A Global Illness-Death Model

Andrew Dolgert adolgert@uw.edu

August 2, 2019

Abstract: to fellow authors

This is a paper about what EpiViz-AT does, not what Dismod-AT does. We don't care about Dismod-AT internals, just its likelihood structure. This will be the first presentation about how the hierarchy works.

Here's the plan. This is a draft, so introduction and conclusions will remain bullet points until a later date. Start with figures and equations, and fill in. Put in too many references early to ensure all readers have access to related articles, and trim later. Hand-drawn figures will be replaced later. Focus on crisp concepts and clear notation.

1 Introduction

State the problem:

- Disease burden of population by age over time
- In multiple countries or multiple regions within a country
- Missing data

Make an argument:

- Respect the mechanism of disease. This is the history of individuals who live in a place. This is Dr. Murray's exhortation to place work within context of individual.
- Use all of the data we know. This is meta-regression.
- Use similarity among countries and within countries to deal with missing data, but don't use spatial methods. Identify sets of similar countries according to Global health criteria.

Small-multiples of two countries with same disease, for two rates, one derived quantity. With measurements somehow. This should show why the mechanism of disease matters. Show how a shift in incidence over time leads to a delayed shift in prevalence.

1. Introduction: No math here. Why & argument.
2. Previous Work: Moving this later can make it a more useful section, but second is the traditional spot.
3. Input and Output: One of the best ways to simplify what this code does is to describe plainly what goes in and what goes out. This is still for the larger community.
4. Two-level model: This is Dismod-AT, but we don't care about its technical marvel. This is a birds-eye view that tells you what it solves, not how it solves.

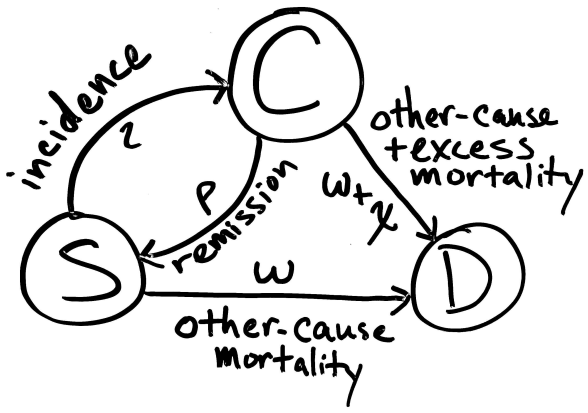


Figure 1: The illness-death model examines the effect of a single disease given background mortality from other conditions. It has three states, susceptible, with-condition, and dead. For the analysis of population health, we treat the rates among these states as depending on time since birth.

5. Global model: What is the "posteriors-to-priors" move that we do, and what does that mean for the results?

6. Examples: Let's see it in action. Put all 195 countries in the appendix. Sensitivity analysis is the hardest part here, but it's very important for making this trustworthy.

7. Conclusions: Be clear about limitations in this section. People will skip from the introduction to here.

2 Previous Work

The illness-death model has been known since before 1900 [42]. Fix and Neyman stated it as a time-inhomogeneous Markov model [18]. There are multiple approaches for estimation with this model. In actuarial science, it is sometimes called the Disability model [14]. For reliability, it is called a repair model [30].

You can use survival analysis. Example in Andersen and Keiding 2002? You can use GLMs with approximate approaches to the nonlinear disease model [42]. You can go full Semi-Markov on the individual level.

As Hoem 1972 describes, there is also a choice about how parametric to be, and we don't want to miss behaviors that show change over time, so we want non-parametric [21].

We could use discrete-time at the lowest level of the model, but that raises more questions about how to deal with incommensurate data. The audience are doctors who understand that disease is a continuous process within a body and within a population.

The use of a time-inhomogeneous ordinary differential equation for disease was advocated by both Keyfitz and Rogers [27] and Jones [25]. It is Jones's piecewise-constant integration that is used by a series of implementations, all known under the name Dismod.

Dismod ii [5] looked at data within a single year, across all ages. There is a book with examples of use of Dismod ii, but it doesn't describe the hierarchical model or likelihood structure [19]. There is an earlier Dismod-AT paper, paper that describes use of the differential equation along cohort time [6]. That implementation is strengthened by automatic differentiation, as described below. The use of AD makes optimization faster but rules out solution by other methods such as decomposition of the operator, flowgraphs [23], or integration by uniformization [45].

3 Input and Output

(Modelers should write this part. Audience is global health researcher deciding whether to use this tool.)

3.1 Input

The input data comes from multiple populations, which are usually multiple countries but can be, for instance, districts within one or more countries. The input is population measurements, usually from surveys, so they are events per exposure. These inputs to the model can be any one of incidence, remission, prevalence, excess mortality, other-cause mortality, cause-specific mortality, all-cause mortality, standardized mortality ratio, or relative risk. The model needs to know, for each record in the data, over what age and time range the value applies, over what population, and an estimation of its error. Specification of priors will be discussed below when describing the Bayesian calculation.

One important distinction is that incidence of disease is usually measured by surveys as incidence among the total population, while hospital data is more likely to report incidence among those susceptible, excluding from the denominator those who already have the disease. These are treated separately by Dismod-AT as total-incidence and susceptible-incidence.

It's possible to pass into the model direct values for hazard rates. These would have to be hazard rates as a function of age. For instance, a hazard rate for remission, as a function of time since diagnosis, would be inappropriate. However, almost all available data is demographic data, and the model is built to handle this.

Disease data is generally less well-known than mortality data, which is often available for all age-groups over all years. The implementation of the global illness-death model assumes the availability of all-cause mortality rates for every age group and year. Because the illness-death compartmental model, in the variant used here, does not include fertility rates, it is more appropriate to use the lifetable mortality rate, ${}_n m_x$, not the crude mortality rate, ${}_n M_x$. It also does not include shocks, meaning accidents and disasters that cause jumps in disease state or living population. Cause-specific mortality is sometimes included when available data sources are judged to be informative by modelers.

Say something about covariates. Include information on patterns of missingness.

3.2 Output

The statistical model's main outputs are the hazard rates, as a continuous value over all ages and times in the model, for incidence, remission, excess mortality, and other-cause mortality. It also produces prevalence and covariate multipliers as continuous values over age and time.

From these continuous values, the model derives any of the common population measurements. Population measurements apply to subpopulations, a term we use here to denote all individuals between a minimum and maximum age from a starting to a finishing time. A population measurement is count of event per exposure for a subpopulation. For instance, the population mortality rate,

$$\frac{\text{death events}}{\text{person-years in age and time range}} \tag{1}$$

is a function of the mortality hazard rate, $\mu(a,t)$. If $\mu(a,t)$ is known, then the population mortality rate is found by integration over age and time, equivalent to Eq. 1,

$${}_{a_f - a_i} m_{a_i}(t_i, t_f) = \frac{\int_{t_i}^{t_f} \int_{a_i}^{a_f} \mu(a,t)\, e^{-\int_0^x \mu(x,t)dx} da\, dt}{\int_{t_i}^{t_f} \int_{a_i}^{a_f} e^{-\int_0^x \mu(x,t)dx} da\, dt}. \tag{2}$$

For a subpopulation, the mortality rate, incidence rate, standardized mortality ratio, and relative risk, are all found by integrating the model's continuous main outputs.

3.3 Interpolation and Graduation

While the statistical model will evaluate goodness-of-fit against demographic data on subpopulations, it runs faster when input data is more similar to its output values. Faster runs are especially helpful for exploratory work with the model.

One of the output values of the model is other-cause mortality, while all-cause mortality and cause-specific mortality are inputs. Cause deletion algorithms create other-cause mortality estimates [4, 3, 7, 8, 9, 10, 34, 37]. And these are those algorithms that take demographic mortality rates and return point estimates of mortality rates or continuous hazard rates for mortality. XXX Go through refs and make decisions.

The model outputs are continuous functions. If the inputs are defined as point estimates, meaning for a single age and time instead of a range of ages and times, then this also speeds comparison with the outputs. Graduation methods can make point estimates. Examples and XXX Go through refs and make decisions [7, 8, 11, 16, 17, 20, 21, 22, 28, 32, 35, 36, 40, 41].

4 Two-level Hierarchical Model

Dismod-AT is a two-level hierarchical estimation [1]. The macro-level is a population, and the micro-level are measurements of the disease state of that population. The global illness-death model, described in Sec. 5, uses Dismod-AT at each of multiple levels of a deep location hierarchy. This section describes those capabilities of Dismod-AT that are used by the global illness-death model. It is not a complete description of its capabilities or methods. A paper about an earlier version of Dismod-AT covers many implementation details in greater depth [6]. The goal here is to explain the structure of its Bayesian model in a compact form, beginning with a description of the differential equation that Dismod-AT solves in its process model.

4.1 Illness-death as a differential equation

As shown in Fig. 1, the illness-death model is a compartmental model with three states and four transitions [27]. The S state represents those alive who may have any disease but the one under study. Any members of the population not susceptible to the disease should be excluded from S. The C state represents those who have the disease. The D state is death. The four transitions are incidence of disease, remission of disease, death from S, and death from C. The last transition we mark as $\omega+\chi$, meaning its rate is always assumed to be at least as much as other-cause mortality, ω, augmented by excess mortality, χ.

To analyze individual-level data or small populations, it would be appropriate to use a stochastic representation of the compartmental model. Keyfitz and Rogers describe a time-inhomogeneous Markov model for illness and death, deriving a Kolmogorov forward equation to describe the change in state occupancy over time for an individual,

$$\frac{dP(x)}{dx} = -\lambda(x)P(x), \tag{3}$$

where x is time since birth, $\lambda(x)$ is a matrix of hazard rates, as a function of age, and $P(x)$ is a vector of probability an individual is in a compartmental state [27].

That same equation, in its large-population, or continuum, limit, has been well-known since before 1900 as a time-inhomogeneous ordinary differential equation for prevalence of disease within a population,

$$\begin{bmatrix} \frac{dS(x)}{dx} \\ \frac{dC(x)}{dx} \\ \frac{dD(x)}{dx} \end{bmatrix} = -\begin{bmatrix} \iota(x)+\omega(x) & -\rho(x) & 0 \\ -\iota(x) & \rho(x)+\omega(x)+\chi(x) & 0 \\ \omega(x) & \omega(x)+\chi(x) & 0 \end{bmatrix} \begin{bmatrix} S(x) \\ C(x) \\ D(x) \end{bmatrix}. \tag{4}$$

Rates, $\lambda_x = (\iota(x), \rho(x), \omega(x), \chi(x))$, are averages over a population, defined continuously across all ages. The combination of an initial prevalence and all rates, defined for all x, determines a unique solution for prevalence over x. This closed population represents a cohort, born at the same time, within a larger population.

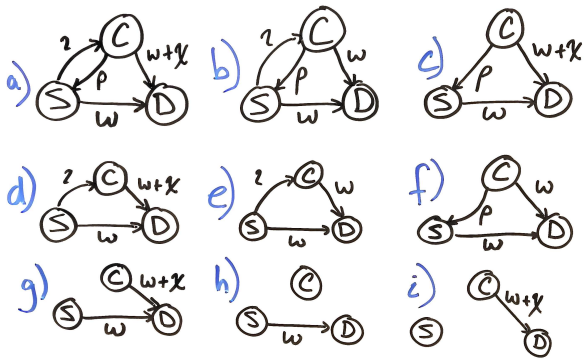
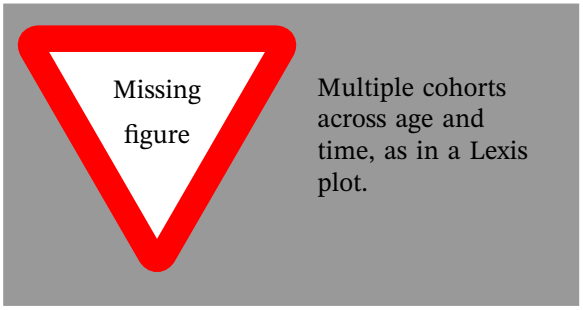


Figure 2: Reduced forms of the illness-death model apply to different diseases. a) Illness-death model b) without excess mortality for disability-only disease c) without incidence for congenital disease d) without remission e) no remission, but only disability f) no incidence, only disability g) condition at birth, as a competing cause of death. h) and i) are the model g) with different initial prevalence completely in S or completely in C. They are both called the two-state model. Case i) can be used to graduate cause-specific mortality over age and time.



The larger population can be thought of as having multiple cohorts, each with its own birth time, t_0, and cohort age, x. While it is possible to derive a partial differential equation for the illness-death model across age $a = x + t_0$ and time $t = x + t_0$, that differential equation is separable when rotated back to cohort time [24], so solving each cohort individually is exactly equivalent. This is not a model for communicable disease, so there is no interaction among cohorts. Neighboring cohorts will be similar because input demographic data spans neighboring cohorts, because all incidence rates are continuous, and because the statistical model will introduce regularization terms across changes in rates by age and by time.

The structure of the differential equation determines the structure of the Dismod-AT process model. The differential equation's inputs are $\lambda_x(t) = (\iota(a,t), \rho(a,t), \omega(a,t), \chi(a,t))$ and prevalence at birth, $l_0(t)$. Its output is prevalence in each state over all ages and times, $l_x(\lambda_x(t), l_0(t), t)$. Therefore, the statistical model must parameterize multiple fields over age and time.

A Markov random field will represent each continuous field over age and time. Its parameters are the value of the field at discrete points on a rectangular lattice in age and time. The continuous field is a bivariate interpolation among the points on the rectangular lattice. Labeling grid points by (i, j), each continuous field will have three priors,

$$v_{i,j} \sim N(\mu, \sigma) \tag{5}$$

$$v_{i+1,j} - v_{i,j} \sim N(\mu, \sigma) \tag{6}$$

$$v_{i,j+1} - v_{i,j} \sim N(\mu, \sigma), \tag{7}$$

which regularize the grid by penalizing differences in age and time. Outside of the support of the (i, j), the value is taken as the nearest value within the support.

Dismod-AT, can calculate on a wide variety of compartmental models created by setting any of

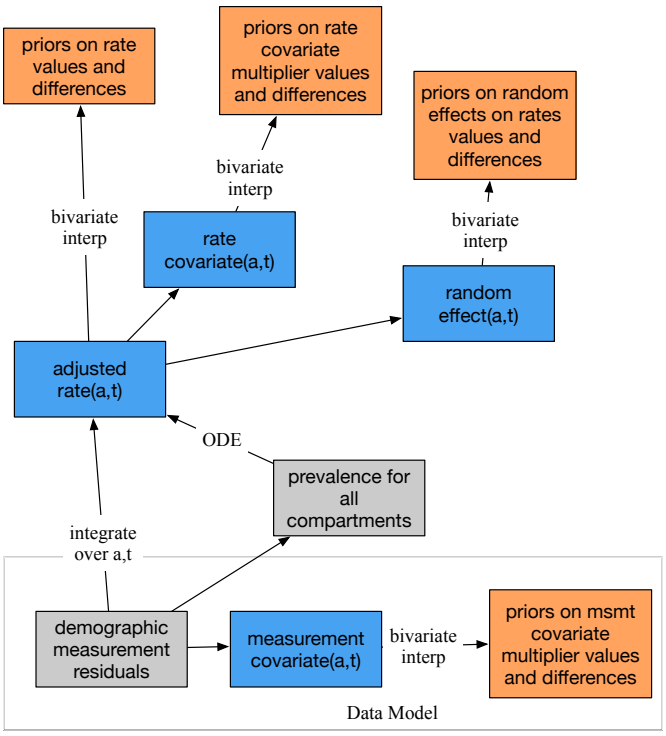


Figure 3: In blue are continuous functions defined by grids of priors. In orange are the priors. Each of the orange boxes also has hyper-priors, η, which multiply standard deviation. The Data Model is clear. How do we represent the hierarchical model and process model? Where do we split this fruitfully?

the transition rates to zero, as shown in Fig. 2. Different variants are useful for different diseases. For instance, congenital defects have an initial prevalence but no incidence. Some diseases have no remission and others no excess mortality. Which topological variant is appropriate for a particular disease is a matter of model selection.

4.2 Hierarchical Bayesian Model

The joint likelihood of a Dismod-AT model can be written as three parts of a hierarchical Bayesian model.

- Data model for demographic data, Z, $P[Z|Y,\Theta]$. Given a known set of rates and prevalence of disease in all countries, what are the expected errors of measurements such as cause-specific mortality rate and and excess mortality.
- The process model isn't a probability. It's the output of the illness-death equation, given its inputs, $y = f(\theta, u)$. Given underlying rates for a disease and how both covariates and population-specific effects modify that rate, what is the hazard rate and prevalence of disease in this population.
- Random effects model, $P[U|\Phi]$, where u_j are the random effects on the value of the rates.
- Parameter model $P[\Theta|\Phi]$. The parameters are the value of each field over age and time to determine input rates and covariates.
- A hyper-parameter model, $P[\Phi]$, which controls expectation about standard deviation of the parameter model.

Each prior distribution can be any one of a number of uniform, Gaussian, Laplace, Student's-t, log-Gaussian, log-Laplace, or log-Student's-t. As such, we denote all distributions by dist(μ, σ) where μ is the mean and σ the standard deviation, omitting other parameters. These distributions can all be constrained within minimum and maximum values, as well.

4.2.1 Hyper-priors

Each set of random variables to represent a surface in age and time has three hyper-priors, ϕ_i, which represent beliefs about prior knowledge of the standard deviation of each control point, so a control point will have a distribution $\text{dist}(\mu, \phi_i\sigma)$.

4.2.2 Parameters Model

The parameters, Θ, are the mean, standard deviation, minimum values and maximum values of the prior distributions on control points for each of

- q_k, the four underlying rates, which are hazard rates without random effects,
- l_0, prevalence at birth,
- α, β, γ, covariate multipliers on rate values, observation values, and standard deviations.

These all have the same prior structure, which defines the continuous function as a bivariate interpolation among a rectangular grid of control points in age and time.

Each continuous function has three hyper-hyper-priors which will multiply the standard deviation of the hyper-priors, (Φ_v, Φ_a, Φ_t). The hyper-priors specify the value of every control point for all continuous functions of age and type. The unnormalized likelihood is

$$P[\Theta|\Phi] \propto \left(\prod_v V_v\right)\left(\prod_a A_a\right)\left(\prod_t T_t\right) \tag{8}$$

where V_v are value priors at each control point, A_a are priors on value differences between neighboring control points in the age direction, and T_t are priors on value differences between neighboring control points in the time direction. This is normalized by dividing it by its integral over all values θ. Each of those value, age-difference, and time-difference priors has the form $\theta \sim \text{dist}(\mu_i, \phi_k\sigma_i)$, where the variables Φ are the mean and variance of the θ, and ϕ_k is the hyperprior.

4.2.3 Random Effects Model

There is a random effect on each rate for each location. Random effects, U, are an example of a Markov random field defining a continuous quantity. There is a grid of control points, each of which has a distribution, $\text{dist}(\mu, \sigma)$, where μ and σ are the parameters, and the standard deviation has the hyper-parameter, ϕ.

4.2.4 Process Model

The process model is the main outcome, Y, which is the hazard rate and the prevalence over all ages and times, expressed as $f(\theta, u)$. The rate for each location comes from the underlying rate for all locations, modified by a covariate multiplier and a random effect,

$$r_k = q_k e^{u_j + \sum_d \alpha_d x_d}. \tag{9}$$

The differential equation calculates prevalence as is then $y = f(r_k)$.

4.2.5 Data Model

The data model $P[Z|Y, \theta]$, is the computation of functionals from rates and prevalence. This happens in two steps. The first calculates a function of prevalence and rates, $h(y(a,t))$. For instance, all-cause mortality is $\omega(a,t) + \chi(a,t)C(a,t)/(S(a,t)+C(a,t))$. The second integrates this over a subpopulation, weighted according to exposure, $w(a,t)$. There is a covariate adjustment, which Dismod-AT calls an covariate multiplier on the measured value, within the integral,

$$g(y(a,t), \beta(a,t)) = \frac{1}{\bar{w}} \int\int w(a,t)h(y(a,t))e^{\sum \beta x} da dt \tag{10}$$

where $\bar{w}$ is normalization on the exposure. For linear distributions,

$$z_i \sim \text{dist}(g_i(l_x(r_k, l_0), r_k, l_0), \sigma + \sum_j \gamma_j x_{ij}) \tag{11}$$

while for log-distributions,

$$z_i \sim \text{dist}(g_i(l_x(r_k, l_0), r_k, l_0), \sigma(1 + \sum_j \gamma_j x_{ij})). \tag{12}$$

The overall likelihood is these multiplied, $Z = \prod_i Z_i$.

4.3 Method of Solution

Deterministic solution called maximum *a posteriori* (MAP) estimation. Mode of the likelihood but no information about the distribution. We get a $\hat{y}$ but not $P[Y|\Theta, \phi, \eta]$.

Integration assumes constant hazard rates for each step [25]. Integrate with automatic differentiation, which means we know the Jacobian, $dy_i/d\theta$, for all measurements and parameters. Then pass this to an optimizer, such as ipopt, snopt.

Comparison with measurements defined over ranges of age and time requires multiple shooting along the characteristic, which is the cohort's age-time line, in order to integrate the functional over the underlying rate.

Slud and Suntornchost cover a broad range of methods of numerical solution [44].

4.4 Estimate of Posteriors

In order to construct a hierarchy of countries, the two-level model should estimate its posterior, $p(\theta|z, \phi)$, but it calculates only the MAP estimate. Skaug and Fournier describe this [43].

The method is to overestimate the uncertainty and simulate draws. We know that, because priors carry information, they always reduce uncertainty in the measured data. Therefore, the method is to:

1. Given measurements, fit to find primary rates, covariates, and random effects, $\hat{y}$.
2. Generate a set of input measurements, $\hat{z}$ from those rates, to match the original set of measurements.
3. Apply input uncertainty for the original measurement values, $z_i - \hat{z} \sim N(0, \sigma(z))$, to the output measurements at the new values. This is an overestimate.
4. Sample from the predicted measurements with the uncertainty of the actual measurements.
5. Fit the problem again to each of the z_i to get a new θ_i.
6. Report the θ_i, from which we can construct estimators of $p(\theta|z)$, uncertainty in the MAP estimate.

The global model will use these posteriors in order to create priors for populations that have insufficient data. Laird calculates univariate nonparametric maximum likelihood [29]. It doesn't look like the simulate command in Dismod-AT captures uncertainty in the priors.

There are challenges to doing this correctly. Some papers discuss how to get the posterior distribution from MAP output [39, 38]. If there are 1500 model variables, which isn't uncommon given that every grid point is three parameters, what are the chances that the posterior distribution for one of those model variables isn't shrunken down to a constraint on the system? Laird looks at repeated samples of univariate data, not multivariate data.

5 Global Model

- The two-level model isn't enough.
- This is spatial in the sense that there is a graph of which countries are related. Clayton and Kaldor use EB for relative risk mapping among districts [15]. This article is a bridge between the EB and spatial methods. Breslow summarizes Empirical Bayesian for spatial mapping [12].
- Two concerns to address: spatial relatedness and how to choose priors for a model that represents an age-time field as splines. It's parametric in that it uses distributions to describe each point but nonparametric in that these distributions describe a spline of hazard rates.
- Construction of a hierarchy from the two-level model. Use Empirical Bayes method [13, 26]. (Everything says to read the Carlin and Louis 2010 book.) Solve for the posterior $\hat{\theta}$ that maximizes $P[\Theta|Z]$ at higher level and use this as priors on rates at the next level down.
- Previous work on using MAP output for empirical Bayes doesn't seem to cover this case of splines. They cover the case of using splines to describe the distribution of a prior, not to describe a field of priors. This is hierarchical, nonparametric, spline-based regression [33], so it uses a nonparametric estimator [29]. Pereyra asks how to get confidence regions in high dimensional spaces [38].
- Aaron says that INLA estimates the posterior distribution using the information matrix, which comes from the Hessian. It seems like Dismod-AT should have that. Can we use that instead of making draws?
- Do we reserve some of the parameters as pure priors? For instance, what if we want to insist that excess mortality rate be below some number for the whole hierarchy? That's done by splitting the priors into some that do and don't get modeled with EB [31]. Don't forget the hyperprior parameters, Φ.
- Need to be very careful about uncertainty of parameters. Must include multiple sources of uncertainty. What about correlation in the parameters? Brad recommends looking at simpler models, such as the multivariate Gaussian / Gaussian model in order to understand this better. For instance, the effect of double-counting data is that variance has $\sqrt{2}$ that you have to take out. I don't know that we're doing this now in Dismod-ODE.
- Let's say we solve at the region level, so there are many regions within a super-region. All measurements are associated with the country or subnational where the measurement was taken. The micro-level for Dismod-AT is the measurement. It the macro-level the region or is it the country or subnational which is within the given region? Are there seven macro-level populations or seven hundred?
- It would make sense to group countries differently for different diseases. Priors act most powerfully at the leaves of the tree.

6 Examples

6.1 Disease description

6.2 Solution

6.3 Sensitivity analysis

Including sensitivity analysis to ensure it's stable and insensitive to choices of priors, as long as those priors follow rules for being uninformative.

Sensitivity to grid, integration points, priors.

Put some example graphs here, complete ones in appendix. One or two diseases. All countries in appendix.

7 Conclusions

- Parsimony versus fidelity. Yes, lots of priors, but they define fields for a model with very few moving parts.
- Estimates burden using data, without putting finger on the scale.
- This is a tool that is used for a statistical investigation of data. Run on one location, then a drill, then a region, then global. Analyze residuals, as for any hierarchical model.
- Clear statement of what is borrowed strength. Real problem with this because data at one time fights regularization in age and time.
- Notice that this isn't semi-Markov. This means remission and excess mortality can behave in different ways for short-duration diseases. Discuss short-duration diseases.
- Some countries have no data. They borrow from neighboring countries in the most-detailed hierarchical model. This makes grouping of countries a strong prior for model selection.
- This model makes clear some classes of disease. They can have different transitions turned off. They can have different initial prevalence.
- Software available [2].

7.1 Limitations

- For a GLM, you might take time with regularization to ensure fitting. This regularizes over age and time and at least 195 countries, so it's hard to tune.
- For what diseases is this model unnecessary or inappropriate?
- Non-parametric, hierarchical, nonlinear, heteroscedastic, with hyper-hyper parameters.
- The weight function should be calculated from prevalence and rates, not given as an input. This gets the exposure wrong for every subpopulation measurement.
- Can include shocks in time-inhomogeneous Markov, but not if you are using an ODE formulation of it. Means we have to use mortality rate without shocks.
- No adaptive steps for integration, or adaptive definition of control points.
- Significant amount of computing resources.
- Assumption of a lot of input data for background mortality, cause-specific mortality.

7.2 Future steps

References

[1] Disease rates as functions of age and time. https://bradbell.github.io/dismod_at/doc/dismod_at.htm. Accessed: 2018-11-18.

[2] Dismod cascade. https://cascade.readthedocs.io/. Accessed: 2018-11-18.

[3] P. K. Andersen. Decomposition of number of life years lost according to causes of death. *Statistics in Medicine*, 32(30):5278–5285, 2013.

[4] Per Kragh Andersen, Vladimir Canudas-Romo, and Niels Keiding. Cause-specific measures of life years lost. *Demographic Research*, 29(December):1127–1152, 2013.

[5] Jan J Barendregt, Gerrit J Van Oortmarssen, Theo Vos, and Christopher JL Murray. A generic model for the assessment of disease epidemiology: the computational basis of dismod ii. *Population health metrics*, 1(1):4, 2003.

[6] Bradley M Bell and Abraham D Flaxman. A statistical model and estimation of disease rates as functions of age and time. *SIAM Journal on Scientific Computing*, 35(2):B511–B528, 2013.

[7] Hiram Beltrán-Sánchez and Samuel H. Preston. A New Method for Attributing Changes in Life Expectancy to Various Causes of Death , with Application to the United States with Application to the United States. *Population Studies Center*, (May, 4):1–26, 2007.

[8] Hiram Beltrán-sánchez, Samuel H Preston, and Vladimir Canudas-romo. An integrated approach to cause-of-death analysis: cause-deleted life tables and decompositions of life expectancy. *Demography Research*, 19(19):1323, 2008.

[9] Hiram Beltran-Sanchez and Samir Soneji. A unifying framework for assessing changes in life expectancy associated with changes in mortality: The case of violent deaths. *Theoretical Population Biology*, 80(1):38–48, 2011.

[10] Marie Pier Bergeron-Boucher, Marcus Ebeling, and Vladimir Canudas-Romo. Decomposing changes in life expectancy: Compression versus shifting mortality. *Demographic Research*, 33(1):391–424, 2015.

[11] Heather Booth and Leonie Tickle. Mortality modelling and forecasting: A review of methods. *Annals of Actuarial Science*, 3(1-2):3–43, 2008.

[12] Norman Breslow. Biostatistics and Bayes. *Statistical Science*, 5(3):269–284, 1990.

[13] Bradley P Carlin and Thomas A Louis. Empirical bayes: Past, present and future. *Journal of the American Statistical Association*, 95(452):1286–1289, 2000.

[14] Marcus C. Christiansen. Multistate models in health insurance. *AStA Advances in Statistical Analysis*, 96(2):155–186, 2012.

[15] David Clayton and John Kaldor. Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping. *Biometrics*, 43(3):671–681, 1987.

[16] J B Copas and S Haberman. NON-PARAMETRIC GRADUATION USING KERNEL METHODS. *Journal of the Institute of Actuaries*, 110(1):135–156, 1983.

[17] Paul H C Ellers and Brian D Marx. Flexible Smoothing with B-splines and Penalties. *Statistical Science*, 11(2):89–102, 1996.

[18] Evelyn Fix and Jerzy Neyman. A simple stochastic model of recovery, relapse, death and loss of patients. *Human Biology*, 23(3):205–241, 1951.

[19] Abraham D Flaxman, Dr Theo Vos, and Christopher JL Murray. *An integrative metaregression framework for descriptive epidemiology*. University of Washington Press, 2015.

[20] Steven Haberman and Arthur E Renshaw. Generalized Linear Models and Actuarial Science. *Journal of the Royal Statistical Society. Series D*, 45(4):407–436, 1996.

[21] Jan M Hoem. The Statistical Theory of Demographic Rates : A Review of Current Developments. *Scandinavian Journal of Statistics*, 3(4):169–185, 1976.

[22] John J Hsieh. Construction of Expanded Continuous Life Tables- A Generalization of Abridged and Complete Life Tables. *Mathematical Biosciences*, 103:287–302, 1991.

[23] Aparna V Huzurbazar. Modeling and analysis of engineering systems data using flowgraph models. *Technometrics*, 42(3):300–306, 2000.

[24] Mimmo Iannelli and Fabio Milner. *The Basic Approach to Age-Structured Population Dynamics*. Springer, 2017.

[25] Bruce L Jones. Actuarial Calculations Using a Markov Model. *Transactions of Society of Actuaries*, 46:227–350, 1994.

[26] Robert E Kass and Duane Steffey. Approximate bayesian inference in conditionally independent hierarchical models (parametric empirical bayes models). *Journal of the American Statistical Association*, 84(407):717–726, 1989.

[27] Nathan Keyfitz and Andrei Rogers. Simplified Multiple Contingency Calculations. *The Journal of Risk and Insurance*, 49(1):59–72, 1982.

[28] Anastasia Kostaki and Vangelis Panousis. Expanding an abridged life table. *Demographic Research*, 5:1–22, 2001.

[29] Nan Laird. Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, 73(364):805–811, 1978.

[30] Nikolaos Limnios and Gheorghe Oprisan. *Semi-Markov processes and reliability*. Springer Science & Business Media, 2012.

[31] Alberto Malinverno and Victoria A. Briggs. Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes. *Geophysics*, 69(4):1005–1016, 2004.

[32] Donald R Mcneil, T James Trussell, and John C Turner. Spline Interpolation of Demographic Data. *Demography*, 14(2):245–252, 2017.

[33] Peter Müller and Fernando A Quintana. Nonparametric Bayesian Data Analysis. *Statistical Science*, 19(1):95–110, 2004.

[34] Claudia Nau and Glenn Firebaugh. A New Method for Determining Why Length of Life is More Unequal in Some Populations Than in Others. *Demography*, 49(4):1207–1230, 2012.

[35] László Németh and Trifon I. Missov. Adequate life-expectancy reconstruction for adult human mortality data. *PLoS ONE*, 13(6):1–8, 2018.

[36] Paul Norman, Alan Marshall, Chris Thompson, Lee Williamson, and Phil Rees. Estimating detailed distributions from grouped sociodemographic data : ' get me started in ' curve fitting using nonlinear models. *Journal of Population Research*, 29(2):173–198, 2017.

[37] Wilma J Nusselder and Caspar W N Looman. Decomposition of differences in health expectancy by cause. *Demography*, 41(2):315–334, 2004.

[38] Marcelo Pereyra. Maximum-a-posteriori estimation with bayesian confidence regions. *SIAM Journal on Imaging Sciences*, 10(1):285–302, 2017.

[39] Marcelo Pereyra, Philip Schniter, Emilie Chouzenoux, Jean-Christophe Pesquet, Jean-Yves Tourneret, Alfred O Hero, and Steve McLaughlin. A survey of stochastic simulation and optimization methods in signal processing. *IEEE Journal of Selected Topics in Signal Processing*, 10(2):224–241, 2016.

[40] P Peristera and A Kostaki. GRADUATION OF MORTALITY DATA USING KERNEL ESTIMATES. pages 561–568, 2004.

[41] A. E. Renshaw and S. Haberman. On the graduations associated with a multiple state model for permanent health insurance. *Insurance: Mathematics and Economics*, 17(1):1–17, 1995.

[42] Hilary L Seal. Multiple Decrements or Competing Risks. *Biometrika*, 64(3):429–439, 1977.

[43] Hans J. Skaug and David A. Fournier. Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. *Computational Statistics and Data Analysis*, 51(2):699–709, 2006.

[44] Eric V. Slud and Jiraphan Suntornchost. Parametric survival densities from phase-type models. *Lifetime Data Analysis*, 20(3):459–480, 2014.

[45] Nico M Van Dijk. Uniformization for nonhomogeneous markov chains. *Operations research letters*, 12(5):283–291, 1992.