Journal of Immunological Methods xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Journal of Immunological Methods

journal homepage: www.elsevier.com/locate/jim

Computational modelling

Bayesian immunological model development from the literature: Example investigation of recent thymic emigrants

Tyson H. Holmes [a,*], David B. Lewis [b]

[a] Stanford Center for Sleep Sciences and Medicine, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, 401 Quarry Road, Stanford, CA 94305-5718, USA

[b] Department 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

ARTICLE INFO

Article history:
Received 18 September 2013
Received in revised form 16 June 2014
Accepted 21 August 2014
Available online xxxx

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

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.

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-Rodriquez 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 δRec-ψJα rearrangements of the TCR-α/δ gene locus; and TRECs may have long-term stability within these relatively long-lived cells (Hazenberg 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).

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

* Corresponding author. Tel.: +1 650 725 2925; fax: +1 650 725 7341.
E-mail addresses: tholmes@stanford.edu (T.H. Holmes), dblewis@stanford.edu (D.B. Lewis).

http://dx.doi.org/10.1016/j.jim.2014.08.001
0022-1759/

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) 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, footnotes are provided throughout the following text. 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 $\mu = \beta_0 \exp(\mathrm{I}[\delta < a]\beta_1(a - \delta))$, where the indicator function $\mathrm{I}[\delta < a] = 1$ 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 $Y = \beta_0 \exp(\mathrm{I}[\delta < a]\beta_1(a - \delta) + E)$. On the natural logarithmic scale,

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

Model E1 is illustrated schematically in Fig. 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 $\mathrm{I}[\delta < a]\beta_1(a - \delta)$. 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 $Z = \varsigma_0 + \varsigma_1 X$. In the statistical context, the spline term $\mathrm{I}[\delta < a]\beta_1(a - \delta)$ 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 population[2] using the simple model given by E1, as described in Section 4 and illustrated in Section 5.

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

[2] Because 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/).

Please cite this article as: Holmes, T.H., Lewis, D.B., Bayesian immunological model development from the literature: Example investigation of recent thymic emigrants, J. Immunol. Methods (2014), http://dx.doi.org/10.1016/j.jim.2014.08.001