Published in final edited form as:

J Immunol Methods. 2014 December ; 414: 32–50. doi:10.1016/j.jim.2014.08.001.

Bayesian Immunological Model Development from the Literature: Example Investigation of Recent Thymic Emigrants†

Tyson H. Holmes^{a,*} and David B. Lewis^b

Tyson H. Holmes: tholmes@stanford.edu; David B. Lewis: dblewis@stanford.edu

^aStanford Center for Sleep Sciences and Medicine, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, 401 Quarry Road, Stanford, California, 94305-5718, USA. Telephone: 1 (650) 725-2925. Fax: 1 (650) 725-7341

^bDepartment of Pediatrics, Stanford Immunology Program, and the Institute for Immunity, Transplantation, and Infectious Disease, Stanford University School of Medicine, CCSR Building, Room 3215A, 269 Campus Drive, Stanford, CA 94305-5366, USA

Abstract

Bayesian estimation techniques offer a systematic and quantitative approach for synthesizing data drawn from the literature to model immunological systems. As detailed here, the practitioner begins with a theoretical model and then sequentially draws information from source data sets and/or published findings to inform estimation of model parameters. Options are available to weigh these various sources of information differentially per objective measures of their corresponding scientific strengths. This approach is illustrated in depth through a carefully worked example for a model of decline in T-cell receptor excision circle content of peripheral T cells during development and aging. Estimates from this model indicate that 21 years of age is plausible for the developmental timing of mean age of onset of decline in T-cell receptor excision circle content of peripheral T cells.

Keywords

Bayesian estimation; Immunological model; Human T-cell development; T-cell receptor excision circle; Naïve CD4 T cell; Recent thymic emigrant

†Abbreviations: RTE - recent thymic emigrant; TREC - T-cell receptor excision circle; TREC-CD4NT - TREC content per naïve CD4 T cell.

© 2014 Elsevier B.V. All rights reserved.

*Corresponding author: tholmes@stanford.edu.

The authors have no competing financial interests to declare.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

The peripheral naïve T-cell compartment is maintained by the *de novo* production of naïve T cells by the thymus and by peripheral naïve T-cell homeostatic proliferation (Kilpatrick et al., 2008). Involution of the thymus during adulthood is well known; but thymic structure also changes during childhood (Flores et al., 1999; Sklair-Levy et al., 2000), with peak cellularity achieved between three to six months of age (Weerkamp et al., 2005). The thymus may provide the newborn and young infant with a “head start”; but as the body grows after birth (Dowling and Hodgkin, 2009), there is a progressive decrease in the proportion of the peripheral T-cell compartment that is recent thymic emigrants (RTEs; Haines et al., 2009). While, theoretically, increased frequency of defects in hematopoietic thymic precursors (Montecino-Rodriguez et al., 2005) could contribute to a decline of peripheral RTE content, this seems unlikely to be significant during development of the healthy child. Prolongation of the naïve T-cell lifespan (Dowling and Hodgkin, 2009) could partially offset the rate of decline in RTE production during youth.

Naïve T cells of thymic origin are particularly enriched in T-cell receptor excision circles (TRECs) derived from δ Rec- ψ Ja rearrangements of the TCR- α/δ gene locus; and TRECs may have long-term stability within these relatively long-lived cells (Hazenbergh et al., 2001). On this basis, TREC content of the peripheral T-cell compartment may be used to assess the contribution of cells of thymic ancestry (i.e., RTEs and their progeny) to the periphery. Throughout, we use a normalized measure of this contribution, TREC content per naïve CD4 T cell (TREC-CD4NT).

Bains et al. (2009a) were interested specifically in CD4 T-cell homeostasis during youth and re-analyzed the data of Douek et al. (2001) on TREC-CD4NT over ages 0 to 20 years. They concluded that mean TREC-CD4NT was “constant” over this period, failing to reject the null hypothesis of no trend with age ($p = 0.11$, Bains et al., 2009a). While limiting analysis to ages 0 to 20 years seems appropriate to their purpose, this yielded a sample size of only 13 individuals. Failure to reject a null hypothesis of no trend in a sample this small could easily be due to very low statistical power, which is far from sufficient evidence of constancy. Moreover, testing the null hypothesis of absolutely no change only bluntly addresses the issue. A more refined approach is to ask at what age does TREC-CD4NT begin to decline and, once initiated, at what rate?

The purpose of this paper is to illustrate the utility of Bayesian methods for quantitatively addressing questions about immunological systems and processes using estimates drawn from the literature; and our approach is illustrated below in detail in application to the question about decline in TREC-CD4NT with development and aging. We begin by proposing a parsimonious mathematical model for this process. Estimation of model parameters begins by making exact, quantitative statements about an initial, most uninformed state of knowledge about parameter values. We then sequentially update those with estimates from the literature, bringing more information to bear on the problem than can be provided by any single data set.

Our objective is to provide the working immunologist with a clear description of the specifics that are necessary to carry out this procedure. This is not simply a conceptual introduction to Bayesian methods of parameter estimation, although that is outlined here too, effectively offering a tutorial on the topic to the immunologist. We go further and provide something more concrete and practical: a specific step-by-step method that allows the immunologist to 1) formulate a biologically-meaningful, statistical model for the specific purpose of applying the method of informative priors that is detailed in Sections 3.2 through 4.2, 2) update estimates of the model's parameters, first from the literature and then from source data, and 3) individually weight those updates based on objective measures of the scientific quality of their sources. For those readers who are interested in additional mathematical and other technical details, extensive footnotes are provided. The online supplement includes code for performing the analysis presented in the paper.

2. Theory

Analysis was performed on an original data set from Douek et al. (2001) as provided by Daniel Douek (Daniel Douek, personal communication). These authors measured TREC content via quantitative PCR in sorted CD4 T cells in 33 healthy individuals. Initially, we fit a regression model to these data using a piecewise linear spline basis (Hastie et al., 2009) with mean age of onset of decline in TREC content represented by a knot ("hinge" or abrupt change in slope) fixed at 20 years of age per the assumption of Bains et al. (2009a) (data not shown). We soon realized that this model may be unnecessarily restrictive, as it presumes that we know when TREC-CD4NT begins to decline prominently in the periphery. That is, instead of being fixed in advance, the location of the change in slope could be estimated from data. Derivation of that model is as follows.

2.1. Mathematical model of TREC-CD4NT decline

Let Y be a random variable representing TREC-CD4NT expressed as the quantity of TREC-bearing cells per 150,000 naïve CD4 T cells (see below for choice of units); and let $\mu(a)$ denote the population mean of Y at age a years.¹ For simplicity of notation, hereafter we simply denote $\mu(a)$ by μ . Consider a simple system in which rate of decline in μ with age in years a is directly proportional to current mean μ , which we can express as $d\mu/da = \beta_1\mu$, for some constant β_1 . Solving, $\mu = \beta_0\exp(\beta_1 a)$, so that β_0 is the population mean at birth, when $a = 0$ (i.e., β_0 represents an *initial condition*). Suppose further that we allow for the possibility, as in Bains et al. (2009a), that no decline ($\beta_1 = 0$) occurs up to an mean age of onset, which here we denote by δ . We incorporate this into our model as $\mu = \beta_0\exp(I[\delta < a]\beta_1(a - \delta))$, where the indicator function $I[\delta < a] = 1$ for ages greater than δ and 0 otherwise. That is, $\mu = \beta_0$ up to and including age δ . In practice, we observe values of Y , not the population mean μ . Thus we need to include a term for the individual variation (residual error) E about the mean as $Y = \beta_0\exp(I[\delta < a]\beta_1(a - \delta) + E)$. On the natural logarithmic scale,

$$\ln Y = \beta_0 + I[\delta < a]\beta_1(a - \delta) + E. \quad (E1)$$

¹Specifically, this is a conditional expectation $\mu(a) = E[Y|a]$.

Model E1 is illustrated schematically in Figure A.1 of the online supplement. Here we will assume E is normally distributed with mean 0 and variance σ^2 (an assumption that we confirmed during model fitting). Model E1 may appear to be linear or, more specifically, piecewise linear because it describes two straight lines joined at a knot where $a = \delta$. However, strictly, the model is nonlinear, due to the presence of the parameter product $I[\delta < a]\beta_1(a - \delta)$. In contrast, a truly linear model would consist of a sum of additive terms, each of a *single* parameter, such as the simple linear regression model $Z = \zeta_0 + \zeta_1 X$. In the statistical context, the spline term $I[\delta < a]\beta_1(a - \delta)$ would be linear if knot δ were a known constant and not a parameter to be estimated. A piecewise linear model with a “corner” at δ may seem like a rough approximation for what must be a much smoother underlying biological process. However, the primary purpose of the present paper is illustration of a method, which is facilitated by decomposing decline into separable components of mean initial level β_0 , mean age of onset δ , and mean rate of decline β_1 . Moreover, the Bayesian method of posterior model averaging (Hartemink et al., 2002) allows us to obtain a “smoothed” estimate of decline on average across the populationⁱⁱ using the simple model given by E1, as described in section 4 and illustrated in section 5.

2.2. Introductory example: Bayesian estimation and informative prior construction

A Bayesian approach (Bolstad, 2007; Leonard and Hsu, 1999) to estimating the parameters of E1 assumes that, in addition to the random residual error E , possible values of the true underlying model parameters can also be conceptualized as random variables. This allows us to assign distributions to these parameters that reflect our state of knowledge or *belief* (or lack thereof) about the plausibility of each possible value of a parameter (Kruschke, 2011, pp. 68–69). The basic approach of Bayesian estimation is illustrated in Figure 1. Following an example given in Rice (1995, pp. 592–593), suppose that random variable Z takes the value 1 if a specific event occurs and 0 otherwise; and that for a set η observations, one wishes to estimate parameter ω , which denotes the probability that any one observation is an event. For instance, suppose that ω represents the probability that a naïve CD4 T-cell is of thymic origin and that η is the quantity naïve CD4 T-cells in a 1- μ L aliquot. We assume that random variable Z follows a binomial distribution with parameters ω and η . If we do not know anything about the value of ω , a Bayesian approach could assume that parameter ω has a uniform *prior* distribution on the interval from 0 to 1, as depicted by the solid line in Figure 1.ⁱⁱⁱ The term “prior” indicates antecedence to information (i.e., data); and this prior distribution is “*uninformative*” in the sense that it contains absolutely no information about the precise location of ω other than its mathematical bounds.^{iv} Suppose that 30 observations (i.e., it is known that $\eta = 30$) are obtained of which 19 are events. We can then *update* the prior distribution with this information to obtain the *posterior* distribution for ω , shown by the dashed line in Figure 1. In an iterative process, this posterior can then serve as a less vague or *informative* prior and, possibly, in turn be updated with additional information to

ⁱⁱBecause the data available to us are cross-sectional, parameter estimates and the fitted curve represent averages across the population (i.e., population marginal). If instead longitudinal data were available, we could use model E1 to estimate the trajectory of an average individual. However, very few studies collect longitudinal data over decades, with rare exceptions such as the Dunedin study (<http://dunedinstudy.otago.ac.nz/>).

ⁱⁱⁱA prior could also be selected for η , but here we are taking η to be known by design to aid exposition.

^{iv}Purely non-informative priors also exist, which take a uniform distribution on $(-\infty, \infty)$, $[0, \infty)$, etc. (Congdon, 2006, p. 4), as appropriate; but because these may introduce artifacts into posterior distributions, these priors will not be considered further here.

yield a new posterior for ω (e.g., based on a new additional sample of 30 observations). Guikema (2007) refers to this process of starting with a fully uninformative prior, updating that prior with data via a likelihood, and then employing that empirically informed prior in the next round of updating as “two-stage” use of a “pre-prior.”

3. Calculation

3.1. Step 1: Immunological model formulation

Estimation proceeds in four steps. In *Step 1*, a particular model for the process of interest needs to be formulated. Section 2.1 contains a detailed example of a particular model’s formulation; and in this section we provide general guidelines for this process. Model formulation is distinctive from *model development*. Formulation is the first step in which biological concepts are formalized as mathematical constructs, whereas development is the more encompassing, iterative process of formulation, testing, revision of formulation, etc. See Motta and Pappalardo (2013) for a general overview of the process of model development. The statistical aspects of model development are discussed in Section 6.4.

A generic representation of a mathematical model of an immunological system is $\mathbf{Y} \leftarrow h[\mathbf{X}|\boldsymbol{\beta}] + \mathbf{E}$.

3.1.1 Inputs $\mathbf{X}$ —An input/output framework is typically assumed, indicating directional relationship, and here represented by the left arrow, between one set of features $\mathbf{X}$ and another set $\mathbf{Y}$. While models of association without directionality are certainly possible, the ultimate purpose of immunological studies is to understand mechanism; and mechanisms entail sequences of events, precedence and possibly causality. The sometimes difficult problem of demonstrating causality is central to model development, as defined above, and can require a combination of manipulative experimental design and/or specialized statistical methods of causal inference (Pearl, 2009) that will not be discussed further here. Models are often written as equations (e.g., Model E1) with inputs on the right and the left-arrow implied by convention.

One of the temptations in building models is to specify many input variables. With the addition of each new input variable, the dimensionality of the model increases. This is problematic at a fixed sample size, because an increase in dimensionality produces a decline in “sampling density” (Hastie et al. 2009, p. 23), which destabilizes estimates of model parameters. One can posit a high-dimensional model and employ variable selection methods to reduce dimensionality (e.g., Ishwaran and Rao, 2005). That empirical process of dimension reduction goes beyond model formulation to model development, as just defined. Presumably that empirical dimension reduction would precede the Bayesian estimation method outlined in the remainder of this paper that borrows information across multiple studies for each parameter to be estimated, which could be a mammoth task for a high-dimensional input $\mathbf{X}$.

3.1.2 Outputs $\mathbf{Y}$ —The simplest models are one-to-one, with one input and one output, such as Model E1. Many-to-one representations are also commonplace (e.g., several input variables employed to predict a single outcome variable). One-to-many and many-to-many

relationships are also possible, where output may be vector valued (e.g., spatial coordinates) or an unordered multi-member set, such as a list of upregulated cytokines, or a network (e.g., cascade). For practical purposes, all inputs and outputs should be variables that can be measured by the study.^v

3.1.3 Model form $h[\cdot]$ —This is the form of the model for the mean—namely, how mean values in output $\mathbf{Y}$ vary in a way that depends upon inputs $\mathbf{X}$. Distinct from this, and the topic of Section 3.1.5, is the model for how the $\mathbf{Y}$ vary about their mean. Models of the mean may take many forms, from lines and planes to curved surfaces, step functions, truth tables (Stoll, 1963), networks, etc. Traditionally, modeling of biological systems has employed difference and differential equations or systems of equations because many components of biological processes from the subcellular to the population level can be represented in terms of rates and fluxes (Edelstein-Keshet, 1988). Also, linear models for the mean have been popular because they require fewer parameters than nonlinear models (Neter et al. 1996, p. 534). Linear models (defined in Section 2.1) are often approximations for the complex underlying processes that they are designed to represent; and, in fact, tools such as series representations (Thomas and Finney, 1996) can be employed to provide varying levels of approximation to nonlinear functions. Nonlinear models are more standard in some fields, such as in pharmacokinetics (Racine-Poon and Wakefield, 1998). More recently, partial differential and cellular automata, as well as cellular automata and lattice Boltzman methods have been combined in “hybrid” modeling techniques, which may be especially useful when modeling across multiple scales with stochastic components (Alemani et al., 2012). Because all models are approximations of biological systems being studied, no one model is necessarily “correct.” Instead, of the many possible models that can provide a close approximation to that system, strongest candidates are those that are parsimonious in structure and that facilitate interpretation. Parsimony in the model for the mean can avoid *overfitting* (Harrell, 2001), where in the model fits the data in the sample on hand very well, including noise that is idiosyncratic to that particular sample; as a result, the fit is unstable and will not validate in new samples from the same population. Complexity can often be reduced through judicious choices in parameterization (Section 3.1.4). The selected form of the model should also facilitate visual display (Tufte, 2001) so that results can be more readily conveyed to the target audience. The rise in graphical models (e.g., Scutari and Strimmer, 2008) may due to the fact that these are quite flexible in terms of the functional forms that they can accommodate but they also lend themselves readily to visual display that aids interpretation.

3.1.4 Model parameters β —Parameters are what define a model in terms of *scale*, both in terms of the mean (Section 3.1.3) and the distribution of output values about the mean (Section 3.1.5). To illustrate, two models for the mean may have identical form but very different scale ($y = 4e^{2x}$ vs. $y = 119e^{8.2x}$). Indeed, the *raison d’être* for parameter estimation, including Bayesian estimation via informative priors, is that parameters are what tie a

^vMeasurability is not always necessary. A variable may be a scientifically necessary element of a model but one that cannot be measured. “Latent” variables can be accommodated by estimation procedures such as the expectation-maximization algorithm in the frequentist setting (Dempster et al. 1977) and data augmentation within Bayesian estimation (Tanner and Wong 1987).

particular model functional form $h[]$ to a particular population being studied via scale, whether that is a particular population of cells, patients, etc.

A model will contain specific unknown parameters to be estimated. Model E1, of parameters β_0 , β_1 , δ , and σ is an example. Those parameters that are not unknown are constants. To avoid unnecessary complexity in estimation, model formulation should always include setting all known parameters to their constant values (Cox et al., 2006). To the extent possible, the model should be constructed of parameters that have direct biological interpretation, as explained for the parameters β_0 , β_1 , and δ of Model E1 in Section 2.1. Parameters without clear interest to the investigation can be broadly termed *nuisance* parameters (Stuart et al. 1999, p. 197), in immunology interest being terms of biological and clinical meaning. At times, nuisance parameters appear in the model of the mean (e.g., θ in Roberts, 1996). However, they mostly arise as higher moments of statistical distributions (dispersion parameters, shape parameters), such as σ of Model E1. As demonstrated below, for example, the process by which we estimate rate of decline β_1 following the mean age of onset δ using parameter estimates drawn from the literature will require introduction of a two additional parameters, φ_1 and φ_2 , respectively. A few such nuisance parameters may be unavoidable and manageable; but, in general, parameterization, both nuisance and biological, should be only rich enough to capture the essential elements of the system being studied. Redundant parameters can also arise, which are parameters that are given different labels but actually represent the same scaling factor. These may arise most often in models that consist of systems of equations. Excess parameters also arise when a model is built with more inputs than are truly necessary to study the system (Section 3.1.1) because as a general rule each new input variable adds at least one additional parameter to the model. Any form of over-parameterization, whether due to nuisance, redundant or excess parameters, can cause overfitting (and ensuing instability), as described in Section 3.1.3, and muddle interpretation.

3.1.5 Residual E—Residuals represent the admission that processes exist that are too trivial to concern us (i.e., “noise”). Model formulation should be complete enough to ensure that residuals do not contain processes that are substantial because this can bias estimates of specified parameters (e.g., Groenwold et al. 2009). In fact, when **E** retains much of the system’s structure, the model formulation is *misspecified*, especially if a statistical distribution is chosen for **E** that ignores that structure. This is the subject of Sections 5.2 and 6.4 on assessment of model fit. Models without unexplained error, in which the entire process is fully understood and specified in $\mathbf{Y} \leftarrow h[\mathbf{X}|\boldsymbol{\beta}]$, are rare to absent in biology.

In order to complete model formulation and progress into model development via parameter estimation (Sections 3.2 through 3.4), a decision needs to be made about the statistical distribution for the residuals **E**. This process begins by asking if the outputs **Y** are measured on continuous or discrete (counts, categories, etc.) scales. Introductory (Rice 1995) or more advanced texts (Stuart and Ord 1994) can then be used to guide selection of a particular parametric distribution. As indicated in Section 3.1.4, parametric distributions contain additional parameters that will often need to be estimated; so choice of distribution will be guided in part by what distribution will be practical to fit without risking overfitting and introduction of too many nuisance parameters. Availability of information—either source

data or published values (Section 3.3)—for estimating those parameters also needs to be preliminarily assessed. Though the menu of available distributions is broad, if none seem suitable then *semiparametric* methods can be employed, one of which is discussed in Section 6.7.1.

3.1.5 Summary of model formulation—In summary, at a minimum, the following series of questions should be addressed when formulating an immunological model for purposes of the method of informative priors outlined in the sections below.

1. What are the *measurable* inputs and outputs to the model? Are inputs and outputs arranged in a meaningful way in terms of directionality, precedence or cause and effect? If either input or output is multivariate, can we, through application of scientific reasoning, reduce the quantity of variables *a priori* to an essential set necessary to address the scientific question of interest or do we need first to apply some form of empirical variable selection? Have any important inputs been omitted, thereby causing the residual to be structured and the model to be misspecified?
2. What is the structural form of the model? Is that form too complex so that some approximation (e.g., linear) will suffice? Will the selected form facilitate biological interpretation and possibly visual display?
3. What are the model parameters? Are all proposed parameters essential and are all essential parameters present? Can any of the parameters be represented by a constant? Are any parameters redundant with one another so that they can be combined? Can the quantity of inputs be dropped to eliminate excess parameters?
4. What particular statistical distribution, if any, will model the error term and what is the rationale for this selection?

3.2. Step 2: Uninformative prior specification

For Step 2, an uninformative prior distribution needs to be identified for each of the parameters. As used here, an uninformative distribution is one which reflects a state of no empirical knowledge about possible values of a parameter. The distribution of a random variable has many properties; often among the three most important for choosing a vague, uninformative prior are its *support*, *expectation*, and *variance*. The support is the domain of possible values that the random variable may assume. For example, a beta distribution, which is bounded below by 0 and above by 1 may be a good choice for a prior of a parameter that is a proportion.^{vi} The variance, as a measure of dispersion, describes how uninformative the prior is. As its variance decreases, the prior distribution becomes more informative and less vague. As a rule of thumb, the prior distribution's variance should be at least on order-of-magnitude larger than the posterior's variance (Spiegelhalter et al., 1996, p. 28). That said, one should avoid making a prior too “flat,” even to the point of being improper^{vii}, as this can create problems with the posterior, including practical problems with

^{vi}The uniform distribution used in the example at the end of Section 2 is a specific type of beta distribution.

^{vii}An improper prior violates the law of total probability by not integrating or summing to 1 over the support of the parameter.

the computation of the posterior, such as slow convergence of the Markov Chain Monte Carlo algorithm (Section 4.2). The uninformative quality of the prior is often also typified by the prior's expectation (i.e., mean), as an uninformative expectation can sometimes be far from where our experience, which is empirical, suggests that mean should be.

An underappreciated problem is that uninformative priors may be unintentionally made informative through use of standard transformations. A classic example arises in logistic regression for the probability of an event. Especially when in doubt, good practice is to plot the distribution of the parameter of interest on the transformed scale to ensure that it has not been made informative by transformation (e.g., plot posterior distribution without applying the likelihood update, Section 3.3). This topic is detailed at length in Seaman et al. (2012).

3.3. Step 3: Empirically updating to construct informative priors

With Step 3, each uninformative prior is updated with empirical information (data)^{viii} to generate an informative prior. Specifically, we want to characterize the distribution of the parameter *conditional* on the observed data. The goal is a posterior distribution. These data can be summary, such as parameter estimates obtained from the literature, or they can be a source sample, with data at the level of individual observations. Below, we illustrate a systematic process wherein uninformative priors are first updated with published parameter estimates and then with source data. Preliminary to writing computer code, the analyst performs the mathematical calculations necessary to update the uninformative priors with parameter estimates from the literature. This step is particularly valuable because these calculations provide an important opportunity for the analyst to think critically about 1) how this first update is being performed and 2) how the resultant posteriors will be used in the next round of updates with source data. Next, the analyst instantiates these posterior distributions into computer code as informative priors. That code reads source data to update these informative priors to yield the final and most informed posterior distribution, from which the final point estimates (and intervals) of the parameters are obtained.

Updating is facilitated by application of Bayes Theorem, which we state here for a vector of p continuous random variables $\mathbf{y}$, because continuous random variables are used throughout the example below:

$$g_1(\boldsymbol{\omega}|\mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\omega})g_0(\boldsymbol{\omega})}{\int_{\Omega} f(\mathbf{y}|\boldsymbol{\omega})g_0(\boldsymbol{\omega})d\boldsymbol{\omega}} \quad (\text{E2})$$

Here $g_1(\boldsymbol{\omega}|\mathbf{y})$ is the *joint distribution*^{ix} of the vector of parameters $\boldsymbol{\omega}$ *posterior* to (i.e., conditioned upon) updating with some observed data (parameter estimate or source) $\mathbf{y}$; $f(\mathbf{y}|\boldsymbol{\omega})$ is the likelihood of the data conditional on the parameters (which will be defined shortly); $g_0(\boldsymbol{\omega})$ is the uninformative joint distribution of the parameters prior to updating with any data (Leonard and Hsu, 1999, p. 106); and Ω is the support of $\boldsymbol{\omega}$. As such, the numerator of this ratio is the product of the likelihood and prior distribution of the parameters (Leonard

^{viii}Uninformative priors may be updated by eliciting subjective information, such as expert opinion (Kadane and Wolfson, 1998), an approach that will not be pursued here.

^{ix}A distribution that describes how two or more random variables vary together is a joint distribution.

and Hsu, 1999, p. 106). In the denominator, integrating over the domain of the parameter vector gives the *prior predictive distribution* of the data (Leonard and Hsu, 1999, p. 106), which is a constant for a given data vector $\mathbf{y}$. Because this is a constant, we can express the relationship between posterior and prior more simply in terms of a proportionality

$$\text{posterior of parameter} \propto \text{likelihood} \times \text{prior distribution of parameter.} \quad (\text{E3})$$

This streamlining simplifies estimation substantially because, in practice^x, updating to form the posterior does not require performing the integration within the denominator of E2.

Since the uninformative prior $g_0(\omega)$ was specified in Step 2, *in step 3 we need to specify the likelihood of the data $f(\mathbf{y}|\omega)$* . The likelihood function defines the plausibility of different hypothesized values of ω for the observed data $\mathbf{y}$ (Edwards, 1992, p. 9; Spiegelhalter et al., 2004, p. 18).^{xi} For example, suppose the datum is a parameter estimate drawn from the literature. A parameter estimator $h_{\omega}(\mathcal{S})$ that summarizes a random sampling $\mathcal{S}$ of data is itself a random variable (because it is a function of one) and thus possesses a distribution. That distribution can define a likelihood. A given sample $\mathcal{S} = \mathbf{S}$ yields estimate $\mathbf{y} = h_{\omega}(\mathbf{S})$. Sampling distributions for many estimators are approximately normal, with approximation to the normal improving with the size of sample $\mathcal{S}$.^{xii} This motivates using the normal

density $f(\mathbf{y}|\omega) = \frac{1}{(\sqrt{2\pi})^p |\Sigma|^{1/2}} e^{-(\mathbf{y}-\omega)' \Sigma^{-1} (\mathbf{y}-\omega)/2}$ for the likelihood in ω . That is, given estimate $\mathbf{y}$, $f(\mathbf{y}|\omega)$ describes how plausible any value of ω is for the observed data $\mathbf{y}$. This is exactly how the likelihood serves to update the prior for ω to yield a posterior distribution of ω that is a distribution of plausibilities of ω given data $\mathbf{y}$.

3.4. Step 4: Differential weighting of updates

3.4.1 Computational method—The final step, *Step 4*, asks, “Are all updates of equal importance; or should some updates be weighted more heavily than others?” An investigator may believe that some sources of information are of greater relevance, higher quality, etc. and wish to differentially weight these sources on that basis. Sources of information may be available that vary considerably in how closely they match the population of interest. One way to improve matching is through introduction of covariates—covariates along which the sampled populations are meaningfully discrepant—into the likelihood (E2) or into weights to adjust all sources toward the population of interest, perhaps even toward an ideal population that is not observed at all (Holmes et al., 2012).

^x...including use of numerical sampling methods (Kruschke 2011, p. 166, and detailed below).

^{xi}The data as observed are fixed and known; so the question is, “What value of the parameter is the most plausible explanation of the observed data?”

^{xii}This is an example of a central limit theorem, of which many variants exist. The most familiar concerns sums of independent and identically distributed random variables (Ross 1998, p. 399); although central limit theorems exist for other circumstances as well, including for independent but not identically distributed random variables (Ross 1998, p. 406). The latter is useful for us here because the various random quantities arising from fitting data to a nonlinear regression model (e.g., the of E1) are often independent but not identically distributed (due to heteroscedasticity). Jennrich (1969) demonstrated that the sampling distribution of a nonlinear regression parameter estimator tends toward a normal distribution approximately, with approximation improving with increasing sample size of the source data.

The challenge with weights is that they are unknowns. Exactly which values should be assigned to each weight for each source of information is completely open to question in each application. Also different means of deriving those weights are available. Some authors advocate for eliciting expert opinion in order to assign weights (Rietbergen et al., 2011). Because such subjective weighting schemes can be particularly difficult to defend, in our view, a less controversial approach is to derive these weights based on some *objective* measures of either data quality or sample relevance. This is discussed at the end of this Section and illustrated in Section 5.3.

Computationally, each updated prior can be weighted using a power prior formulation (Ibrahim and Chen, 2000). Specifically, let $g_0(\omega)$ denote the uninformative prior for some parameter ω ; and let the product $f(\mathbf{y}_1|\omega)^{w_1} f(\mathbf{y}_2|\omega)^{w_2} \dots f(\mathbf{y}_m|\omega)^{w_m}$ denote the likelihood for the m initial updates (sources) of respective weights $\mathbf{w} = \{w_1, w_2, \dots, w_m\}$ and data $\mathbf{y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m\}$. Then, the updated power prior is

$$g_1(\omega|\mathbf{w}, \mathbf{y}) \propto f(\mathbf{y}_1|\omega)^{w_1} f(\mathbf{y}_2|\omega)^{w_2} \dots f(\mathbf{y}_m|\omega)^{w_m} g_0(\omega), \quad (\text{E4})$$

which is posterior relative to $g_0(\omega)$.^{xiii} Formulation E4 not only shows how weights are applied but demonstrates the point made above about the iterative nature of the updating process, here with iterations being based on data $\mathbf{y}_1$ then on $\mathbf{y}_2$ then on $\mathbf{y}_3$, etc. by taking their product.

3.4.2 Specific weighting schemes—Given this computational method for applying weights, what are some computational techniques for deriving the weights to apply? Suppose we have $i = 1, 2, \dots, m$ sources.

1. *Equal weighting* $\mathbf{w} = \{1, 1, \dots, 1\}$: Here no differential weighting is applied. Some might see this as the most objective weighting but it does ignore completely any information available about variability in relevance or reliability of information sources.
2. *Unequal weighting based on reliability* $\mathbf{w} = \left\{ \frac{1}{\vartheta_1}, \frac{1}{\vartheta_2}, \dots, \frac{1}{\vartheta_m} \right\}$: In this context, “reliability,” as denoted by ϑ_i , is measured in terms of the amount of measurement error present in a data source. Technical error, or the standard deviation among technical repeats σ_T , when reported, is a very useful measure of reliability $\vartheta = \sigma_T$, when converted to common units. Weighting based on reliability also applies to *reporting* error, wherein parameter estimates are reported as an interval rather than as a point value (e.g., average cell count ranged from 10^3 and 10^4 per mL). This motivates a precision-based differential weighting scheme of $w \stackrel{\text{def}}{=} \sqrt{12}/(u-l)$, which is the reciprocal of the standard deviation of a uniform random variable on interval $[l, u]$. Section 5.3 illustrates application of precision-based weights as well as a form of reliability weighting based on years since publication.

^{xiii}Weights may contain error, especially if empirically derived. Propagation of this uncertainty through to $g_1(\omega|\mathbf{w}, \mathbf{Y})$ could be rigorously explored but will not be investigated here.

3. Unequal weighting based on relevance $\mathbf{w} = \left\{ \frac{1}{E_1}, \frac{1}{E_2}, \dots, \frac{1}{E_m} \right\}$: The use of covariates was mentioned earlier in this Section as a means of regression adjustment via the likelihood for differences in relevance among source's sampled populations. Suppose instead covariates $x_1, x_2, \dots, x_p$ are to be used to form weights. For the i^{th} source, $E_i \stackrel{\text{def}}{=} \sqrt{\sum_{k=1}^p (x_k - x_{ki})^2}$, where x_k and x_{ki} are the values of the k^{th} covariate in the population of interest and i^{th} source, respectively.

3.4.3 The functional role of weighting—Differential weighting should be standard when multiple sources of information are being employed for estimation of the parameters of a model. Moreover, different types of objective weighting schemes should be applied in order to assess just how sensitive conclusions are to each source and means of weighting. We advocate for erring toward being more rather than less inclusive of possible sources of information that may pertain to estimation of the model's parameters and then allow objectively-derived differential weights reveal which, if any, particular sources are of especially limited utility in terms of reliability and/or relevance.

3.5. Detailed example

The remainder of this section (Calculation), Section 4 (Methods) and Section 5 (Results) detail implementation of these steps. Estimation of model E1 is used to illustrate the process.

3.5.1. Detailed example: uninformative prior specification—Having defined our model (E1) of developmental decline in TREC-CD4NT in Section 2, Step 2 of estimation of the parameters begins with specification of a vague, uninformative prior for each of the four parameters. 1) We assume parameter β_0 has a normal prior distribution $g_0(\beta_0)$ of mean $\ln(150,000/2)$ TREC-CD4NT and variance 10^3 . Douek et al. (2001) express TREC-CD4NT as “TREC per μg naïve T cell DNA” and report that “1 μg of genomic DNA is equivalent to 150,000 cells.” 150,000/2 falls halfway between all or none of the observed naïve CD4 T cells per μg of naïve T-cell DNA being TREC-CD4NT. The large variance employed here, which meets the criterion of being at least an order of magnitude larger than the posterior variance (data not shown), reflects the uninformative state in which nothing empirical is known about β_0 . 2) Regarding β_1 , while initial suspicion is that TREC-CD4NT should decline after onset so that $\beta_1 < 0$, an uninformative, vague prior is a normal distribution $g_0(\beta_1)$ of mean equal to zero and variance (10^3) , which also meets the criterion of being at least an order of magnitude larger than the posterior variance (data not shown). 3) For mean age of onset δ , an uninformative and vague prior is a uniform distribution $g_0(\delta)$ on $(0, 126)$, where $\psi = 126$ years is currently theorized to be the maximum of human longevity (Weon and Je, 2009).^{xiv} Use of a uniform distribution admits that, prior to gathering any empirical information on mean age of onset, any age between 0 and 126 years is equally likely to be the mean age of onset of decline in TREC-CD4NT. The expectation for this prior is $126/2$

^{xiv}Weon and Je (2009) do make reference to empirical data to extrapolate to 126; so, in this sense, our prior distribution here for δ is not purely uninformative. In theory, one could designate the upper bound ψ on human lifespan to be a nuisance parameter of the prior distribution of δ . This may not be necessary because a uniform distribution on $(0, 126)$ is, without question, vague with regard to mean age of onset.

years; and a uniform distribution has vague variance, here $\frac{126^2}{12} = 1,323$ squared years. 4) Finally, for the variance parameter σ^2 in TREC-CD4NT about the mean μ of E1, in the extreme case, half of all observations are $\ln(1)$ (i.e., 1 in 150,000 of a person's naïve CD4 T cells per μg of naïve T-cell DNA is TREC-bearing) and half are $\ln(150,000)$ (i.e., all 150,000 of a person's naïve CD4 T cells per μg of naïve T cell DNA are TREC-bearing). An uninformative prior for σ^2 is a uniform distribution $g_0(\sigma^2)$ on the interval of $(0, (\ln(150,000)/2)^2)$. Importantly, notice that for β_0 and σ^2 we are not using observed data but only the measurement units for TREC-CD4NT from Douek et al. (2001), as detailed above. What is important to the method here is that a consistent scale is used for all parameters; and what is of essential interest biologically is the *proportion* of total naïve CD4 T cells that are TREC-CD4NT. The denominator of 150,000 is simply a handy scaling factor for converting this proportion to a cell count. A different scaling factor could be used instead if used consistently throughout the update process for these parameters.

3.5.2. Detailed example: empirically updating priors

3.5.2.1. Detailed example: updating mean at birth: A Bayesian approach allowed incorporation of information from other published studies into the fit of the regression model by updating the above uninformative priors ("Step 3" from earlier in this section). Updating a normal prior distribution for a parameter is especially straightforward with data drawn from a normal distribution (i.e., via a normal likelihood), because prior distribution, likelihood and posterior distribution are all normal (Rice, 1995). Recall that, as defined here, β_0 is the mean quantity of TREC-bearing cells per 150,000 naïve CD4 T cells at birth. With population variance σ^2 for TREC-CD4NT (of model E1), the sample mean, which is our sample estimator $\hat{\beta}_0$ for β_0 ,^{xv} has variance σ^2/n in a sample of size n . Schönland et al. (2003) provide a sample *median* estimate of 13.4% TREC-CD4NT in cord blood from $n = 45$ newborns. Under the assumption that sampling of TREC-CD4NT values is from a normally distributed population (on the natural-logarithm scale used here), the population mean and median are equal, due to symmetry; however, the variance of the sample median is wider than that of the sample mean, σ^2/n . The variance for the sample median of Schönland et al. (2003) can be written as

$$\sigma_{med}^2 \approx \frac{\left(\frac{\sigma^2}{45}\right)}{\frac{2}{\pi} + \frac{1}{45} \left(\frac{4}{\pi} - 1\right)} \quad (\text{E5})$$

per the approximate variance inflation factor for the sample mean versus the sample median from a normal distribution that is given in Stuart et al. (1999, p. 9). Because the sample median will tend toward being normally distributed when calculated from observations drawn from a normal distribution (Chu, 1955), we can update the uninformative prior for β_0 using an approximate likelihood f_1 for a normal distribution on this observed datum of $\hat{\beta}_0 = \ln(150,000 \times 0.134)$ of expectation β_0 , and variance σ_{med}^2 as

^{xv}While Bayesian estimation is becoming more common, most parameter estimates published in the literature are obtained via frequentist (non-Bayesian) methods, such as maximum likelihood. If $\ln Y$ is sampled from a normal distribution of mean β_0 , the maximum likelihood estimate $\hat{\beta}_0$ of β_0 is the sample mean.

$$f_1(\hat{\beta}_0|\{\beta_0, \sigma^2\}) \approx \left(2\pi \left(\frac{\left(\frac{\sigma^2}{45}\right)}{\frac{2}{\pi} + \frac{1}{45}\left(\frac{4}{\pi} - 1\right)}\right)\right)^{-\frac{1}{2}} \exp\left(\frac{-(\ln(150,000 \times 0.134) - \beta_0)^2}{2\left(\frac{\sigma^2}{45}\right) / \left(\frac{2}{\pi} + \frac{1}{45}\left(\frac{4}{\pi} - 1\right)\right)}\right). \quad (\text{E6})$$

3.5.2.2. Detailed example: updating rate of decline: Hug et al. (2003) report on the rate of decline in TREC-CD4NT in $n = 53$ healthy donors. They observed an exponential rate of decline from approximately 10^5 TREC per 10^6 peripheral CD4 T cells at age 18 years to 10^4 TREC per 10^6 peripheral CD4 T cells at age 60 years (Jürgen Haas, personal communication). This gives a system in two equations:

$$\begin{aligned} \ln\left(\frac{10^5}{\frac{p}{100}10^6} \times (1.5 \times 10^5)\right) &= \hat{\gamma} + 18\hat{\beta}_1, \\ \ln\left(\frac{10^4}{\frac{p}{100}10^6} \times (1.5 \times 10^5)\right) &= \hat{\gamma} + 60\hat{\beta}_1, \end{aligned} \quad (\text{E7})$$

where p is the percentage of CD4 T cells that are naïve; and the conversion to the denominator of Douek et al. (2001) of 1.5×10^5 naïve CD4 T cells is included (again, to make sure that consistent scaling is used throughout).^{xvi} E7 reduces to a system of two

equations in two unknowns if we take $p = \frac{1.342}{0.0579} + \left(100 - \frac{1.342}{0.0579}\right) e^{-0.0579a}$, which is a fitted function from Jackola and Hallgren (1998) that describes the change in percentage naïve CD4 T cells with age a in years. System E7 solves to give $\hat{\beta}_1 \approx -0.0387$.^{xvii} Estimator E7 of slope parameter β_1 is nonlinear, with the variance ϕ_1 among estimates of β_1 being an implicit function of the four estimators that yielded values of 10^5 TREC per 10^6 peripheral CD4 T cells, 10^4 TREC per 10^6 peripheral CD4 T cells, 0.0579, and 1.342. If we assume a central limit theorem applies to estimator E7 of β_1 , we can update the prior for β_1 using likelihood f_2 for a normal distribution on the observed datum of $\hat{\beta}_1 \approx -0.0387$, expectation β_1 , and variance ϕ_1 . The approximate likelihood is then

$$f_2(\hat{\beta}_1|\{\beta_1, \phi_1\}) = \frac{1}{\sqrt{2\pi\phi_1}} \exp\left(\frac{-(\hat{\beta}_1 - \beta_1)^2}{2\phi_1}\right). \quad (\text{E8})$$

3.5.2.3. Detailed example: updating age of onset: Updating mean age of onset of decline δ poses more of a challenge. Other than what can be estimated from the data set of Douek et al. (2001), the literature appears to contain no direct information regarding δ . Information is available, however, on some measures related to TREC-CD4NT. For example, Zhang et al. (1999) report sample means by consecutive age groups for TREC content per peripheral blood mononuclear cell. Based on their fit of a nonlinear regression model to a sample of $n = 532$ “normal” controls, they conclude, “Between the teen years and age 20–25, a sharp

^{xvi}The estimate $\hat{\gamma}$ is the estimate of the intercept γ , where $\hat{\beta}_0 = \gamma$. Parameter γ is not of interest biologically unless $\delta = 0$. See Figure A. 1 in the online supplement.

^{xvii}Various applications are available for obtaining solutions to systems of nonlinear equations, including the freeware wxMaxima (© 2004–2013 Andrej Vodopivec).

drop of ~1–1.5 logs was noted, followed by a gradual decline (downslope of ~0.03/year) thereafter.” From this we derive an estimate of $\hat{\delta} = 19$ years, which is the midpoint of 13 to 25 years.^{xviii} Weerkamp et al. (2005) worked further upstream, documenting the relationship between thymic cellularity and age. Among their findings, they observed levels of T-cell precursors (CD34⁺CD1a⁺ thymocytes) to be highest in children of ages 3 to 6 months, giving an estimate of $\hat{\delta} = 3/8$ years on a sample size of $n = 66$. Being based on precursors, this estimate is doubtless biased low. So why use it? Because it does provide some rough information about the general locale of δ and that provision of information could help to reduce the variance in the estimate of δ obtained after all updates have been applied to form the final posterior distribution. This is an example of a *bias-variance trade-off*, which is encountered in many parameter-estimation settings (Section 6.2).

How do we formulate a likelihood for these estimates? To begin, estimates from Zhang et al. (1999) and Weerkamp et al. (2005) were re-scaled for fitting purposes through division by 126 years (i.e., current maximum human longevity) to avoid calculations that involve large and/or small numbers and thereby improve numerical stability of computational algorithms. Next, an appropriate distribution for $\hat{\delta}$ needs to be selected. The prior and likelihood must share a common support, here $[1, 126]$. Given that estimates based on samples from a doubly-bounded support may converge in distribution toward the normal even at very small sample sizes (e.g., Stuart and Ord, 1994, p. 389), one could formulate a likelihood for $\hat{\delta}$ based on a doubly truncated normal distribution (Stuart and Ord, 1994, p. 185). From the observed means of $\hat{\delta} \approx \{ \frac{3}{8}, 19 \}$, of expectation $\{ \beta_1, \beta_2 \}$ and variance $\{ \frac{\phi_1}{n}, \frac{\phi_2}{n} \}$, we have the likelihood

$$f_3(\langle \hat{\delta} \rangle | \{ \delta, \phi_2 \}) = \left(\frac{\frac{1}{\sqrt{2\pi(\frac{\phi_2}{532})}} \exp\left(\frac{-(\frac{19}{126} - \langle \delta \rangle)^2}{2(\frac{\phi_2}{532})}\right)}{\Phi\left(\frac{1 - \langle \delta \rangle}{\sqrt{\frac{\phi_2}{532}}}\right) - \Phi\left(\frac{-\langle \delta \rangle}{\sqrt{\frac{\phi_2}{532}}}\right)} \right) \left(\frac{\frac{1}{\sqrt{2\pi(\frac{\phi_2}{66})}} \exp\left(\frac{-(\frac{3/8}{126} - \langle \delta \rangle)^2}{2(\frac{\phi_2}{66})}\right)}{\Phi\left(\frac{1 - \langle \delta \rangle}{\sqrt{\frac{\phi_2}{66}}}\right) - \Phi\left(\frac{-\langle \delta \rangle}{\sqrt{\frac{\phi_2}{66}}}\right)} \right), \quad [\text{E9}]$$

where $\langle \delta \rangle = \delta/126$ and $\Phi()$ is the cumulative distribution function for the standard normal distribution (i.e., mean of zero and standard deviation of one). This likelihood is a product of two terms corresponding to the two independent, observed data (i.e., updates) of 19 years and 3/8 years.

3.5.2.4. Detailed example: updating dispersion parameters: Unlike the other four parameters, ϕ_1 and ϕ_2 are not components of model E1. Rather, as mentioned earlier, each is a nuisance parameter. As such, we have not yet specified their uninformative priors. We do know that $0 < \phi_1$ and $0 < \phi_2$. Gamma and inverse gamma distributions have this support and are often employed as priors for dispersion parameters such as ϕ_1 and ϕ_2 (Congdon, 2006, p. 69). Consider a gamma distribution in parameters a and b of expectation $\frac{a}{b}$ and variance $\frac{a}{b^2}$.

^{xviii}: Why use the midpoint? One rationale is that 19 years is an uninformative expectation, being that of a uniform distribution on interval (13, 25).

We chose uninformative gamma priors $g_0(\varphi_1)$ and $g_0(\varphi_2)$ with $a = 1$ and $b = 1/10$. These priors are suitably vague because each prior's variance of 100 is an order of magnitude larger than its expectation of 10.

Have we overlooked likelihoods for σ^2 , φ_1 and φ_2 ? No; and to understand why not, recall E2, which we re-write here as a proportionality.

$$g_1(\omega|\mathbf{y}) \propto f(\mathbf{y}|\omega)g_0(\omega). \quad [\text{E10}]$$

Recall also that ω is the *vector* containing all parameters, $\omega = \{\beta_0, \beta_1, \delta, \sigma^2, \varphi_1, \varphi_2\}$, which defines $g_0(\omega)$ as the joint prior for all parameters, $f(\mathbf{y}|\omega)$ as the joint likelihood for all parameters, and $g_1(\omega|\mathbf{y})$ as the joint posterior for all parameters. One approach is to formulate $g_0(\omega)$ as the product of the separate priors.

$$g_0(\omega) = g_0(\beta_0)g_0(\beta_1)g_0(\delta)g_0(\sigma^2)g_0(\varphi_1)g_0(\varphi_2). \quad [\text{E11}]$$

By writing this as a product of the separate (i.e., univariate) priors, we are assuming that the parameters are distributed independently of one another (Congdon, 2006, p. 69). This is commonly done but is by no means necessary. For example, we could have formulated a joint multivariate normal prior for β_0 and β_1 . However, that would require introduction of another nuisance parameter^{xix} to define the correlation between β_0 and β_1 . While the product formulation for the joint prior may be chosen for convenience, a product formulation of the individual likelihoods

$$f(\mathbf{y}|\omega) = f_1(\hat{\beta}_0|\{\beta_0, \sigma^2\}) f_2(\hat{\beta}_1|\{\beta_1, \phi_1\}) f_3(\hat{\delta}|\{\delta, \phi_2\}) \quad [\text{E12}]$$

is the indisputably correct choice given that each likelihood is based on a separate sample of data and these separate samples are from mutually independent sources. Importantly, unlike the joint prior, each of these likelihoods potentially updates two parameters. E6 contains β_0 and σ^2 , E8 contains β_1 and φ_1 , and E9 contains δ and φ_2 .

3.5.2.5. Detailed example: updating dispersion parameters—standard errors provided:

We say *potentially* because updates can only happen if data are provided that further inform what we know about the parameter. Parameters σ^2 , φ_1 and φ_2 are the variance parameters from normal distributions. Unlike many distributions (e.g., gamma), the variance is not a function of the mean for the normal distribution. All of our data updates (i.e., E6, E8 and E9) only provide data on the mean; and thus do not inform estimates of the variance parameters. In section 4, we provide a final update with some source data and this does provide information about the variance parameter σ^2 . In contrast, we never explicitly update φ_1 and φ_2 ; but then these are not parameters of model E1; instead they are nuisance parameters as described in Section 3.1.4. To place this in perspective, the worked example

^{xix}Another option would be to factor the joint prior by conditioning as $g_0(\beta_0, \beta_1) = g_0(\beta_0|\beta_1)g_0(\beta_1)$. For example, see Congdon (2006, p. 70). For an introduction to conditional probability, see Ross (1998, Chapter 3).

of section 3.5 may be somewhat unusual in that so little information was available to update variance parameters (σ^2 , φ_1 and φ_2); and this is because, somewhat remarkably, none of the parameter estimates from the literature employed here were reported with estimated standard errors or sample standard deviations.^{xx}

In most applications, at least some reported parameter estimates will be accompanied by a standard error or standard deviation. This is valuable information about the reliability of an estimate. As a simple example, suppose a sample mean of sample size n was reported in the literature that could serve as an estimate of β_0 ; and the standard error of the mean $\frac{\sigma^2}{n}$ was also reported. If sampling was from a normal distribution, the ratio $\frac{\beta_0 - \beta_0}{\sigma^2/n}$, which incorporates information about the mean as well as information about the variance, could be updated with a likelihood for a Student's t distribution on $n - 1$ degrees of freedom (Rice, 1995, pp. 177–182). All else equal, parameter estimates based on larger sample sizes are more stable and thus more informative. As sample size (and degrees of freedom) increases, Student's t distributions become narrower and more concentrated about the mean. This results in a narrower posterior distribution. In a sense, updating with this likelihood performs another form of differential reliability weighting (Section 3.4), with weights being a function of sample size.

3.5.3. Detailed example: differential weighting of literature updates—Step 4

entails weighting the update process. For the primary analyses of Section 5, we set

$w_i \stackrel{\text{def}}{=} w \stackrel{\text{def}}{=} 1$, for $j = 1, 2, \dots, m$; but we vary the w_i in some additional analyses in Section 5.3. Ibrahim and Chen (2000) employed the restriction $0 \leq w_i \leq 1$. We relax this to $0 \leq w_i$ to permit greater flexibility in weighting the data set of Douek et al. (2001) relative to other information from the literature. Weights greater than one cause a result from the literature to be weighted more heavily than the source data.

3.5.4. Detailed example: updating with source data—With Steps 1 through 4

completed, we are now ready to update using source data and obtain our final estimate of the posterior distribution for each parameter of model E1. The source data permits updates of all four parameters of interest $\{\beta_0, \beta_1, \delta, \sigma^2\}$. Specifically, the likelihood for these data is

$$f_d(\mathbf{z} | \{\beta_0, \beta_1, \delta, \sigma^2\}) = \prod_{\mathbf{z}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(\ln y_i - (\beta_0 + I[\delta < a]\beta_1(a - \delta)))^2}{2\sigma^2} \right) \quad [\text{E13}]$$

where $\mathbf{z} = \{\ln y_1, \ln y_2, \dots, \ln y_{33}\}$ is the vector of 33 observations in the sample from Douek et al. (2001). Unlike f_1, f_2 , and f_3 that were products in no more than two terms, f_d is a product in 33 terms; although each term is based on a single datum. Combining all four steps, we can write

^{xx}What can be done, however, is to differentially weight based on measures of precision of measurement, as described in Section 3.5.3 and illustrated in Section 5.

$$g_1(\boldsymbol{\omega}|\mathbf{y}) \propto f_1(\hat{\beta}_0|\{\beta_0, \sigma^2\})f_2(\hat{\beta}_1|\{\beta_1, \phi_1\})f_3(\hat{\delta}|\{\delta, \phi_2\})f_d(\mathbf{z}|\{\beta_0, \beta_1, \delta, \sigma^2\}) \times g_0(\beta_0)g_0(\beta_1)g_0(\delta)g_0(\sigma^2)g_0(\phi_1)g_0(\phi_2). \quad \text{[E14]}$$

4. Materials and Methods

4.1. Sampling the multivariate posterior

Application of Bayes Theorem (E2) through analytic (mathematical) calculations is intractable for all but the simplest problems (e.g., Figure 1). Often particularly intractable is the integral of the denominator of E2. However, even with the proportionality representation (E3), the full mathematical expression of the unnormalized posterior (i.e., the product of the likelihoods and the priors) is complex. This would quickly become evident if one tried to write out the full expression for the details of the likelihoods and priors on the right side of E14. A much more practicable alternative is to employ numerical methods. Among the various options available, we prefer Markov chain Monte Carlo (MCMC) via the Metropolis algorithm because it can be employed for such a wide variety of problems. We discuss an alternative approach in Section 6.6.

The overall approach, detailed in Kruschke (2011) and elsewhere, is to numerically sample from many locations on the unnormalized posterior to approximate the full posterior distribution. This requires the ability to evaluate the unnormalized posterior at any location, which is enabled by the fact that we have specified all of the components on the right-hand side of E14. Locations are sampled randomly in sequence; and this process of random sampling from a particular model (here the unnormalized posterior) is an example of a Monte Carlo method (Dodge, 2003, p. 266). At each step in this sequence, the Metropolis algorithm currently resides at a particular location on the unnormalized posterior. The next location is obtained via a proposal distribution. If that next location evaluates to a higher value, the algorithm always moves to that location; otherwise the algorithm only moves to that location with probability less than one. If the algorithm does not move to that new location, an alternative new location is randomly selected by the proposal distribution, that new location is evaluated, and the process continues. This process of evaluating the next possible location solely based on the current location (and not earlier locations in the sequence) is the *Markov property*. In the long run, after many moves, an accurate representation of the unnormalized posterior distribution emerges. The number of moves required before sampling becomes representative of the unnormalized posterior is known as the *burn-in* period; so burn-in samples are often discarded. Even after the burn-in period, successive samples that are close together in the sampling sequence may be highly correlated and not representative of something closely approximating independently sampled observations from the unnormalized posterior. That is why, after burn-in, it can be helpful to *thin* the sampling by only retaining every m^{th} sample value, where m may be in the 10's, 100's or even the 1000's, the choice of which depends on how slowly the proposal distribution moves the algorithm over the unnormalized posterior. So, clearly, choice of the burn-in period, proposal distribution, and thinning rate are among the important factors to consider when tuning the algorithm to obtain "good mixing," which is the state where the retained sequence of sampled values provide very close approximation to a sequence of independent observations from the unnormalized posterior.

4.2. Markov chain Monte Carlo implementation

For estimating the parameters of model E1, we implemented Metropolis MCMC in SAS® v. 9.3 and v 9.4 (SAS® Institute, Cary, NC). We provide our code in the online supplement; and that code includes many comments to assist the user. We used a wide Student's-*t* (3 degrees of freedom) proposal distribution to help move MCMC sampling more rapidly throughout the posterior. Good mixing was achieved using 75,000 burn-in iterations followed by 10^6 sampling iterations thinned at a rate of 1/1000, giving a retained total sample size of 1,000 observations on the unnormalized posterior (Figure A.2 of online supplement). Good mixing is also evident from the fact that correlation among sampled values was indistinguishable from zero (Table A.1 of online supplement), indicating that each sampled point was providing approximately independent information about the unnormalized posterior distribution. A Geweke test (Geweke, 1992) failed to detect a change in mean estimates of parameters over the post-burn-in period ($p > 0.16$). Finally, three different samplings (chains) were run, each with different starting values for the parameters and for the seed of the pseudo-random number generator. All three chains converged on very similar posterior estimates for each parameter (Table A.2 of online supplement).

The Stanford University Research Compliance Office Institutional Review Board determined that this research does not meet the requirements of human subjects research because it did not entail “obtaining or receiving private individually identifiable information.”

5. Results

5.1. Parameter estimates

Estimates of the model's parameters are given in the top four rows of Table 1. Table 1 provides two statistics from the sampled ($n = 1,000$) unnormalized posterior distribution: the mean of the posterior for each parameter and 95% credible intervals for each parameter. The particular type of credible interval reported is a highest density interval (HDI). Credible intervals are interpreted differently than confidence intervals, where the latter are frequentist, not Bayesian. A 95% HDI spans that portion of the posterior that represents 95% of the values in which we have strongest belief (Kruschke, 2011, p. 40). The mean quantity of TREC-CD4NT at birth is estimated to be ~ 19,167, which expressed as a percentage is ~ 1–2.7%, with a credible interval of 16389 to 22538 or 10.9% to 15%. Our belief is high (95%) that % TREC-CD4NT at birth is within this narrow range. Mean age of onset of decline in TREC-CD4NT is estimated at ~ 21.5 years of age but has a relatively wide 95% credible interval of 13.2 to 30.1 years. Once begun, rate of decline is estimated to be pronounced. Accounting for the transformed scale of parameters' estimates (i.e., the natural logarithm) we have $y = e^{\beta_0 + \beta_1(a - 126\delta)}$. From this, MCMC simulations estimate that, after onset, percentage decline (95% credible interval) in the first year is ~ 5.3% (3.8%, 6.7%) and during the first 10 years is ~ 41.7% (32.3%, 50.0%).

5.2. Smoothed estimate of fitted curve

Model E1 is probably unrealistic in the sense that it has a sharp corner at age δ years. To obtain a more realistic depiction of this developmental process, we employed posterior model averaging (Hartemink et al., 2002). Specifically, each of the 1,000 draws from the posterior consists of a set of four values corresponding to the four model parameters $\{\beta_0, \beta_1, \delta, \sigma^2\}$. Each such set of four values defines a particular piecewise linear function that starts at a particular value of TREC-CD4NT at birth, remains there until a specific age, and then declines at particular rate. Each draw also has associated with it a posterior density value, which serves as a measure of our strength of belief in the particular *set* of values of the parameters in that draw. This allows us to calculate a *belief-weighted* average of each of these 1,000 piecewise linear functions; and that weighted average is the smooth solid line of Figure 2. This smooth line is, in effect, what is most plausible based on our current beliefs. Also included in Figure 2, for reference, is the estimate of the rate of decline in TREC-CD4NT adapted from Hug et al. (2003), shown by the dashed curve, as well as the point estimates of TREC-CD4NT at birth from Schönland et al. (2003) and the age of onset from Zhang et al. (1999) and Weerkamp et al. (2005). The solid circles are the observed data from Douek et al. (2001). Additional methods of assessing model fit are discussed in Section 6.4.

5.3. Differential weighting of literature sources

These estimates do contain uncertainty of two types. That uncertainty is reflected in part within the credible intervals of Table 1; however, these intervals do not capture all uncertainty because they are conditional upon the particular weights assigned to the various published values used in forming the overall estimate of each parameter. As discussed in Section 3, we initially chose to weight each data source equally. Equal weighting presumes that we have no reason to consider one data source as providing more reliable or relevant information about a parameter than another.

To illustrate an alternative, and perhaps equally valid, weighting scheme, consider estimation of β_0 . Expressed as a percentage, the estimates of this parameter from Douek et al. (2001) and Schönland et al. (2003) are 8.4% and 13.4%, respectively. This difference is doubtless partly due to sampling error (limited sample size) but also may be due to improvements in assay methods. Suppose we fit the model again but weigh the more recent estimate more heavily. Specifically, the middle four rows of Table 1 show estimates from the model if we apply power prior coefficients that are proportional to the reciprocal of years before present: 13/13 for Douek et al. (2001) and 13/11 for Schönland et al. (2003). Comparing this result to that under equal weighting (top four rows of Table 1), recency-based differential weighting shifted estimates only slightly.

We can also differentially weight the different sources of information on mean age of onset. The results of Zhang et al. (1999) suggested that decline occurred “Between the teen years and age 20–25...,” a wide interval. The data of Douek et al. (2001) are provided in units of years. Weerkamp et al. (2005) observed levels of T-cell precursors (CD34⁺CD1a⁺ thymocytes) to be highest in children of ages 3 to 6 months. In each of these three cases, we have an estimate that falls somewhere within an interval but we do not know exactly where within that interval (i.e., interval censored). In fact, to our knowledge, all locations within

the interval are equally plausible. Using the reliability-based weights of Section 3.4, we obtain $w_1 = \sqrt{12}/(25-13) = 1/\sqrt{12}$, $w_2 = \sqrt{12}/(l+1-l) = \sqrt{12}$, and $w_3 = \sqrt{12}/(1/2-1/4) = 8\sqrt{3}$ for Zhang et al. (1999), Douek et al. (2001), and Weerkamp et al. (2005), respectively. Relative weights do not change if all weights are multiplied by a constant. This can simplify calculation by setting at least one weight equal to unity. Dividing by $\sqrt{12}$ gives $w_1 = 1/12$, $w_2 = 1$, $w_3 = 4$. A pronounced reduction in the age in onset of decline in TREC-CD4NT results, from 21.5 years without differential weighting (top four rows of Table 1) to 7.27 years with reliability-based weighting (bottom four rows of Table 1), due to the relatively fine-grained scale of measurement in Weerkamp et al. (2005). In their model of thymopoiesis, Ye and Krischner (2002) estimated peak production of RTEs to be at 1 year of age; and Bains et al. (2009b) also report peak thymic export at approximately 1 year of age. Of course, thymic production and export are only one component of TREC-CD4NT. Inclusion of Weerkamp et al. (2005) presumes that the developmental timing of peak cellularity demarcates the age of onset in decline in TREC-CD4NT. All that said, even with reliability-based weighting, the HDI on age of onset in decline in TREC-CD4NT still encompasses 21.5 years of age. The estimated rate of decline following onset ($\hat{\beta}_1$) is attenuated to -0.0386 when the estimated age of onset is earlier (bottom four rows of Table 1). That is, for decline in TREC-CD4NT, onset is estimated to start earlier and then proceed more slowly with age.

6. Discussion

6.1. Immunological implications of results

We illustrate the use of Bayesian methods to fit a developmental model of peripheral TREC-CD4NT and we informed parameter estimation for that model using published findings plus source data. This example application estimates that TREC-CD4NT begins to decline by roughly 21.5 years of age and that decline thereafter may be rapid, such that TREC-CD4NT is reduced to half of birth level by about 34 years of age. These findings agree approximately with the assumption of Bains et al. (2009a) that TREC-CD4NT remains constant over the first 20 years of life. Even when we remove the timing of peak cellularity (Weerkamp et al., 2005), the estimated mean age of onset is roughly 22 years, which remains close to 21.5 years of age. Our results also agree with the theoretical work of Murray et al. (2003) who found the principal source of naïve T cells to shift from the thymus to the periphery at age 20 years. In contrast, we estimate that TREC-CD4NT will be higher at 16.9% at age 55 and they estimate output at this age to be somewhat lower at 5%. Based on these results, decline in TREC-CD4NT appears to lag behind decline in thymic mass, which appears to peak around ten to twelve years of age (Ogiu et al, 1997). When we apply reliability-based weighting (Section 5.3), the estimate of the age of onset of decline in TREC-CD4NT shifts to approximately the seventh year of life; although, as we state in Section 5.3, the 95% HDI on this estimate still includes 21.5 years of age. That is, even with accounting for variable reliability of data sources, a mean age of onset of decline of ~ 21.5 years of age remains within plausible range. Moreover, as discussed in Section 5.2, use of posterior model averaging produces a plausible description of onset of decline that is not sudden at age 21.5 but rather a gradual process centered around this age (solid line of Figure 2).

Clearly our results are not the end of inquiry on this topic; nor should they be. Other reports may be published that provide additional information on parameters of model E1 and those data could be used as additional updates to stabilize estimates even further, allowing investigators to pinpoint even more precisely the age at which decline in TREC-CD4NT begins. This in turn facilitates much more focused study. For example, as credible intervals continue to narrow on estimates of age of onset, investigators can transition from asking, “When does decline begin?” to asking, “Why does decline begin in early adulthood?” Importantly, this transition may be possible without performance of any new study that is specifically designed to estimate age of onset of decline in TREC-CD4NT, for the reasons given in section 6.2. Also, what our basic model (E1) provides is a description of the timing of major lifetime onset of decline in TREC-CD4NT. More fine-grained modeling could elucidate details of age-related trends within younger and older populations. Research will also continue to be needed to obtain estimates of onset of decline for specific medical subpopulations which may have divergent levels of TREC-CD4NT (e.g., Figure 4 of Sandgaard et al., 2014). Estimates for specific medical subpopulations could use informative priors that are, at least in part, updated using results from the more general medical population. See Section 6.2.

6.2. Advantages of Bayesian approach

The Bayesian estimation framework permits rigorous quantitative synthesis of several studies and therefore has the potential to yield more reliable estimates than studies based on single, small samples. *Of particular utility is that each study need not provide information on all model parameters.* For example, Schönland et al. (2003) only provided information on β_0 , Zhang et al. (1999) only provided information on, Hug et al. (2003) only provided information on β_1 , etc. These were combined with the information provided on all parameters from the sample data of Douek et al. (2001) to yield an integrated set of estimates. Despite being drawn from multiple sources, estimation is a synthesis, in that each source of information on one parameter impacts estimation of the other parameters. One implication of this is that Bayesian techniques can be used to fit models of highly complex immunological systems that contain more parameters than can be measured by any single study.

Though we introduce Bayes Theorem for continuous distributions, the method of Sections 3.2 through 3.4 are equally applicable to discrete distributions. A discrete version of the Theorem is given below in E15. Admixtures are possible as well. The example of Section 2.2 employed a continuous uniform distribution (a type of beta distribution) for the prior and a discrete distribution (binomial) for the likelihood.

Not every parameter estimate drawn from the literature for this purpose is a perfect match to the problem at hand. We addressed some of those possible limitations in Section 5 through differential weighting of data sources based on recency of publication and measurement precision. Also, bias is a concern. For instance, Weerkamp et al. (2005) measured precursors to TREC-CD4NT; and so naturally we expect that these estimates will err on being too early in age. That said, using these estimates does move us from a state of essentially no knowledge (onset being anywhere from birth to age 126 years) to a position closer to where

age of onset is believable; or, more precisely, our distribution of what ages are believable is made more focused by the estimate from Weerkamp et al. (2005), even if that focus is not exactly where onset is most believable. This is an example of a *bias-variance trade-off*; we reduce the variance or spread in believable values for age of onset but that comes at the cost of being misled slightly (bias). The important point being that not every estimate taken from the literature needs to be perfectly matched to the parameter of interest. An approximate surrogate may also be helpful. Even so, this estimate from Weerkamp et al. (2005) could be dropped for final estimation of Model E1's parameters because, in addition to introducing bias, it also widens the 95% HDI for δ (data not shown).

Another bias-variance trade-off is in operation as well. Biological systems are inherently complex and a correspondingly complex model may permit a more realistic (i.e., less biased) representation of the system being studied. However, that greater complexity often requires a richer parameterization; but parameter estimates from these richer models are typically less stable. Therefore the mathematical model employed needs to be parsimonious, limited to the essential features. For example, here parameterization was limited to where development begins (β_0), when it begins to change (δ), and how rapidly it declines thereafter (β_1). Though parsimonious, this model is nonlinear due to the presence of the parameter product $\beta_1 I[\delta < a](a - \delta)$, as mentioned in Section 2. Nonlinearity did not pose any difficulties for the Bayesian fitting algorithm. Good mixing of the Markov chain Monte Carlo algorithm was obtained for all parameter estimates (online supplement Figure A.2). That this method can accommodate nonlinear models means that it offers wide latitude for estimating parameters of immunological systems. Demonstrated in Section 5 was how even an approximate model such as E1, with a probably unrealistically sharp corner at age of onset, can be smoothed, using posterior averaging, to generate a potentially more realistic depiction of developmental decline of TREC-CD4NT. Still, as mentioned in Section 6.1, models of finer-grained changes in TREC-CD4NT with age could be explored and this would likely require a richer parameterization.

6.3. Relaxing parametric assumptions

The robustness of Bayesian estimates can be explored further. Historically, a criticism of Bayesian methods derives from their fully parametric nature. For example, specific standard parametric distributions (e.g., normal) are typically used to specify priors and likelihoods. That said, the menu of parametric distributions is extensive and continues to grow. Also, software applications permit users to specify parametric distributions that are not “built in.” Even so, situations can certainly arise where standard parametric distributions will not suffice. To address this, *mixtures* of distributions can prove useful. Mixtures can be of a finite or an effectively infinite quantity of component distributions; and identification, the ability to estimate model parameters, may require use of constraints or initial empirical updates of priors (Congdon, 2006, pp. 187–217). See Costa et al. (2011) and Lin et al. (2013) for recent applications. Another approach, applicable to models in fewer parameters, is grid approximation (Section 6.6) in which the prior (Kruschke 2011, p. 105) and, even more importantly for the method of the present paper, the likelihood do not need to be discrete versions of any particular parametric distribution (i.e., *distribution-free*).

6.4. Model development

As defined in Section 3.1, model formulation is the first step in which biological concepts are formalized as mathematical constructs, whereas model development is the more encompassing, iterative process of formulation, testing, revision of formulation, etc. (Motta and Pappalardo, 2013). As illustrated in Section 2.1, model development should be governed by immunological theory. Section 3.1 detailed the first critical step of model development—formulation. In Sections 3.1 and 6.2, parsimony in model parameterization was emphasized to facilitate stable parameter estimation as well as model interpretability.

A purely theory-driven approach to model development may only allow the investigator to winnow options down to a *set* of plausible models. Congdon (2006) discusses various options for *empirical* (i.e., data-driven) model selection (selecting one model from the set) or for averaging competing models. Section 5.2.1 describes assessing model fit by a form of posterior averaging. Fit may also be assessed by sampling from the posterior predictive distribution. This entails simulating new observations from the posterior fit of the model and comparing sets of simulated posterior data $\{y^*\}_i$ against original data y_i for the i^{th} observation. Simulation is straightforward to implement (e.g., `preddist` statement in `proc mcmc` of SAS® v.9.4, SAS Institute, Cary, NC). A useful measure of fit in terms of bias and stability, due to Laud and Ibrahim (1995), is $C \stackrel{\text{def}}{=} (E[y_i^*|y_i] - y_i)^2 + \text{Var}[y_i^*|y_i]$ (Congdon, 2006, p. 47), where $E[y_i^*|y_i]$ and $\text{Var}[y_i^*|y_i]$ are the respective mean and variance of posterior observations corresponding to the i^{th} original observation. Smaller values of C indicate better fit. One approach is to calculate C while varying choice of uninformative prior, choice of weights for different data sources for the informative prior, choice of likelihood function and even functional form of the model. The bias term $(E[y_i^*|y_i] - y_i)^2$ can be explored further through statistical agreement analyses (e.g., intraclass correlation coefficient) of estimates of $E[y_i^*|y_i]$ versus y_i . Of course, model fit is not a purely statistical issue: fit is poor if posterior parameter estimates are biologically implausible, especially if the HDIs do not contain any parameter values that are biologically plausible.

When model fit is poor, three possible causes are the data sources for the informative priors, choice of parametric distributions for likelihood and/or prior, and model formulation.

1. If an informative prior is the result of likelihood updates from multiple sources of information, some alternative differential weighting schemes should be explored (Section 3.4). Also, further research may be needed on additional sources of information for generating the informative prior. Studies on otherwise unrelated topics may provide useful sources of information in their control groups (e.g., Zhang et al., 1999).
2. If choice of parametric distribution may be limiting, generalized cross entropy and generalized maximum entropy, which are semiparametric (Section 6.7.2), could be used to update the prior.
3. The current parameterization of the model may not be sufficiently flexible. Any unnecessary constraints on existing parameters should be lifted. Parameters can be added to increase model flexibility but, ideally, only if those parameters have

biological interpretation; otherwise nuisance parameters are introduced (Section 3.1.4). If different formulations are examined, one can compare fits of one model against another using Bayes factors, as detailed in Kruschke (2011).

6.5. Other applications

In this paper, we provide two example applications of Bayesian estimation, a brief conceptual example in Section 2.2 and a very detailed example in Sections 3 through 5. However, these examples are far from all-encompassing of the types of data sets to be encountered by working immunologists. In this section, we discuss three other broad classes of data structures: very small, longitudinal and wide data.

Very small data sets—The method of Sections 3.1 through 3.4, which permits integration of data across studies, may be particularly important to apply to areas of investigation where sample size per study will, by necessity, always be quite small. One prominent example is the immunology of the developing fetus (Vermijlen et al., 2010), including thymic development, wherein any form of invasive *in utero* sampling (e.g., of blood via cordocentesis) carries especially high risks to research participants. More broadly, medical studies in pediatric populations tend to have small sample sizes. Recently, interest has been emerging in the use of Bayesian methods in this setting, employing adult data in combination with data from children to boost statistical power in these small samples (Schoenfeld et al., 2009). The work that we present in this manuscript does that too, although in a more biologically plausible way. Schoenfeld et al. (2009) use adult data for rough prior approximation of pediatric parameters. The method outlined in sections 2, 3 and 4 above does not use adult data as a rough approximation to conditions within children but instead models the process of development from childhood into adulthood, using adult data as biologically *complementary information*; and we do so using the special utility of Bayesian methods. The work of Schoenfeld et al. (2009) is to be commended for addressing an important practical problem in research. The current method takes that same idea a step further by building the Bayesian estimation process on a specific biological model. When an investigator's source data set is especially small, an extra thorough evaluation should be made of different, plausible, *objective* weighting schemes (Section 3.4) to ensure that findings based on the resultant parameter estimates are robust. Especially in very small samples, only covariates of low measurement error should be employed for covariate-based weights (Section 3.4).

Longitudinal data sets—This structure arises whenever repeated measurements are obtained on individuals over time (e.g., see Ringhoffer et al., 2013, for an application to naïve T-cells of thymic origin), a design that is used widely to characterize the temporal kinetics of immune features. The complexity added by this design is that measurements made on the same individual are often correlated. As such, this requires that model formulation (Section 3.1.5 in particular) accounts for the correlation structure. A direct approach is to specify a correlation matrix for the residuals where that matrix is populated with the correlation parameter(s). An uninformative prior (e.g., inverse Wishart) and likelihood are then specified for this matrix (Daniels and Pourahmadi, 2002). An alternative

or complementary approach is hierarchical Bayesian analysis, which is described in Spiegelhalter et al. (2004).

Wide data sets—Wide data structures arise when the quantity of immune features measured exceeds the sample size. These are now commonplace in immunological applications due to widespread use of high-dimensional assays (e.g., cytokine profiling). This structure is exaggerated in “pilot” studies (i.e., small samples) that also collect data on non-immunological features (e.g., Alcaide et al. 2013). One widespread approach to the analysis of wide data sets is dimension reduction; and one popular form of dimension reduction is feature selection. Feature selection can be accomplished in a Bayesian estimation setting through use of a “spike and slab” prior. This prior is a mixture of an informative prior (the spike) and an uninformative prior (the slab). The informative prior is often centered on zero and designed to shrink estimates of parameters for features that are purely noise to zero, thereby accomplishing feature selection. See Ishwaran and Rao (2005) for details. This type of analysis could be useful early in model development (Section 6.4) for generating a parsimonious formulation (Section 3.1) which could then be used in subsequent analyses (Section 3.3).

6.6. Alternative method for characterizing the posterior

Sections 4.1 and 4.2 detailed one particular means of characterizing the posterior. A number of other approaches have been developed; but one that avoids MCMC and its attendant tuning altogether and may thereby be computationally useful to the working immunologist is *grid approximation* (Kruschke 2011, pp. 101–115). For this technique, continuous distributions of priors and likelihoods are approximated by fine grids. That is, each distribution is made discrete by breaking the support of the variable into many intervals of equal width and assigning two values to each interval: 1) the value of the variable (parameter) at the midpoint of the interval and 2) a probability that is the integral (or sum) of the probability distribution over that interval. Kruschke (2011, pp. 103–104) walks through this construction in more detail. Probabilities can be obtained using cumulative distribution functions available in SAS® (SAS Institute, Cary, NC), R (R Core Team, 2014), and elsewhere. A discrete version of Bayes Theorem can then be applied:

$$q_1(\omega|\mathbf{y}) = \frac{m(\mathbf{y}|\omega)q_0(\omega)}{\sum_{\Omega} m(\mathbf{y}|\omega)q_0(\omega)} \quad (\text{E15})$$

In E15, $q_1(\omega|\mathbf{y})$ is the discrete fine-grid approximation of the posterior distribution $g_1(\omega|\mathbf{y})$. Application is simplest if the same grid is used to approximate likelihood $m(\mathbf{y}|\omega)$ and prior $q_0(\omega)$. Typically, grid approximation is applied to models in a just a few parameters (e.g., $p = 1, 2$ or 3) because, if n is the quantity of intervals for each parameter, the quantity of cells formed by p parameters is n^p (Kruschke 2011, p. 118). For example, a fine grid with $n = 100$ requires a total grid of one million cells for three parameters. For models in few parameters, grid approximation is computationally much simpler than MCMC, in basic application requiring some lines of basic code to construct the grid, assign probabilities to the resultant cells, apply E15 and construct posterior HDIs on parameters. Also, one can tailor low-dimensional grid structures so they do not extend into physiologically impossible

combinations of values for the immune parameters. To accomplish that, an approximating grid would need to be trimmed to exclude those nonsense regions and then renormalized to ensure that probabilities across all cells again sum to one, separately for likelihood and prior. For models in two parameters, heatmap technology (e.g., see available functions and packages in R; R Core Team, 2014) can be used to generate especially clear visual displays of joint posterior distributions, which facilitates interpretation.

6.7. Alternatives to fully Bayesian estimation

6.7.1 Generalized cross entropy (GCE)—The foundations for GCE are information theoretic and, in particular, built upon Kullback and Leibler's (1951) concept of cross entropy $\mathfrak{C}$ as a quantitative measure of *discrepancy* between two distributions,

$$\mathfrak{C} \stackrel{\text{def}}{=} \sum_{j=1}^{\varpi} p_j \ln \left[\frac{p_j}{q_j} \right] \quad (\text{E16})$$

where $0 \leq q_j \leq 1$, $\sum_j q_j = 1$, and $0 \leq p_j \leq 1$, $\sum_j p_j = 1$, are the respective prior and posterior probabilities of the j^{th} event. In very broad terms, GCE estimation proceeds by minimizing $\mathfrak{C}$ with respect to the posterior distribution, the p_j , subject to constraints on the p_j imposed by the data (inputs $\mathbf{X}$ and outputs $\mathbf{Y}$), the model, and the prior—the q_j . In other words, in place of the likelihood, optimization (here a minimization) is of this constrained $\mathfrak{C}$, which generates a posterior distribution for each model parameter. Without data and model constraints, $\mathfrak{C}$ is minimized when $p_j = q_j$ for all j . As with Bayesian analysis, model and data can “pull” the posterior away from the prior distribution. Also as in Bayesian analysis, each parameter β_i , $i = 1, \dots, \kappa$, has its own prior and (marginal) posterior distribution; so, for a multi-parameter model ($1 < \kappa$), we write p_{ij} and q_{ij} . Following model formulation ($\mathbf{Y} \leftarrow h[\mathbf{X} | \boldsymbol{\beta}] + \mathbf{E}$, Section 3.1), analysis proceeds by specifying a discrete *support* $\{x_{ij}\}_{j=1}^{\varpi}$ for the i^{th} parameter, where the x_{ij} are often spaced at equal intervals and should span all feasible values of the i^{th} parameter. Each x_{ij} is then assigned respective prior q_{ij} and posterior p_{ij} probabilities so that β_i updates from $\tilde{\beta}_i = \sum_{j=1}^{\varpi} q_{ij} x_{ij}$ to $\hat{\beta}_i = \sum_{j=1}^{\varpi} p_{ij} x_{ij}$. If that prior is fully uninformative, $q_{ij} = 1/\varpi$ for all i (i.e., uniform mass distribution), which is commonly where the prior is initially set (analogous to Section 3.2). Model fitting also requires specification of a support and prior for the residuals $\mathbf{E}$ (Section 3.1.6). See Golan et al. (1996) for more details.

GCE is by design *semiparametric* in that modeling entails parameters but the analyst does not need to specify parametric distributions (e.g., normal distribution) for the outputs $\mathbf{Y}$ (and residuals $\mathbf{E}$; Section 3.1). This is in stark contrast to traditional Bayesian specifications, which largely rely on fully parametric formulations for priors and likelihoods (but see Sections 6.3 and 6.6). In fact, because, with GCE, supports for residuals $\mathbf{E}$ and parameters $\boldsymbol{\beta}$ are discrete, the analyst has complete freedom for specification of distributions.

However, this flexibility may come at a cost in that entropy-based parameter estimators may be quite sensitive to details of the specification of the support set $\{x_{ij}\}_{j=1}^{\varpi}$ (Léon et al.,

1999). Choice of support set for GCE is a topic that remains under development. We recommend that a support set should typically be dense (i.e., $20 < \varpi$). This makes GCE less approximate. Minimum and maximum values need to be biologically justified (e.g., physiological bounds); and all other support points can be spaced at regular intervals between these two outer bounds.

Informative priors for GCE can be constructed using the closely-related method of generalized maximum entropy (GME). We illustrate through an example. Suppose that one parameter of a model of development of the thymus is mass at birth. According to Leonidas (1998), thymic mass is “generally” 10 g to 15 g at birth and peaks at a maximum of ~ 40 g during human development. On this basis, a dense set of support values is 0.5, 1.5, 2.5, ..., 38.5, 39.5. The support of 0.5 is for the interval [0, 1), 1.5 is for the interval [1, 2), etc. Ogiu et al. (1997) provide means and standard deviations for thymic mass at birth of 12.1 ± 4.8 for males and 12.7 ± 4.9 for females from respective sample sizes of 50 and 63. Guikema (2007) suggests using the sample mean to construct a GME informative prior; but admits that this does little to stabilize the posterior estimate of the parameter. For that reason, we suggest a possibly novel approach using the sample mean $\hat{\beta}$ and its sample variance $\hat{\sigma}_{\hat{\beta}}^2$ to create an informative prior. Assuming a sex ratio at birth of 51.5% males (Parkes 1926),

$$\begin{aligned}\hat{\beta} &= 0.515 \times 12.1 + 0.485 \times 12.7, \\ \hat{\sigma}_{\hat{\beta}}^2 &= 0.515^2 \left(\frac{4.8^2}{50} \right) + 0.485^2 \left(\frac{4.9^2}{63} \right).\end{aligned}\quad (\text{E17})$$

Applying one-sided Chebyshev inequalities,

$$\begin{aligned}P\{\hat{\beta} \geq \beta + k\} &\leq \frac{\sigma_{\hat{\beta}}^2}{\sigma_{\hat{\beta}}^2 + k^2} \\ P\{\hat{\beta} \leq \beta - k\} &\leq \frac{\sigma_{\hat{\beta}}^2}{\sigma_{\hat{\beta}}^2 + k^2},\end{aligned}\quad (\text{E18})$$

for constant $0 < k$ and $0 \leq \sigma_{\hat{\beta}}^2 < \infty$ (Ross 1998, pp. 412–414). These very general inequalities might be useful for constructing informative priors because they permit incorporating information about the distribution of the random variable $\hat{\beta}$ *without any further assumptions* than those listed here. That is, we do not need to assume that, say, $\hat{\beta}$ has a normal sampling distribution as we typically would if updating was based on a parametric likelihood. Further, instead of thinking of E18 as probability statements about $\hat{\beta}$, we could treat these as *fiducial* (Stuart et al. 1999, pp. 156–162) statements of the relative plausibilities of different values of β , given the observed data $\hat{\beta}$ and $\hat{\sigma}_{\hat{\beta}}^2$.^{xxi} We obtain an informative prior by maximizing entropy $\mathfrak{E} \stackrel{\text{def}}{=} \sum_{j=1}^{\varpi} q_j \ln[q_j]$ with respect to the q_j , $j \in \{1, \dots, \varpi\}$, subject to constraints:

^{xxi}Fiducial theory is not without controversy. See Stuart et al. 1999.

$$\begin{aligned}
& \max_{q_j, j \in \{1, \dots, \varpi\}} \left\{ \sum_{j=1}^{\varpi} q_j \ln[q_j] \right\}, \text{ such that} \\
& 0 \leq q_j \leq 1, \quad \sum_{j=1}^{\varpi} q_j = 1, \\
& \sum_{j=1}^{\varpi} q_j x_j = \hat{\beta} \\
& \sum_{j=1}^{10} q_j \leq \frac{\hat{\sigma}_{\beta}^2}{\hat{\sigma}_{\beta}^2 + (12.1 - 10)^2}, \quad \sum_{j=16}^{40} q_j \leq \frac{\hat{\sigma}_{\beta}^2}{\hat{\sigma}_{\beta}^2 + (15 - 12.1)^2}
\end{aligned} \tag{E19a}$$

The first row of constraints, which we encountered above, ensures that the $\{q_j\}_{j=1}^{\varpi}$ constitute a proper mass function (i.e., sum to one). The second row constrains the mean of the prior to be equal to the sample mean. The last row imposes the one-sided Chebyshev fiducial

constraints per E18. In this last row, we have used the sample variance $\hat{\sigma}_{\beta}^2$ because σ_{β}^2 is unknown. Additional Chebyshev fiducial constraints can be imposed at other pairs of values (e.g., at 11/14 and 12/13 in addition to at 10/15).

$$\begin{aligned}
& \sum_{j=1}^{11} q_j \leq \frac{\hat{\sigma}_{\beta}^2}{\hat{\sigma}_{\beta}^2 + (12.4 - 11)^2}, \quad \sum_{j=15}^{40} q_j \leq \frac{\hat{\sigma}_{\beta}^2}{\hat{\sigma}_{\beta}^2 + (14 - 12.4)^2} \\
& \sum_{j=1}^{12} q_j \leq \frac{\hat{\sigma}_{\beta}^2}{\hat{\sigma}_{\beta}^2 + (12.4 - 12)^2}, \quad \sum_{j=14}^{40} q_j \leq \frac{\hat{\sigma}_{\beta}^2}{\hat{\sigma}_{\beta}^2 + (13 - 12.4)^2}
\end{aligned} \tag{E19b}$$

Adding constraints narrows (makes more informative) the prior but also boosts computational burden. The resultant informative prior for $\hat{\beta}$ is concentrated between 10 g and 15 g (Figure 3).

GCE has some disadvantages compared to the fully Bayesian approach. Due to its discrete nature, GCE is necessarily approximate in applications in which a continuous distribution (prior) may be applicable. GCE is not as well-known as Bayesian methods, especially in medical disciplines, which might pose some challenges in publication of findings. Also, unlike the wide array of software available for Bayesian analysis (SAS®, R, WinBUGS, etc.), to our knowledge only a single software application exists specifically for implementing GCE and GME: the proc entropy procedure is available in SAS®/ETS (SAS Institute, Cary, NC). That said, for the investigator who is willing to write custom code, extension to the nonlinear setting is possible in proc optmodel in SAS®/OR (SAS Institute, Cary, NC) and elsewhere. Code (proc optmodel) for performing E19 is available upon request from the first author (THH). We recommend bootstrapping (Efron and Tibshirani 1993) parameters' standard errors.

6.7.2 Mixed methods—Sections 3.2 and 3.3 described a fully Bayesian process for formulating an uninformative prior and using data to update that prior to an informative prior. In addition to a Bayesian approach, Guikema (2007) outlines some alternative non-Bayesian techniques for development of empirically informed priors: bootstrapped method of moments, bootstrapped maximum likelihood, and maximum entropy. Because Bayesian and non-Bayesian estimation are used in combination, these can be termed *mixed methods*.

Mixed methods may be helpful when Bayesian construction of an informative prior is biasing (Section 6.4). In particular, Guikema (2007) advocates for use of maximum entropy to construct an informative prior because, based on his work, this conservative approach tends to introduce the *least bias* into posterior estimates, among the various methods that he examined, including fully Bayesian. We illustrate construction of an informative prior using maximum entropy in Section 6.7.1, E19. A simpler version, and that which Guikema (2007) suggests, would drop the constraints derived from the Chebyshev inequalities.

6.8. A summary flowchart of the estimation process

Figure 4 provides a guided summary through the methods discussed in the current paper. This flowchart reads from upper left to lower right. Along the way, key decisions are highlighted and options provided for the analyst. In the center of this chart is the centerpiece of the method, empirical updating to generate informative priors. In the bottom lower right, if model fit is poor, reformulation is an option, which involves returning to the start of the process in the upper left corner. In our view, this type of workflow will become increasingly commonplace in immunological studies, because it allows each investigation to build upon existing work in a thoroughly quantitative and objective way.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by research award (R01 AI-083757 and R01 AI-100121 to DBL) from the NIAID/NIH. NIAID/NIH did advise on study design of the research award through their review process but, for this report, did not participate in collection, analysis or interpretation of data; writing of the report; nor the decision to submit this report for publication. THH and DBL wrote the manuscript. THH performed the statistical analyses. We thank Dr. Kellie Takagi for editorial assistance and constructive comments on various drafts; and we are grateful to an anonymous early reviewer for motivating derivation of model E1 as a solution to a differential equation.

References

1. Alcaide ML, Parmigiani A, Pallikkuth S, Roach M, Freguja R, Negra MD, Bolivar H, Fischl MA, Pahwa S. Immune activation in HIV-infected aging women on antiretrovirals-implications for age-associated comorbidities: A cross-sectional pilot study. PLoS ONE. 2013; 8:e63804. [PubMed: 23724003]
2. Bains I, Antia R, Callard R, Yates AJ. Quantifying the development of the peripheral naïve CD4+ T-cell pool in humans. Blood. 2009a; 113:5480–5487. [PubMed: 19179300]
3. Bains I, Thiébaud R, Yates AJ, Callard R. Quantifying thymic export: Combining models of naïve T cell proliferation and TCR excision circle dynamics gives an explicit measure of thymic output. J Immunol. 2009b; 183:4329–4336. [PubMed: 19734223]
4. Bolstad, WM. Introduction to Bayesian Statistics. John Wiley & Sons, Inc; Hoboken: 2007.
5. Chu JT. On the distribution of the sample median. Ann Math Stat. 1955; 26:112–116.
6. Congdon, P. Bayesian Statistical Modeling. 2. John Wiley & Sons, Ltd; The Atrium: 2006.
7. Costa IG, Roeder HG, do Rego TG, de Carvalho FDAT. Predicting gene expression in T cell differentiation from histone modifications and transcription factor binding affinities by linear mixture models. BMC Bioinform. 2011; 12(SUPPL 1)
8. Cox GM, Gibbons JM, Wood ATA, Craigmiles J, Ramsden SJ, Crout NMJ. Towards the systematic simplification of mechanistic models. Ecol Model. 2006; 198:240–246.

9. Daniels MJ, Pourahmadi M. Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika*. 2002; 89:553–566.
10. Dodge, Y. *The Oxford Dictionary of Statistical Terms*. Oxford Press; Oxford: 2003.
11. Douek DC, Betts MR, Hill BJ, Little SJ, Lempicki R, Metcalf JA, Casazza J, Yoder C, Adelsberger JW, Stevens RA, Baseler MW, Keiser P, Richman DD, Davey RT, Koup RA. Evidence for increased T cell turnover and decreased thymic output in HIV infection. *J Immunol*. 2001; 167:6663–6668. [PubMed: 11714838]
12. Dowling MR, Hodgkin PD. Modelling naïve T-cell homeostasis: consequences of heritable cellular lifespan during ageing. *Immunol Cell Biol*. 2009; 87:445–456. [PubMed: 19290017]
13. Edelstein-Keshet, L. *Mathematical models in biology*. Random House; New York: 1988.
14. Edwards, AWF. *Likelihood*. The Johns Hopkins University Press; Baltimore: 1992. expanded
15. Efron, B.; Tibshirani, RJ. *An Introduction to the Bootstrap*. Chapman & Hall; New York: 1993.
16. Flores KG, Li J, Sempowski GD, Haynes BF, Hale LP. Analysis of the human thymic perivascular space during aging. *J Clin Invest*. 1999; 104:1031–1039. [PubMed: 10525041]
17. Geweke, J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernardo, JM.; Berger, JO.; Dawid, AP.; Smith, AFM., editors. *Bayesian Statistics*. Vol. 4. Clarendon Press; Oxford: 1992.
18. Golan, A.; Judge, GG.; Miller, DJ. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. John Wiley & Sons; Chichester: 1996.
19. Groenwold RHH, Hak E, Hoes AW. Quantitative assessment of unobserved confounding is mandatory in nonrandomized intervention studies. *J Clin Epidemiol*. 2009; 62:22–28. [PubMed: 18619797]
20. Guikema SD. Formulating informative, data-based priors for failure probability estimation in reliability analysis. *Reliab Eng Syst Saf*. 2007; 92:490–502.
21. Haines CJ, Giffon TD, Lu L, Lu X, Tessier-Lavigne M, Ross DT, Lewis DB. Human CD4⁺ T cell recent thymic emigrants are identified by protein tyrosine kinase 7 and have reduced immune function. *J Exp Med*. 2009; 206:275–285. [PubMed: 19171767]
22. Harrell, FE, Jr. *Regression Modeling Strategies*. Springer; New York: 2001.
23. Hartemink AJ, Gifford DK, Jaakkola TS, Young RA. Combining location and expression data for principled discovery of genetic regulatory network models. *Pac Symp Biocomput*. 2002; 7:437–449. [PubMed: 11928497]
24. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2. Springer; New York: 2009.
25. Hazenberg MD, Verschuren MC, Hamann D, Miedema F, Dongen JJ. T cell molecular excision circles as markers for recent thymic emigrants: basic aspects, technical approach, and guidelines for interpretation. *J Mol Med*. 2001; 79:631–640. [PubMed: 11715066]
26. Holmes TH, Nichols DA, Thomander D, Kushida CA. A method for estimating normative distributions for study-specific populations of clinical trials. *Contemp Clin Trials*. 2012; 33:445–449. [PubMed: 22138103]
27. Hug A, Korporal M, Schroder I, Jurgen H, Glatz K, Storch-Hagenlocher B, Wildemann B. Thymic export function and T cell homeostasis in patients with relapsing remitting multiple sclerosis. *J Immunol*. 2003; 171:432–437. [PubMed: 12817027]
28. Ibrahim JG, Chen M. Power prior distributions for regression models. *Stat Sci*. 2000; 15:46–60.
29. Ishwaran H, Rao JS. Spike and slab variable selection: frequentist and Bayesian strategies. *Ann Statist*. 2005; 33:730–773.
30. Jackola DR, Hallgren HM. Dynamic phenotypic restructuring of the CD4 and CD8 T-cell subsets with age in healthy humans: a compartmental model analysis. *Mech Ageing Dev*. 1998; 105:241–264. [PubMed: 9862233]
31. Kilpatrick RD, Rickabaugh T, Hultin LE, Hultin P, Hausner MA, Detels R, Phair J, Jamieson BD. Homeostasis of the naïve CD4⁺ T-cell compartment during aging. *J Immunol*. 2008; 180:1499–1507. [PubMed: 18209045]
32. Kullback J, Leibler RA. On information and sufficiency. *Ann Math Stat*. 1951; 22:79–86.
33. Kruschke, JK. *Doing Bayesian Data Analysis*. Academic Press; Burlington: 2011.

34. Laud P, Ibrahim J. Predictive model selection. *J Roy Stat Soc Ser B.* 1995; 57:247–262.
35. Léon Y, Peeters L, Quinqu M, Surry Y. The use of maximum entropy to estimate input-output coefficients from regional farm accounting data. *J Agric Econ.* 1999; 50:425–439.
36. Leonard, T.; Hsu, JSJ. *Bayesian Methods: an Analysis for Statisticians and Interdisciplinary Researchers.* Cambridge University Press; Cambridge: 1999.
37. Leonidas JC. The thymus: from past misconception to present recognition. *Pediatr Radiol.* 1998; 28:275–282. [PubMed: 9569257]
38. Lin L, Chan C, Hadrup SR, Froesig TM, Wang Q, West M. Hierarchical Bayesian mixture modelling for antigen-specific T-cell subtyping in combinatorially encoded flow cytometry studies. *Stat Appl Genet Mol Biol.* 2013; 24:1–23.
39. Motta S, Pappalardo F. Mathematical modeling of biological systems. *Briefings Bioinform.* 2013; 14:411–422.
40. Montecino-Rodriguez E, Min H, Dorshkind K. Reevaluating current models of thymic involution. *Semin Immunol.* 2005; 17:356–361. [PubMed: 15967677]
41. Murray JM, Kaufmann GR, Hodgkin PD, Lewin SR, Kelleher AD, Davenport MP, Zaunders JJ. Naive T cells are maintained by thymic output in early ages but by proliferation without phenotypic change after age twenty. *Immun Cell Biol.* 2003; 81:487–495.
42. Neter, J.; Kutner, MH.; Nachtsheim, CJ.; Wasserman, W. *Applied linear statistical models.* 4. WCB McGraw-Hill; Boston: 1996.
43. Ogiu N, Nakamura Y, Ijiri I, Hiraiwa K, Ogiu T. A statistical analysis of the internal organ weights of normal Japanese people. *Health Phys.* 1997; 72:368–383. [PubMed: 9030838]
44. Parkes AS. The mammalian sex ratio. *Hum Reprod.* 1926; 2:1–51.
45. Pearl J. Causal inference in statistics: an overview. *Stat Surv.* 2009; 3:96–146.
46. R Core Team. *R: a language and environment for statistical computing.* R Foundation for Statistical Computing; Vienna, Austria: 2014. <http://www.R-project.org/>
47. Racine-Poon A, Wakefield J. Statistical methods for population pharmacokinetic modelling. *Stat Methods Med Res.* 1998; 7:63–84. [PubMed: 9533262]
48. Rietbergen C, Klugkist I, Janssen KJM, Moons KGM, Hoijtink HJA. Incorporation of historical data in the analysis of randomized therapeutic trials. *Contemp Clin Trials.* 2011; 32:848–855. [PubMed: 21729767]
49. Rice, JA. *Mathematical Statistics and Data Analysis.* 2. Duxbury: Press Belmont; 1995.
50. Ross, S. *A first course in probability.* 5. Prentice Hall; Upper Saddle River: 1998.
51. Ringhoffer S, Rojewski M, Döhner H, Bunjes D, Ringhoffer M. T-cell reconstitution after allogeneic stem cell transplantation: Assessment by measurement of the sjTREC/ β TREC ratio and thymic naive T cells. *Haematologica.* 2013; 98:1600–1608. [PubMed: 23585532]
52. Roberts MG. The dynamics of bovine tuberculosis in possum populations, and its eradication or control by culling or vaccination. *J Animal Ecol.* 1996; 65:451–464.
53. Sandgaard KS, Lewis J, Adams S, Klein N, Callard R. Antiretroviral therapy increases thymic output in children with HIV. *AIDS.* 2014; 28:209–214. [PubMed: 24072195]
54. Schoenfeld DA, Zheng H, Finkelstein DM. Bayesian design using adult data to augment pediatric trials. *Clin Trials.* 2009; 6:297–304. [PubMed: 19667026]
55. Schönland SO, Zimmer JK, Lopez-Benitez CM, Widmann T, Ramin KD, Goronzy JJ, Weyand CM. Homeostatic control of T-cell generation in neonates. *Blood.* 2003; 102:1428–1434. [PubMed: 12714521]
56. Scutari, M.; Strimmer, K. Introduction to graphical modeling. In: Stumpf, MPH.; Balding, DJ.; Girolami, M., editors. *Handbook of Statistical Systems Biology.* John Wiley & Sons, Ltd; Chichester: 2008.
57. Seaman JW, Seaman JW Jr, Stamey JD. Hidden dangers of specifying noninformative priors. *Am Stat.* 2012; 66:77–84.
58. Sklair-Levy M, Agid R, Sella T, Strauss-Liviatan N, Bar-Ziv J. Age-related changes in CT attenuation of the thymus in children. *Pediatr Radiol.* 2000; 30:566–569. [PubMed: 10993542]
59. Spiegelhalter, DJ.; Abrams, KR.; Myles, JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation.* John Wiley & Sons, Inc; Chichester: 2004.

60. Spiegelhalter, D.; Best, N.; Gilks, W.; Inskip, H. Hepatitis: a case study in MCMC methods. In: Gilks, W.; Richardson, S.; Spiegelhalter, D., editors. *Markov Chain Monte Carlo in Practice*. Chapman and Hall; London: 1996. p. 21-44.
61. Stoll, RR. *Set theory and logic*. W. H. Freeman; San Francisco: 1963.
62. Stuart, A.; Ord, JK. *Kendall's Advanced Theory of Statistics. Volume I, Distribution Theory*. 6. Arnold; London: 1994.
63. Stuart, A.; Ord, JK.; Arnold, S. *Kendall's Advanced Theory of Statistics, Volume 2A, Classical Inference and the Linear Statistical Model*. 6. Arnold; London: 1999.
64. Thomas, GB., Jr; Finney, RL. *Calculus and Analytic Geometry*. 9. Addison-Wesley Publishing Company; Reading: 1996.
65. Tufte, E. *The Visual Display of Quantitative Information*. 2. Graphics Press; Cheshire: 2001.
66. Vermijlen D, Brouwer M, Donner C, Liesnard C, Tackoen M, Van Rysselberge M, Nicolas T, Goldman M, Marchant A, Willems F. Human cytomegalovirus elicits fetal $\gamma\delta$ T cell responses in utero. *J Exp Med*. 2010; 207:807–821. [PubMed: 20368575]
67. Weerkamp F, de Haas EF, Naber BAE, Comans-Bitter WM, Bogers JJC, van Dongen JJM, Staal FJT. Age-related changes in the cellular composition of the thymus in children. *J Allergy Clin Immunol*. 2005; 115:834–840. [PubMed: 15806007]
68. Weon BM, Je JH. Theoretical estimation of maximum human lifespan. *Biogerontology*. 2009; 10:65–71. [PubMed: 18560989]
69. Ye P, Kirschner DE. Reevaluation of T cell receptor excision circles as a measure of human recent thymic emigrants. *J Immunol*. 2002; 168:4968–4979. [PubMed: 11994448]
70. Zhang L, Lewin SR, Markowitz M, Lin H, Skulsky E, Karanickolas R, He Y, Jin X, Tuttleton S, Vesanen M, Spiegel H, Kost R, van Luzen J, Stellbrink H, Wolinsky S, Borkowsky W, Palumbo P, Kostrikis LG, Ho DD. Measuring recent thymic emigrants in blood of normal and HIV-1 infected individuals before and after effective therapy. *J Exp Med*. 1999; 190:725–732. [PubMed: 10477556]

References

1. Congdon, P. *Bayesian Statistical Modeling*. 2. John Wiley & Sons, Ltd; The Atrium: 2006.
2. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc Ser B*. 1977; 39:1–38.
3. Jennrich RI. Asymptotic properties of non-linear least squares estimators. *Ann Math Statistics*. 1969; 40:633–643.
4. Kadane JB, Wolfson LJ. Experiences in elicitation. *J Roy Stat Soc Ser D: The Statistician*. 1998; 47:3–19.
5. Kruschke, JK. *Doing Bayesian Data Analysis*. Academic Press; Burlington: 2011.
6. Ross, S. *A first course in probability*. 5. Prentice Hall; Upper Saddle River: 1998.
7. Stuart, A.; Ord, JK.; Arnold, S. *Kendall's Advanced Theory of Statistics, Volume 2A, Classical Inference and the Linear Statistical Model*. 6. Arnold; London: 1999.
8. Tanner MA, Wong WH. The calculation of posterior distributions by data augmentation. *J Amer Stat Assoc*. 1987; 82:528–540.
9. Weon BM, Je JH. Theoretical estimation of maximum human lifespan. *Biogerontology*. 2009; 10:65–71. [PubMed: 18560989]

Highlights

- Instruction is provided on use of Bayesian methods to synthesize results from immunological literature.
- Theory, concepts and step-by-step methods are detailed for the working immunologist.
- Techniques are provided for weighting results based on objective measures of scientific strength.
- An example application to decline in thymic output of naive CD4 T cells is provided.
- Example's results suggest that rapid decline of thymic output initiates at ~ 21.5 years of age.
- Bayesian methods offer broad utility for estimating parameters of immunological models.

Fig. 1. Illustration of Bayesian estimation

A posterior distribution (- - - -) for parameter ω is obtained by updating the uniform prior distribution (—) for ω using the observed ratio of 19 events in $\eta=30$ trials. In this hypothetical example, ω is the probability that a naïve CD4 T-cell is of thymic origin and that η is the quantity of naïve CD4 T-cells in a 1- μ L aliquot. Curves show relative plausibilities (vertical axis) of different values of this probability ω before (—) and after (- - - -) updating with the data on 19 events (naïve T-cells).

Fig. 2. Graphical summary of fit of the regression model of TREC-CD4NT development (E1)
 A = estimate of β_0 from Schönland et al. (2003); B = estimate of δ from Zhang et al. (1999);
 C = estimate of δ from Weerkamp et al. (2005); --- = estimate of slope from Hug et al.
 (2003); ● = data of Douek et al. (2001); — = posterior-averaged Bayesian estimate of
 the developmental model. Data are reproduced from Figure 6A (p. 6667) of Douek *et al.*
 (2001). Copyright 2001. The American Association of Immunologists, Inc.

Fig. 3. Informative prior for thymic mass at birth derived using generalized entropy methods rather than Bayesian methods

Solid line (—) describes the relative plausibilities of different values for mean thymic mass at birth based on the published findings of Ogiu et al. (1997) and on ranges of values for individual thymic masses given in Leonidas (1998).

Fig. 4.
Flowchart for practitioners.

Table 1
Posterior estimates of model parameters and their 95% HDI credible intervals

β_0 , σ^2 , δ and β_1 are the perinatal mean, residual variance, age of the onset of decline, and rate of decline following onset, respectively. Parameter estimates for β_0 and δ are given on the back-transformed scale and are in units quantity of TREC-bearing cells per 150,000 naïve CD4 T cells and years, respectively. Parameter estimates for σ^2 and β_1 are provided on the natural-logarithm transformed scale because it is on this scale that homogeneity of variance and linear decline respectively apply. Section 5 of the text provides examples of estimated rates of decline on the back-transformed scale. Each parameter has 95% probability of occurring within the credible interval provided, with probability here representing a measure of strength of belief. Estimates are provided for three different power prior weighting schemes: equal, proportional to the reciprocals of years since publication, and as the reciprocals of measurement reliabilities. See Section 5.3 for details.

Power Prior Weights	Parameter	Estimate	95% Credible Interval
Equal	$\exp(\beta_0)$	19167	(16389, 22538)
	σ^2	0.3016	(0.1657, 0.4889)
	126δ	21.5	(13.2, 30.1)
	β_1	-0.0543	(-0.0696, -0.0394)
Proportional to the reciprocals of years since publication for β_0	$\exp(\beta_0)$	19181	(16293, 22017)
	σ^2	0.2946	(0.1496, 0.4475)
	126δ	21.6	(12.6, 29.1)
	β_1	-0.0548	(-0.0709, -0.0409)
Reciprocals of measurement reliability for δ	$\exp(\beta_0)$	20628	(16831, 25058)
	σ^2	0.3734	(0.1901, 0.5883)
	126δ	7.27	(0.01, 21.7)
	β_1	-0.0386	(-0.0583, -0.0263)