is the number of individuals aged x, who move to Health Region r from another Health Region in Alberta, between July 1, of year t-1, and June 30 of year t. (i.e. the inter-regional migration gain)

$IRML_{t-1,t,x,r}$ is the number of individuals aged x, who move from Health Region r to another Health Region in Alberta, between July 1, of year t-1, and June 30 of year t. (i.e. the inter-regional migration loss)

It is not possible, given the current data, to know at what age the individual actually moved from one region to another. For example: a male appears in region 2 on June 30, 1996, at the age of 15, and appears in region 3 on June 30, 1997, at age 16. We assume that it is equally likely that this person moved at the age 15 as he did at the age of 16. Therefore,

$$IRMG_{t-1,t,x,r} = \frac{1}{2} \times (POP_IRM_{t,x,r} + POP_IRM_{t,x+1,r})$$

For example; to calculate the number of males who moved into region 2 from July 1, 1995 to June 30, 1996 at the age of 15, we would average one half of the individuals, with a status of IRM, who were aged 15 on June 30, 1996, and one half of the individuals, with a status of IRM, who were aged 16 on June 30, 1996.

Similarly,

$$IRML_{t-1,t,x,r} = \frac{1}{2} \times (POP_IRM_{t-1,x,r} + POP_IRM_{t-1,x-1,r})$$

For the open-ended age group 90+, the calculation for net inter-regional migration gains and losses is as follows:

$$IRMG_{t-1,t,90+,r} = \frac{1}{2} \times (POP_IRM_{t,90,r}) + POP_IRM_{t,91+,r}$$

and

$$IRML_{t-1,t,90+,r} = \frac{1}{2} \times (POP_IRM_{t-1,89,r}) + POP_IRM_{t-1,90+,r}$$

For External Migration gains and losses, the calculation details are as follows for ages 1 to 89.

$EXMG_{t-1,t,x,r}$ is the number of individuals aged x, who move to Health Region r from outside of Alberta, between July 1, of year t-1, and June 30 of year t. (i.e. the external migration gain).

$EXML_{t-1,t,x,r}$ is the number of individuals aged x, who move from Health Region r to somewhere outside of Alberta, between July 1, of year t-1, and June 30 of year t. (i.e. the external migration loss).

$DTHS_{t-1,t,x,r}$ is the number of deaths from July 1, of year t-1, to June 30 of year t, of individuals aged x in Health Region r.

$$EXMG_{t-1,t,x,r} = \frac{1}{2} \times (POP_NEW_{t,x,r} + POP_NEW_{t,x+1,r})$$

$$EXML_{t-1,t,x,r} = \frac{1}{2} \times (POP_EXIT_{t-1,x,r} + POP_EXIT_{t-1,x-1,r}) - DTHS_{t-1,t,x,r}$$

For the open-ended age group 90+, the calculation for external migration gains and losses is as follows:

$$EXMG_{t-1,t,90+,r} = \frac{1}{2} \times (POP_NEW_{t,90,r}) + POP_NEW_{t,91+,r}$$

$$EXML_{t-1,t,90+,r} = \frac{1}{2} \times (POP_EXIT_{t-1,89,r}) + POP_EXIT_{t-1,90+,r} - DTHS_{t-1,t,90+,r}$$

The calculation for migration gains and losses of individuals' aged 0 is more involved than for those between 1 and 89 years. The majority of individuals aged 0 with a status of NEW, are not migrants, but new births. Also one cannot determine if a new individual at age 0 that is not a new birth, is actually migrating from another RHA or from outside of Alberta. The migration for age 0 is then a mix of external and inter-regional migration.

$NMG_{t-1,t,x,r}$ is the number of individuals aged 0, who move to Health Region r from somewhere outside of Health Region r, between July 1, of year t-1, and June 30 of year t.

$NML_{t-1,t,x,r}$ is the number of individuals aged 0, who move leave Health Region r to go somewhere outside of Health Region r, between July 1, of year t-1, and June 30 of year t.

$$NMG_{t-1,t,0,r} = \frac{1}{2} \times (POP_NEW_{t,1,r}) + (POP_NEW_{t,0,r}) + \frac{1}{2} \times (POP_IRM_{t,1,r}) - BRTH_{t-1,t,r}$$

$$NML_{t-1,t,0,r} = \frac{1}{2} \times (POP_EXIT_{t-1,0,r}) + \frac{1}{2} \times (POP_IRM_{t-1,0,r}) - DTHS^{*}_{t-1,t,0,r}$$

*not all deaths for individuals aged 0 should be included, but only the deaths of those aged 0, whose date of birth was prior to June 30 of year t-1. This is to avoid subtracting off the death of infants who were both born and died within the period from July 1 of year t-1 to June 30 of year t.

Appendix 5: Methodology for modeling and projecting migration

The methodology for modeling and projecting inter-regional migration losses by year of age, is summarized in the following steps.

1) Construct an array of inter-regional migration losses $IRML_{ijk}$ where

   i ranges across year of age from 1 to I,
   j ranges across different regions, 1 to J,
   k ranges across different calendar years 1 to K.

   Detail about the calculation of IRML is in Appendix 2.

2) Collapse the array of inter-regional migration losses across regions to generate a matrix of provincial inter-regional migration (i.e. the number of people moving inter-regionally in Alberta)

$$IRML_{i.k} = \sum_{j=1}^{J} IRML_{ijk}$$

3) Apply the singular value decomposition to $IRML_{i.k}$, giving

   $IRML_{i.k} = AD\,Y'$, where A is an I by K matrix of age components, D is a K by K matrix of singular values, and Y is a K by K matrix of time components.

4) Determine the number of components, N, needed to appropriately reproduce the matrix $IRML_{i.k}$

$$\hat{IRML}_{i.k} = \sum_{n=1}^{N} A_n D_n Y'_n$$

   A_n is the n^{th} component (column) of A, D_n is the n^{th} element along the ordered diagonal of singular values of D, and Y_n is the n^{th} component of Y.

5) Distribute the values of $\hat{IRML}_{i.k}$ across J regions according to the regions population distribution

   $C1_{ijk} = \hat{IRML}_{i.k} \times W_{.j.}$ for all values of j = 1 to J, where

$$W_{.j.} = \sum_{k=1}^{K}\sum_{i=1}^{I} P_{ijk} \;/\; \sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{i=1}^{I} P_{ijk}$$

   $W_{.j.}$ is constant across all i and k in each region.

6) Compute the array of residuals $RES1_{ijk} = IRML_{ijk} - C1_{ijk}$

7) Collapse $RES1_{ijk}$ across K years

$$RES1_{ij.} = \frac{\sum_{k=1}^{K} RES1_{ijk}}{K}$$

8) Apply the singular value decomposition to $RES1_{ij.}$, giving

$RES1_{ij.} = ADR'$, where A is an I by J matrix of age components, D is a J by J matrix of singular values, and R is a J by J matrix of RHA components.

9) Determine the number of components, N, needed to appropriately reproduce the matrix $RES1_{ij.}$

$$\hat{RES}1_{ij.} = \sum_{n=1}^{N} A_n D_n R'_n$$

A_n is the n^{th} component (column) of A, D_n is the n^{th} element along the ordered diagonal of singular values of D, and R_n is the n^{th} component of R.

10) Uncollapse the values of $\hat{RES}1_{ij.}$ across k years so $C2_{ijk} = \hat{RES}1_{ij.}$ for all k=1 to K.

11) Compute the array of residuals $RES2_{ijk} = RES1_{ijk} - C2_{ijk.}$

12) Collapse $RES2_{ijk}$ across I ages by summing across ages,

$$RES2_{.jk} = \sum_{i=1}^{I} RES2_{ijk}$$

13) Apply the SVD to $RES2_{.jk}$ (apply SVD to the K by J matrix since J < K)

$RES2_{.jk} = YDR'$, where Y is an K by J matrix of time components, D is a J by J matrix of singular values, and R is a J by J matrix of region components.

14) Determine the number of components, N, needed to appropriately reproduce the matrix $RES2_{.jk}$

$$\hat{RES}2_{.jk} = \sum_{n=1}^{N} Y_n D_n R'_n$$

Y_n is the n^{th} component (column) of Y, D_n is the n^{th} element along the ordered diagonal of singular values of D, and R_n is the n^{th} component of R.

15) Distribute the values of $\hat{RES}2_{.jk}$ across I years of age so

$C3_{ijk} = \hat{RES}2_{.jk} \times W_{ij.}$ for all i=1 to I, j=1 to J,

where $W_{ij.} = \sum_{k=1}^{K} P_{ijk} \; / \; \sum_{k=1}^{K}\sum_{i=1}^{I} P_{ijk}$

16) The original matrix of inter-regional migration losses $IRML_{ijk}$ is then estimated as

$\hat{IRML}_{ijk} = C1_{ijk} + C2_{ijk} + C3_{ijk}$, where C1 describes how total people moving inter-regionally over time changes with year of age; C2 describes how regional inter-regional migration losses vary across year of age, and C3 describes how regional inter-regional migration losses change over time.

17) The final step is to project the time components forward p years, throughout the projection period. $C1_{ijk}$ and $C3_{ijk}$ are generated for the k=K+1 to k=K+p future years using the projected values of the time components, while $C2_{ijk}$ remains invariant throughout the projection period. The projected arrays of inter-regional migration losses are then calculated as

$\hat{IRML}_{ijk} = C1_{ijk} + C2_{ijk} + C3_{ijk}$ for all k=K+1 to K+p

18) Inter-regional migration gains are handles in the same manner as losses, except in step 5, when distributing total people moving inter-regionally across regions, the allocations are applied differently.

For losses, the total number of people moving inter-regionally is allocated based on the region's population, assuming that inter-regional losses in a region would, all things being equal, occur based on a regions population. If the losses are distributed in this manner, then gains are distributed differently, namely

$C1_{ijk} = \hat{IRMG}_{i.k} \times Q_{j.}$ for all values of j = 1 to J, where

$Q_{j.} = \sum_{z \neq j} \{W_{.z.} \times W_{.j.} / (\sum_{z \neq j} W_{.z.})\}$ where

$W_{.j.} = \sum_{k=1}^{K}\sum_{i=1}^{I} P_{ijk} \; / \; \sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{i=1}^{I} P_{ijk}$

Appendix 6: Extended cohort component model

The formulas used to calculate projected populations for individuals with the disease, and those at risk of developing the disease are shown below. The formulas shown are for males, and all of the values in the formula are intended to be sex-specific. For example, there would be a value of $P^{DIS}_{t-1,x-1,r}$ for males, and another for females. The only difference in the form of the equations, when applying the model for females, would be replacing the value of "1- Prop_F_EXMG" (for males) with "Prop_F_EXMG" (for females). The equations are shown separately for ages x= 1 to 89 and the open ended age group 90+. The equations for the birth cohort (age x=0) are not shown here since they have not been needed to date (i.e. model has only been applied to date for diseases that are for older populations).

Projecting disease prevalence, ages x= 1 to 89:

$$P^{DIS}_{t,x,r} = P^{DIS}_{t-1,x-1,r} \times S^{DIS}_{t-1,x-1,r} + \quad (1)$$

$$\frac{1}{2}\left(P^{AR}_{t-1,x-1,r} \times I_{t-1,x-1,r} \times \frac{1}{4}\left(S^{DIS}_{t-1,x-1,r} + S^{DIS}_{t,x-1,r} + S^{DIS}_{t-1,x,r} + S^{DIS}_{t,x,r}\right)\right) + \frac{1}{2}\left(P^{AR}_{t-1,x-1,r} \times I_{t-1,x-1,r}\right) + \quad (2)$$

$$\frac{1}{2}\left\{\frac{1}{2}\left(EXMG_{t-1,t,x-1,r} \times INCNEW_{t-1,t,x-1,r}\right) \times \left(1 - Prop_F_EXMG_{t-1,t,x-1,r}\right) + \frac{1}{2}\left(EXMG_{t-1,t,x,r} \times INCNEW_{t-1,t,x,r}\right) \times \left(1 - Prop_F_EXMG_{t-1,t,x,r}\right)\right\} + \quad (3)$$

$$\frac{1}{2}\left\{IRMG^{DIS}_{t-1,t,x,r} \times \left(1 - Prop_F_IRMG_{t-1,t,x,r}\right) + IRMG^{DIS}_{t-1,t,x-1,r} \times \left(1 - Prop_F_IRMG_{t-1,t,x-1,r}\right) - \quad (4)$$

$$\frac{1}{2}\left\{IRML^{DIS}_{t-1,t,x,r} \times \left(1 - Prop_F_IRML_{t-1,t,x,r}\right) + IRML^{DIS}_{t-1,t,x-1,r} \times \left(1 - Prop_F_IRML_{t-1,t,x-1,r}\right)\right\} - \quad (5)$$

$$\frac{1}{2}\left\{EXML^{DIS}_{t-1,t,x,r} \times \left(1 - Prop_F_EXML_{t-1,t,x,r}\right) + EXML^{DIS}_{t-1,t,x-1,r} \times \left(1 - Prop_F_EXML_{t-1,t,x-1,r}\right)\right\} \quad (6)$$

The above equation for calculating the population (prevalence) with the disease can be broken into six parts:

(1) The number of people with the disease surviving from year t-1 to year t
(2) The number of people from the at-risk population developing the disease, and surviving, from June 30 of year t-1 to June 30 of year t. It is assumed that people develop the disease uniformly throughout the year (June 30 of year t-1 to June 30 of year t), and there is an equal chance someone aged x-1 may develop the disease at age x-1 versus age x. The survival rate is then averaged for ages x and x-1 in each of years t-1 and t. Furthermore, the rate is applied to only half of those developing the disease since it is assumed new cases are developed uniformly throughout the year.
(3) The number of people migrating into Alberta from July 1 of year t-1 to June 30 of year t, who develop the disease that year.
(4) The inter-regional migration gains of people with the disease.
(5) The inter-regional migration losses of people with the disease.
(6) The external migration losses of people with the disease.

The list below defines each element of the equation, including calculation details when needed:

- $P^{AR}_{t,x,r}$ is the population at risk of developing the disease, living in region r, aged x at June 30 of year t.

- $P^{DIS}_{t,x,r}$ is the population (prevalence) of people with the disease, living in region r, aged x at June 30 of year t.

- $S^{DIS}_{t-1,x-1,r}$ is the survivorship ratio for those with the disease. It is the probability that someone with the disease, living in region r, aged x-1 at June 30 of year t-1, will live for one more year. Calculation details are provided in appendix 3.

- $I_{t-1,x-1,r}$ is the incidence risk. It is the probability that someone living in region r, aged x-1 and at risk of developing the disease on June 30 of year t-1 will develop the disease from July 1 of year t-1 to June 30 of year t. It is calculated as

$$I_{t-1,x-1,r} = \frac{NEW_{t-1,t,x-1,r}}{P^{AR}_{t-1,x-1,r}},$$

  where $NEW_{t-1,t,x-1,r}$ is the number of people from $P^{AR}_{t-1,x-1,r}$ who develop the disease from July 1 of year t-1 to June 30 of year t.

- $EXMG_{t-1,t,x-1,r}$ is the number of individuals aged x-1, migrating into region r from outside of Alberta, between July 1 of year t-1 and June 30 of year t. Calculation details for external migration are shown in appendix 4.

- $INCNEW_{t-1,t,x,r}$ is the rate at which new migrants aged x develop the disease from July 1 of year t-1 to June 30 of year t. It is calculated as

$$INCNEW_{t-1,t,x,r} = \frac{NEWMIG_{t-1,t,x,r}}{\frac{1}{2} \times EXMG_{t\text{-}1,t,x,r}},$$

where $NEWMIG_{t-1,t,x,r}$ is the number of individuals in region r, developing the disease at age x, from July 1 of year t-1 to June 30 of year t, who were not in the at-risk population on June 30 of year t-1.

- $Prop_F_EXMG_{t-1,t,x,r}$ is the proportion of external migration gains into region r, aged x, between July 1 of year t-1 to June 30 of year t who are female.

- $IRMG^{DIS}_{t-1,t,x,r}$ is the number of individuals aged x, with the disease migrating into region r (from another region in Alberta) between July 1 of year t-1 and June 30 of year t. It is calculated as

$$IRMG^{DIS}_{t-1,t,x,r} = IRMG_{t-1,t,x,r} \times \frac{1}{2}\left\{\frac{P^{DIS}_{t-1,x,r^*}}{P^{AR}_{t-1,x,r^*} + P^{DIS}_{t-1,x,r^*}} + \frac{P^{DIS}_{t-1,x-1,r^*}}{P^{AR}_{t-1,x-1,r^*} + P^{DIS}_{t-1,x-1,r^*}}\right\},$$

where r^* refers to all eight regions outside of region r. The right side of the equation takes an estimate of the proportion of people in the province outside of region r that have the disease and multiplies it by the number of inter-regional migration gains.

$IRMG_{t-1,t,x,r}$ is the number of individuals aged x, migrating into region r from another region in Alberta, between July 1 of year t-1 to June 30 of year t. Calculation details for inter-regional migration are shown in appendix 4.

- $Prop_F_IRMG_{t-1,t,x,r}$ is the proportion of inter-regional migration gains into region r (came from another region of Alberta), aged x, between July 1 of year t-1 to June 30 of year t who are female.

- $IRML^{DIS}_{t-1,t,x,r}$ is the number of individuals aged x, with the disease moving out of region r (into another region of Alberta) between July 1 of year t-1 to June 30 of year t. It is calculated as

$$IRML^{DIS}_{t-1,t,x,r} = IRML_{t-1,t,x,r} \times \frac{1}{2}\left\{\frac{P^{DIS}_{t-1,x,r}}{P^{AR}_{t-1,x,r} + P^{DIS}_{t-1,x,r}} + \frac{P^{DIS}_{t-1,x-1,r}}{P^{AR}_{t-1,x-1,r} + P^{DIS}_{t-1,x-1,r}}\right\},$$

The right side of the equation takes an estimate of the proportion of people in region r that have the disease and multiplies it by the total number of inter-regional migration losses.

$IRML_{t-1,t,x,r}$ is the number of individuals aged x, migrating out of region r to another region in Alberta, between June 30 of year t-1 to June 29 of year t-1. Calculation details for inter-regional migration are shown in appendix 4.

- $Prop_F_IRML_{t-1,t,x,r}$ is the proportion of inter-regional migration losses from region r (into another region of Alberta), aged x, between July 1 of year t-1 to June 30 of year t who are female.

- Typically in practice, it is assumed that $Prop_F_IRML_{t-1,t,x,r} = Prop_F_IRMG_{t-1,t,x,r}$ for all regions of Alberta. In other words, $Prop_F_IRM_{t-1,t,x,r}$ is the proportion of females moving inter-regionally in Alberta at age x, is used for each region, for both gains and losses.

- $EXML^{DIS}_{t-1,t,x,r}$ is the number of individuals aged x with the disease moving out of region r to somewhere outside of Alberta, between July 1 of year t-1 to June 30 of year t. It is calculated as

$$EXML^{DIS}_{t-1,t,x,r} = EXML_{t-1,t,x,r} \times \frac{1}{2}\left\{\frac{P^{DIS}_{t-1,x,r}}{P^{AR}_{t-1,x,r} + P^{DIS}_{t-1,x,r}} + \frac{P^{DIS}_{t-1,x-1,r}}{P^{AR}_{t-1,x-1,r} + P^{DIS}_{t-1,x-1,r}}\right\},$$

  The right side of the equation takes an estimate of the proportion of people in region r that have the disease and multiplies it by the total number of external migration losses.

  $EXML_{t-1,t,x,r}$ is the number of individuals aged x, migrating out of region r to somewhere outside of Alberta, between July 1 of year t-1 to June 30 of year t. Calculation details for inter-regional migration are shown in appendix 4.

- $Prop_F_EXML_{t-1,t,x,r}$ is the proportion of external migration losses out of region r, aged x, between July 1 of year t-1 to June 30 of year t who are female.

- Typically in practice, it is assumed that $Prop_F_EXML_{t-1,t,x,r}$ and $Prop_F_EXMG_{t-1,t,x,r}$ are the same for all regions for each projection year. This assumption can be altered if required.

Projecting disease prevalence, open-ended age interval x=90+:

$$P^{DIS}_{t,90+,r} = (P^{DIS}_{t-1,89,r} + P^{DIS}_{t-1,90+,r}) \times S^{DIS}_{t-1,89+,r} + \qquad (1)$$

$$\frac{1}{2}(P^{AR}_{t-1,89,r} \times I_{t-1,89,r} \times \frac{1}{2}(S^{DIS}_{t-1,89+,r} + S^{DIS}_{t,89+r})) + \frac{1}{2}(P^{AR}_{t-1,89,r} \times I_{t-1,89,r}) +$$

$$\frac{1}{2}(P^{AR}_{t-1,90+,r} \times I_{t-1,90+,r} \times \frac{1}{2}(S^{DIS}_{t-1,89+,r} + S^{DIS}_{t,89+,r})) + \frac{1}{2}(P^{AR}_{t-1,90+,r} \times I_{t-1,90+,r}) + \qquad (2)$$

$$\frac{1}{2}\left\{\frac{1}{2}(EXMG_{t-1,t,89,r} \times INCNEW_{t-1,t,89,r}) \times (1 - Prop_F_EXMG_{t-1,t,89,r})\right\} +$$

$$\frac{1}{2}(EXMG_{t-1,t,90+,r} \times INCNEW_{t-1,t,90+,r}) \times (1 - Prop_F_EXMG_{t-1,t,90,r}) + \quad (3)$$

$$\frac{1}{2}\left\{ IRMG^{DIS}_{t-1,t,89,r} \times (1 - Prop_F_IRMG_{t-1,t,89,r}) \right\}$$
$$+ IRMG^{DIS}_{t-1,t,90+,r} \times (1 - Prop_F_IRMG_{t-1,t,90+,r}) - \quad (4)$$

$$\frac{1}{2}\left\{ IRML^{DIS}_{t-1,t,89,r} \times (1 - Prop_F_IRML_{t-1,t,89,r}) \right\}$$
$$+ IRML^{DIS}_{t-1,t,90+,r} \times (1 - Prop_F_IRML_{t-1,t,90+,r}) - \quad (5)$$

$$\frac{1}{2}\left\{ EXML^{DIS}_{t-1,t,89,r} \times (1 - Prop_F_EXML_{t-1,t,89,r}) \right\}$$
$$+ EXML^{DIS}_{t-1,t,90+,r} \times (1 - Prop_F_EXML_{t-1,t,90+,r}) \quad (6)$$

The elements of the equation for the 90+ age group are the same as those defined above.

Projecting at-risk population, ages x= 1 to 89:

$$P^{AR}_{t,x,r} = P^{AR}_{t-1,x-1,r} \times (1 - I_{t-1,x-1,r}) \times S^{AR}_{t-1,x-1,r} + \quad (1)$$

$$\frac{1}{2}\left\{ (EXMG_{t-1,t,x-1,r} \times (1 - \frac{1}{2} \times INCNEW_{t-1,t,x-1,r})) \times (1 - Prop_F_EXMG_{t-1,t,x-1,r}) + \right.$$

$$\left. \frac{1}{2}(EXMG_{t-1,t,x,r} \times (1 - \frac{1}{2} \times INCNEW_{t-1,t,x,r}) \times (1 - Prop_F_EXMG_{t-1,t,x,r}) \right\} + \quad (2)$$

$$\frac{1}{2}\left\{ IRMG^{AR}_{t-1,t,x,r} \times (1 - Prop_F_IRMG_{t-1,t,x,r}) + \right.$$
$$\left. IRMG^{AR}_{t-1,t,x-1,r} \times (1 - Prop_F_IRMG_{t-1,t,x-1,r}) \right\} - \quad (3)$$

$$\frac{1}{2}\left\{ IRML^{AR}_{t-1,t,x,r} \times (1 - Prop_F_IRML_{t-1,t,x,r}) + \right.$$
$$\left. IRML^{AR}_{t-1,t,x-1,r} \times (1 - Prop_F_IRML_{t-1,t,x-1,r}) \right\} - \quad (4)$$

$$\frac{1}{2}\left\{ EXML^{AR}_{t-1,t,x,r} \times (1 - Prop_F_EXML_{t-1,t,x,r}) + \right.$$
$$\left. EXML^{AR}_{t-1,t,x-1,r} \times (1 - Prop_F_EXML_{t-1,t,x-1,r}) \right\} \quad (5)$$

The above equation for calculating the population (prevalence) with the disease can be broken into five parts:

(1) The number of people at risk of developing the disease on June 30 of year t-1, who do not develop the disease from July 1 of year t-1 to June 30 of year t, and survive to June 30 of year t.
(2) The number of people migrating into Alberta from June 30 of year t-1 to June 30 of year t, who do not develop the disease that year.
(3) The inter-regional migration gains of people who are at risk of developing the disease.
(4) The inter-regional migration losses of people who are at risk of developing the disease.
(5) The external migration losses of people who are at risk of developing the disease.

The elements of the above equation for projecting at-risk population that have not already been previously defined above are:

- $S^{AR}_{t-1,x-1,r}$ is the survivorship ratio for those at risk of developing the disease. It is the probability that someone at risk of developing the disease, living in region r, aged x-1 at June 30 of year t-1, will live for one more year. Calculation details are provided in appendix 3.

- $IRMG^{AR}_{t-1,t,x,r}$ is the number of individuals aged x, at risk of developing the migrating into region r (from another region of Alberta) between July 1 of year t-1 to June 30 of year t. It is calculated as

$$IRMG^{AR}_{t-1,t,x,r} = IRMG_{t-1,t,x,r} \times \frac{1}{2}\left\{\frac{P^{AR}_{t-1,x,r^*}}{P^{AR}_{t-1,x,r^*} + P^{DIS}_{t-1,x,r^*}} + \frac{P^{AR}_{t-1,x-1,r^*}}{P^{AR}_{t-1,x-1,r^*} + P^{DIS}_{t-1,x-1,r^*}}\right\},$$

  where r^* refers to all eight regions outside of region r. The right side of the equation takes an estimate of the proportion of people in the province outside of region r that are at risk of developing the disease and multiplies it by the number of inter-regional migration gains.

- $IRML^{AR}_{t-1,t,x,r}$ is the number of individuals aged x, at risk of developing the disease migrating into region r (from another region of Alberta) between July 1 of year t-1 to June 30 of year t. It is calculated as

$$IRML^{AR}_{t-1,t,x,r} = IRML_{t-1,t,x,r} \times \frac{1}{2}\left\{\frac{P^{AR}_{t-1,x,r}}{P^{AR}_{t-1,x,r} + P^{DIS}_{t-1,x,r}} + \frac{P^{AR}_{t-1,x-1,r}}{P^{AR}_{t-1,x-1,r} + P^{DIS}_{t-1,x-1,r}}\right\},$$

  The right side of the equation takes an estimate of the proportion of people in region r that are at risk of developing the disease and multiplies it by the total number of inter-regional migration losses.

- $EXML^{AR}_{t-1,t,x,r}$ is the number of individuals aged x, at risk of developing the disease moving out of region r to somewhere outside of Alberta, between July 1 of year t-1 to June 30 of year t. It is calculated as

$$EXML^{AR}_{t-1,t,x,r} = EXML_{t-1,t,x,r} \times \frac{1}{2}\left\{\frac{P^{AR}_{t-1,x,r}}{P^{AR}_{t-1,x,r} + P^{DIS}_{t-1,x,r}} + \frac{P^{AR}_{t-1,x-1,r}}{P^{AR}_{t-1,x-1,r} + P^{DIS}_{t-1,x-1,r}}\right\},$$

The right side of the equation takes an estimate of the proportion of people in region r that are at risk of developing the disease and multiplies it by the total number of external migration losses.

The second part of the equation (2) is derived using the following logic:

Since $INCNEW_{t-1,t,x,r} = \frac{NEWMIG_{t-1,t,x,r}}{\frac{1}{2} \times EXMG_{t-1,t,x,r}}$,

(i.e. $NEWMIG_{t-1,t,x,r} = \frac{1}{2}(EXMG_{t-1,t,x,r} \times INCNEW_{t-1,t,x,r})$

And since $EXMG_{t-1,t,x,r} = NEWMIG_{t-1,t,x,r} + NOT_NEWMIG_{t-1,t,x,r}$

($EXMG_{t-1,t,x,r}$ can be broken into those migrants that do develop the disease during the year and those who do not)

then $NOT_NEWMIG_{t-1,t,x,r} = EXMG_{t-1,t,x,r} - NEWMIG_{t-1,t,x,r}$

Therefore $NOT_NEWMIG_{t-1,t,x,r} = EXMG_{t-1,t,x,r} \times (1 - 0.5 \times INCNEW_{t-1,t,x,r})$

where $NOT_NEWMIG_{t-1,t,x,r}$ is the number of individuals in region r, who do not develop the disease from July 1 of year t-1 to June 30 of year t, and were not in the at-risk population on June 30 of year t-1 (i.e. they migrated in at age x).

Glossary of notation

I – Incidence risk, defined as the probability that someone at risk in Alberta as of mid-year, will develop the disease in the upcoming year.

P – Population in Alberta. The superscript AR and DIS are added if referring to the at-risk or disease prevalent populations.

NEWCASE – Number of at-risk people in Alberta as of mid-year developing the disease in the upcoming year.

NEWMIG – Number of new migrants, who were not in Alberta as of mid-year, and develop the disease in the upcoming year.

INCNEW – Incidence rate of new migrants, defined as the rate at which a new migrant into Alberta develops a disease.

EXMG – External migration gains, defined as the number of people migrating into Alberta from outside the province during a given year. The superscript AR and DIS are added if referring to the at-risk or disease prevalent populations.

EXML – External migration losses, defined as the number of people migrating out of Alberta during a given year. The superscript AR and DIS are added if referring to the at-risk or disease prevalent populations.

IRMG – Inter-regional migration gains, defined as the number of people migrating into one Alberta health region from another Alberta health region. The superscript AR and DIS are added if referring to the at-risk or disease prevalent populations.

IRML – Inter-regional migration losses, defined as the number of people migrating out of one Alberta health region to another Alberta health region. The superscript AR and DIS are added if referring to the at-risk or disease prevalent populations.

ASDR – Age-specific death rate, defined as the rate at which persons die. The superscript AR and DIS are added if referring to the at-risk or disease prevalent populations.

D – Number of deaths in a given calendar year. The superscript AR and DIS are added if referring to the at-risk or disease prevalent populations.

Prop_F_IRML – Proportion of inter-regional migration losses that are female. The same notation is applied for inter-regional migration gains, and external migration gains and losses.

S – Survivorship ratio, defined as the probability of surviving one more year. The superscript AR and DIS are added if referring to the at-risk or disease prevalent populations.