16

Bayesian Modeling for Economic Evaluation Using "Real World Evidence"

Gianluca Baio

University College London, UK

This chapter presents an example of the use of Bayesian modeling with "real world evidence". This terminology indicates an increasingly popular body of evidence typically collected in the post-marketing framework and, usually, under observational conditions. Typical examples include population registries, cohort studies, or, more generally "electronic health records". We focus our attention to the case of statistical modeling for economic evaluation of an intervention on the basis of time-to-event outcomes, where the information provided by the available individual level data (e.g. from a Phase III experimental study) is immature and thus benefits from integration with external (population level) sources of information. We present a case study and discuss the advantages of the Bayesian approach.

16.1 Introduction

Generally speaking, we refer to "post-marketing" as the stage of pharmaceutical development after a landmark Phase III study has been conducted and the drug has received approval for marketing. In fact, there may be circumstances where a drug is actually given a license earlier in the development process, for example on the back of a Phase II study or a single-arm trial (Hatswell et al., 2016). In any case, the post-marketing phase seeks to collect and use information from "*real world evidence*" (RWE), i.e. the prescription, consumption, safety and efficacy patterns associated with a given pharmaceutical intervention as observed from data obtained from routine practice.

There are several interesting aspects to the evaluation of pharmaceutical products on the back of post-marketing, or RWE, both from the regulator and the manufacturers point of view, which have an impact in the way the statistical analysis is designed and conducted.

Firstly, while it is not impossible that regulators such as the U.S. Food and Drug Administration (FDA) or the European Medicine Agency (EMEA) require further randomized study after a drug has entered a market, post-marketing studies are often based on *observational* evidence. This implies obvious considerations regarding the possibility for bias, e.g. due to self-selection of individuals into different treatment strategies, which becomes embedded in the data collection. In addition, it is possible that there be delays in market entry between major jurisdictions — for example, there is some evidence to suggest that the US market may be more receptive to drugs licensed with non-Phase III evidence (Hatswell et al., 2016; Djulbegovic et al., 2018). In this case, it is possible that other regulators may use the actual post-marketing data from the US to complement experimental evidence be-

fore making a decision. This has also relevant implications particularly when the focus of the investigation is in the assessment of the "real world" safety of the drug — a situation often termed "surveillance" or "pharmacovigilance" (for some interesting descriptions of these issues within a Bayesian setting, see Dumouchel, 1999; Ahmed et al., 2009; Madigan et al., 2010; Prieto-Merino et al., 2011).

Secondly, post-marketing data may often be used to construct a wider network of evidence, whose main outcome is the economic evaluation of a set of interventions. In a typical situation, one of the drugs being assessed may be newer and there may be experimental data (perhaps from a Phase III trial) available to quantify its effectiveness, possibly against standard of care or placebo. However, there may be other drugs already existing on the market to target the same disease or condition; as mentioned in Chapter 15, from the economic point of view, it is important to contrast the new option with all the relevant alternatives. For this reason, post-marketing data on the other drugs available can be used to allow for indirect comparisons. RWE can be also used for planning of a clinical development programme (e.g. a Phase III study for a new drug), to inform estimates of treatment effects, again in a network meta-analytic context (Martina et al., 2018). Much of this is addressed in Chapter 14.

Thirdly, a related area in which the use of RWE is particularly important — and, crucially, in which the application of Bayesian methods is pivotal — is the economic evaluation of interventions for which the main clinical outcome is represented by a suitable time-to-event, e.g. oncological drugs. This is interesting because often studies such as those for cancer drugs trials produce relatively immature data, where a large proportion of the sample is subject to censoring (Latimer, 2011). Despite the inherent limited amount of evidence present in these data, the economic evaluation process requires the modellers to *extrapolate* the resulting survival curves over a long-term horizon, in order to assess the economic performance of the interventions being compared.

In this chapter we focus particularly on the last case described above. We present a general framework for the integration of experimental and observational studies (possibly derived by post-marketing evidence, or simply from individual level data sets recording information on ancillary aspects of the model). Firstly, in Section 16.2 we briefly review the main features of data classified as RWE. Then, in Section 16.3, we present the general modeling framework that can be used in the case of time-to-event data and the specific issues encountered when using these kind of data for economic evaluation. Section 16.4 presents a case study based on the work of Benaglia et al. (2015); this is not meant to be an exhaustive representation of the possibilities associated with Bayesian modeling using RWE, but it is chosen here because it allows to showcase some of the advantages provided by the inbuilt flexibility and modularity of Bayesian analysis. Finally, Section 16.5 provides some general conclusions and outlooks for future research.

16.2 Real World Evidence

Real World Evidence (RWE) is a terminology increasingly often used to describe observational data obtained using Electronic Health Records (EHR). These are collections of large-scale administrative healthcare data (e.g. GP consultation records or hospital admission/mortality records). Such data are reasonably easy to access and represent a relatively inexpensive means of obtaining a large, representative sample of the general population, when compared to specifically-designed epidemiological studies or RCTs. Typically, EHRs are collected for administrative purposes but there is much scope for these data to be used

for research, increasingly often under the paradigm of *personalized medicine*. This terminology is used to indicate a model of health care provision based on the customization of medical decisions, treatments, practices, or products being tailored to the individual patient — examples include "big data", including those based on genomic measurements (often termed "biobanks").

The main strengths and challenges of EHRs can be summarized as follows.

16.2.1 Primary care databases

Primary care databases can be generally defined as collection of de-identified individual level data from a network of general practices across a given jurisdiction. Primary care data are often linked to a large number of other health-related data, for example pharmaceutical prescriptions, laboratory exams, mortality, etc, in order to provide a longitudinal and representative population health dataset. They can be summarized as having the following characteristics.

- Cover a large proportion of the general population and are linked longitudinally;
- Data can be modeled at the individual patient, GP practice, regional and/or national level, thus allowing (at least theoretically) for hierarchical structures and levels of information to be accounted for in the analysis;
- Patient self-selection may occur, to some degree; in addition, the level of missing data present can be substantial, for some variables.

16.2.2 Mortality/Hospital admission registries

Population registries are usually compiled and maintained by the national statistical authorities (e.g. the US Census Bureau or the European Union Eurostat and its member states affiliates). Essentially every country has an official statistics office collecting data on the relevant population and often these include health-related measures. Classical examples include mortality counts or hospital admission records (e.g. the Hospital Episodes Statistics data set in the UK, or Medicare in the US). The main features of these types of registry data include the following aspects.

- Cover the whole general population, ensuring statistical power and representativeness;
- Variables such as age, sex, date of event, address and ethnicity are routinely recorded, but typical confounding factors are not, which makes the data less useful at the individual level unless a link with external cohorts/surveys is established;
- Data are often used at a small and consistent geographical level (e.g. Middle Super Output Areas, MSOAs, as defined by the Census in the UK) for disease surveillance and risk assessment. They are generally free from missing data issues.

RWE is increasingly popular in fields such as economic evaluation of health-care interventions, particularly in the case where the underlying evidence upon which not just the statistical, but the entire economic model is built (see Chapter 15) is limited — for example, because of short follow up that does not allow to capture the long term effects of a given intervention/drug.

This is often the case when the main clinical outcome is represented by time-to-event variables. We turn our attention to this case in the following, first by describing the specific issues associated with economic evaluation and the general modeling framework.

16.3 Economic modeling and survival analysis

The use of survival modeling has been discussed in Chapter 10. Here we focus on the specific case of survival analysis as embedded in a wider economic modeling — see Chapter 15 and, among others, Spiegelhalter et al. (2004) and Baio (2012).

While interventions that impact upon survival (e.g. cancer drugs) form a high proportion of the treatments appraised by agencies such as the National Institute for Health and Care Excellence (NICE; Latimer, 2011, 2013), modeling for survival analysis in health economics may be challenging. The main reason is that, in order to quantify accurately the long term economic benefits of a new intervention, it is necessary to estimate the mean survival time (rather than usual summaries, such as the median time). In fact, economic evaluation is is concerned with *decision-making*, rather than inference and to do this, we need to estimate the population average benefits of any treatment (i.e. expressed as a function of the mean survival curve). Thus, we usually need to extrapolate the observed survival curves to a (much) longer time horizon than there are data available (see Figure 16.1).

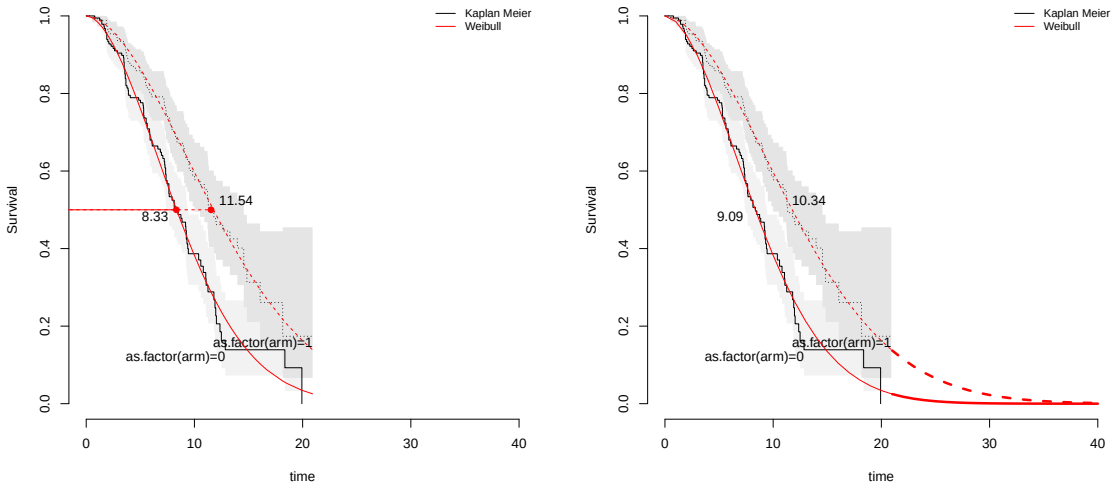

FIGURE 16.1: Survival analysis is generally concerned with determining the *median* time-to-event, as shown in panel (a); in this particular case, the survival curves are presented for two treatment arms. In the controls, the median survival time is 8.33, while it is 11.54 in the active treatment arm. In health economic evaluation, it is necessary to expand the possibly limited time horizon from the experimental data, in order to compute the *mean* time-to-event, e.g. as in panel (b). In this particular analysis, the distance between the two means (9.09 and 10.34 in the control and active treatment) is shorter than in terms of medians.

To this aim, a parametric approach to survival analysis is usually followed and is recommended by NICE technical support document (TSD; Latimer, 2011). This suggests a systematic process in which several parametric models are fitted to the available data in order to select the "best" one, which is then used to produce the relevant output for the economic model, i.e. the most important part of the analytic framework.

16.3.1 General modeling structure

In a study whose main outcome is represented by a time-to-event variable, the observed data consist at least of the pair (t_i, d_i), for individuals $i = 1, \ldots, n$, where: $t_i > 0$ is the observed time at which the event under study occurs; and d_i (for "dummy" variable) is an event indicator, taking value 1 if the event occurs and t_i is indeed observed, or 0 when the $i-$th individual is "censored" (and thus the actual value of t_i is effectively missing). This is usually referred to as "right" censoring and it is usually considered in economic evaluation, although, particularly in the case of RWE, it is possible that individuals are subject to other forms of censoring (for example in the case where the time of the very first measurement might be unknown).

In any case, the observed data are modeled using a suitable probability distribution characterized by a density $f(t \mid \boldsymbol{\theta})$ and defined as a function of a vector of relevant parameters $\boldsymbol{\theta} = (\mu(\boldsymbol{x}), \alpha(\boldsymbol{x}))$. Here we consider: a vector of potential covariates $\boldsymbol{x}$ (e.g. age, sex, trial arm, etc.); a *location* parameter $\mu(\boldsymbol{x})$, which indicates the mean or the scale of the distribution; and a (set of) ancillary parameter(s) $\alpha(\boldsymbol{x})$, which describes its shape or variance. While it is possible for both μ and α to explicitly depend on the covariates $\boldsymbol{x}$, usually the formulation is simplified to assume that these only affect directly the location parameter.

Since $t > 0$, we often model the location parameter using a generalized linear model (GLM)

$$\eta_i = g(\mu_i) = \sum_{j=0}^{J} \beta_j x_{ij} [+ \ldots], \tag{16.1}$$

where $g(\cdot)$ is typically the logarithm and $x_{i0} = 1$ for all i, so that β_0 is the intercept — notice however that the choice of the function $g(\cdot)$ depends on the underlying modeling assumptions. Generally speaking, (16.1) can be extended to include additional terms — for instance, we may want to include random effects to account for repeated measurements or clustering as e.g. done in Chapter 9. We indicate this possibility using the $[+ \ldots]$ notation. When using a Bayesian framework, the model needs to be completed by specifying suitable prior distributions for the parameters.

We can use this general framework to complement the information coming from the available individual level data for one of the interventions being assessed, while including in the wider statistical model a number of "modules", which cover other sources of information. By doing so, we may be able to augment the limited evidence (e.g. because of censoring or short follow up) available in the main data set. We show one such example in the next section.

16.4 Case study: Implantable cardioverter defibrillators for the secondary prevention of cardiac arrythmia

Benaglia et al. (2015) consider the case of implantable cardioverter defibrillators for the secondary prevention of cardiac arrhythmia. In particular, following extensive meta-analysis of existing experimental studies, the two interventions being compared to perform a full economic evaluation are: Anti-arrhythmic drugs (AAD; which in this case can be considered as the *reference* treatment); and: Implantable cardioverter defibrillators (ICDs; the treatment we wish to evaluate in terms of its economic performance in contrast to the reference). The most interesting aspect of this case study is the use of multiple source of evidence, including RWE to complement limited information provided by the available time-to-event data.

The main source of data is a UK cohort study consisting of 535 patients implanted with ICDs between 1991 and 2002 (with average age at implant of 60); notice that this data set only includes (partial) information on only one of the two treatments under consideration. In addition, as is often the case (and even more so, in the case of data obtained from relatively small Phase II/III studies of highly innovative drugs), such data provide limited information because of the follow up available is not long enough. In the current case, the observed follow up of just over 10 years allows us to produce estimates of the time to event (sudden cardiac death) based on survival curves reaching about 0.37 on the Kaplan-Meier estimates (see Figure 16.2).

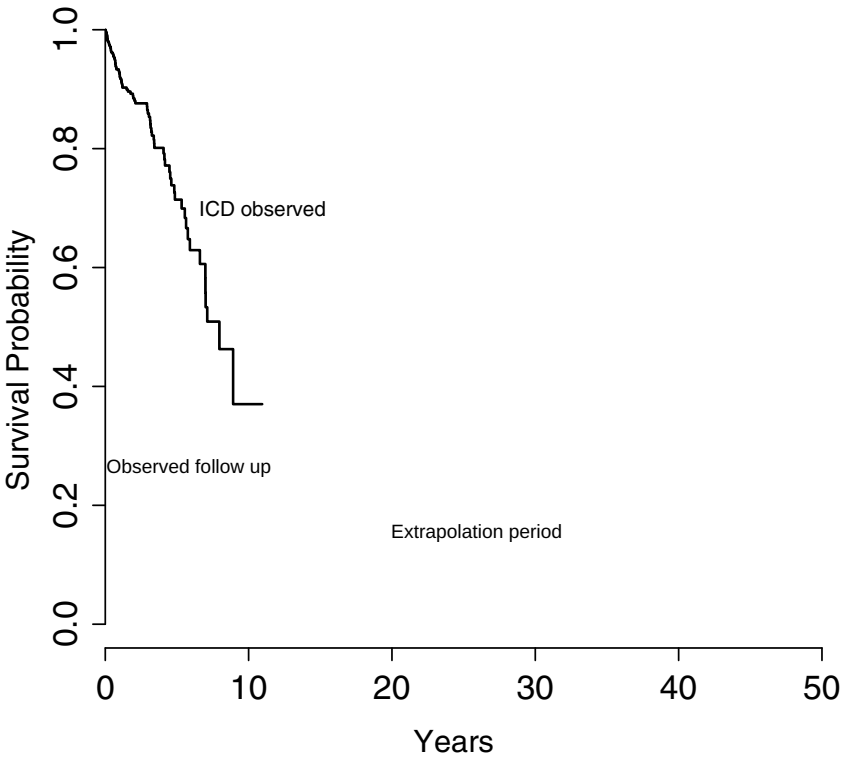

FIGURE 16.2: Kaplan-Meier curve for the observed cohort data on the 535 patients implanted with ICDs. The observed data only reach a survival proportion of about 0.37 over the actual follow up. The $x-$axis covers the actual time horizon required for the economic evaluation (extrapolating over the long term consequences). Adapted from Benaglia et al. (2015)

This means that, although the data may be representative of the overall underlying population, the information available is definitive for only about 60% of the target individuals. For the remaining proportion, we do not have evidence to determine the time at which the event will occur (if it indeed does occur).

For the purposes of the economic evaluation, however, we need to consider two further complications. Firstly, as mentioned above, because of the necessity to extrapolate data over the long term horizon, we typically model data such as those available in this case using parametric models; once the parameters have been estimated based on the observed data, the survival curve can be computed for any required time. However, as happens in this case, a large part of the survival curve will be by necessity based on the extrapolation. In fact, often the underlying data are so "immature" that they do not even reach the median

time during the observed follow up (particularly for longer-term outcomes, such as overall survival). Consequently, it is likely that the extrapolation be extremely sensitive to the parametric model chosen to fit the observed data. NICE suggest testing a set of "standard" models, including popular Weibull, Gamma, Gompertz and Log-Normal, but ultimately, the choice over the actual goodness of fit is based on untestable assumptions.

This is where the use of RWE can come handy — particularly when incorporated in a full Bayesian approach. The main idea is that external data can be used to "anchor" the extrapolated curve by adding information, for example, on the general population characteristics. In this particular case, individual level data are also available from age-sex matched national population registry, produced by the UK statistical authority (Office for National Statistics, ONS). The intuitive rationale behind the use of these data is that the general population is arguably healthier than that considered in the cohort (which is affected by cardiac arrythmia and thus is potentially at greater risk of death). Thus, the general population mortality can act as a sort of upper bound to the survival curve in the cohort. The combination of these two sources of information has the capability of stabilising the inference and the extrapolation from the parametric model of the cohort data.

The easiest way to perform this anchoring is perhaps to assume some form of *proportional hazards* of mortality

$$h_{\text{ICD}}(t) = e^{\gamma} h_{\text{UK}}(t),$$

where $h_{\text{ICD}}(t)$ and $h_{\text{UK}}(t)$ are the hazard function for the ICD and the general UK population, respectively. This would imply that overall mortality in the cohort could be related to overall mortality in the general population by a constant factor e^{γ} (for some parameter γ to be estimated).

However, in a case such as the present, this assumption would be hardly tenable: ICD patients are in fact likely to be perhaps at much greater risk of arrythmia death, because of their very nature. In addition, given that it is likely that the proportion of deaths caused by arrythmia changes over time, this assumption of constant proportionality is almost certain to induce bias over the extrapolation period.

To overcome this issue, Benaglia et al. (2015) build on work presented in Demiris et al. (2015) and use a poly-hazard model for the observed survival times in the cohort data. In a nutshell, poly-hazard models extend the basic set up of a survival model by accounting for the possibility that in fact the observed times are the result of a mixed data generating process, depending on several independent components. For example, we may consider that the occurrence of the event under study depends on M independent causes and that we are willing to model each using a suitable Weibull distribution. Using standard mathematical relationships linking the density of a time-to-event variable to the survival and hazard functions, indicated as $S(t)$ and $h(t)$, respectively (see Chapter 10), we obtain

$$\begin{aligned} f(t_i \mid \boldsymbol{\theta}) &= h(t_i)^{d_i} S(t_i) \\ &= \left[\sum_{m=1}^{M} \alpha_m \mu_{im} t_i^{\alpha_m - 1} \right]^{d_i} \left[\exp\left(- \sum_{m=1}^{M} \mu_{im} t_i^{\alpha_m} \right) \right], \end{aligned}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)^{\top}$ and $\boldsymbol{\theta}_m = (\alpha_m, \mu_{im})^{\top}$ are the shape and scale for the $m-$th component of the mixture. This model is termed a "Poly-Weibull" distribution.

Using population registry mortality data grouped by causes of death (arrhythmia vs all other causes), we can apply the Poly-Weibull model to effectively assume

$$\begin{aligned} h_{\text{ICD}}(t) &= h^{\text{arr}}_{\text{ICD}}(t) + h^{\text{oth}}_{\text{ICD}}(t) \\ &= e^{\gamma} h^{\text{arr}}_{\text{UK}}(t) + h^{\text{oth}}_{\text{UK}}(t) \\ &= e^{\gamma} \alpha_1 \mu_1 t^{\alpha_1 - 1} + \alpha_2 \mu_2 t^{\alpha_2 - 1} \end{aligned} \tag{16.2}$$

Equation 16.2 encodes the (less restrictive) assumption that it is only mortality for arrythmia that varies differentially between the cohort and the general population, while mortality for all other causes is assumed identical in the two. This makes the symplifying assumption of proportionality in the hazards more tenable.

The way in which the information derived from the registry is combined with the cohort data is the following. The registry data consist of two subsets $(\boldsymbol{t}_1, \boldsymbol{d}_1)$ and $(\boldsymbol{t}_2, \boldsymbol{d}_2)$, indicating the survival times and censoring indicators for those who die for arrythmia (subscript/component 1) and all other causes (subscript/component 2), respectively. We can model these two subsets independently using, e.g. a Weibull model

$$(t_{kj}, d_{kj}) \sim \text{Weibull}(\mu_k, \alpha_k)\mathbb{I}(d_{jk},)$$

for each of the $j = 1, \ldots, J$ individuals included in the registry and for $k = 1, 2$. Here we use the notation $\mathbb{I}(d_{jk},)$ to indicate the potential censoring process — if the $j-$th individual in the registry data dies for other causes, then they will be censored for arrythmia-related mortality and vice-versa. Taking full advantage of the modularity of the Bayesian approach, we model the registry and the cohort data using shared parameters (μ_k, α_k), which will be estimated by explicitly combining the two sources of information.

The second complication to do with the wider objective of the analysis being the full economic evaluation is that we want to estimate the *incremental* benefits of the intervention under investigation (ICD) against the reference (AAD). Thus, it is necessary to include additional evidence into the modeling framework; this comes in the form of aggregated summaries produced by the meta-analysis of published studies. These summaries suggest an estimate for the hazard ratio of ICDs vs AAD of 0.72 with 95% interval (0.60 — 0.87) for all causes mortality and of 0.50 with 95% interval (0.37 — 0.67) for arrythmia-related mortality.

While this information is obtained from a standard, frequentist analysis, it is possible to combine this information with some fairly general assumptions in order to turn it into an informative distribution capable to capture the current level of uncertainty on the hazard ratio of ICDs vs AAD. Specifically, Benaglia et al. (2015) assume Normality on the log scale for the hazard ratio of arrhythmia-related mortality and select suitable informative values for the mean and standard deviation (on the log scale) to encode the results of the meta-analysis. Simple computations using statistical software allows us to determine that a $\text{N}(-0.693, 0.148^2)$ produces a distribution such as that shown in Figure 16.3. The picture shows a histogram obtained using 100 000 simulations from the assumed Normal distribution; it is easy to simply exponentiate each of the simulated values and then produce the output shown in Figure 16.3, which depicts the implied distribution for the hazard ratio of ICDs vs AAD. As is possible to see, the mean of the distribution is, as requested, around 0.5 and most of the mass is indeed included in the interval (0.37 – 0.67).

We can use this information to build up a model to estimate mortality in the AAD population — one relatively simple way of doing so is to again assume a poly-hazard structure, where all cause mortality is again identical with the age- and sex-matched general population, while the arrhythmia-specific mortality is proportional to the general population. In this case, Benaglia et al. (2015) model

$$h_{\text{AAD}}(t) \quad = \quad e^{\delta+\gamma}\alpha_1\mu_1 t^{\alpha_1-1} + \alpha_2\mu_2 t^{\alpha_2-1}. \tag{16.3}$$

Equation 16.3 encodes these assumptions as well as the fact that the incremental arrhythmia-related mortality (in comparison to the general population) is described by a combination of two factors. The "baseline" extra log hazard ratio (derived in the ICDs cohort in comparison to the UK population), γ is summed to the log hazard ratio δ, which represents the increase in mortality for AAD vs ICDs. It is possible to turn the distribution shown in Figure 16.3(a) into one describing the inverse log hazard ratio (i.e. the incremental

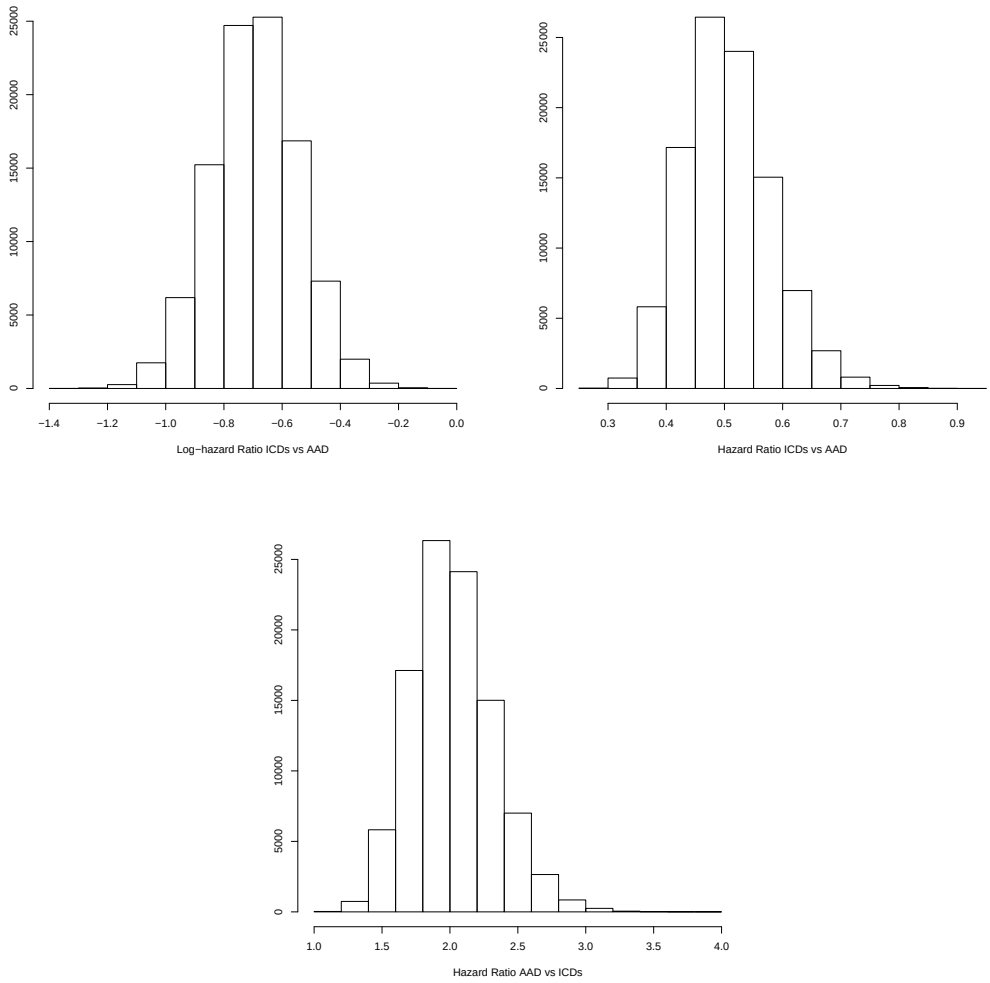

FIGURE 16.3: Panel (a) shows a histogram of a vector of 100 000 simulations from the informative $\mathrm{N}(-0.693, 0.148)$; when exponentiated, these values produce the histogram in panel (b), which shows the implied distribution for the hazard ratio of ICDs vs AAD. Finally, panel (c) shows the histogram of the implied distribution for the hazard ratio of ADD vs ICDs, obtained by simply exponentiating the inverse of the values simulated and graphed in panel (a)

effect of AAD vs ICDs), by simply modeling $\delta \sim \mathrm{N}(0.693, 0.148^2)$ — Figure 16.3(c) shows the resulting histogram for 100 000 simulations from this distribution, rescaled to represent the actual hazard ratio of AAD vs ICDs.

Figure 16.4 shows the results obtained by Benaglia et al. (2015). As is possible to see, the Poly-Weibull model performs better than the standard Weibull (which would assume a single hazard for all causes of mortality), particularly in the estimation for the AAD "arm". Using a simple Weibull model would artificially inflate the survival curve, thus potentially reducing the benefits of ICDs. The general population survival curve acts as an upper limit, guiding the process of extrapolation beyond the observed follow up — notice that there is

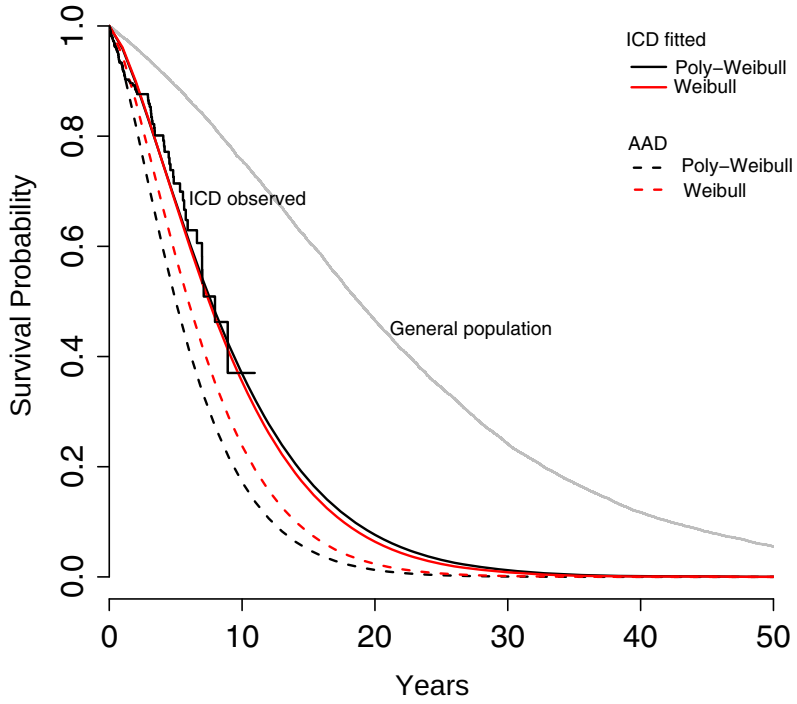

FIGURE 16.4: Survival curves for the two interventions, over a long-term horizon of 50 years. Both the observed individual-level ICD data from the cohort and the aggregated level data for AAD are modeled using a simple Weibull and a Poly-Weibull distribution. The general population (registry) data are also included in the analysis to "anchor" the extrapolation of the survival curves beyond the observed follow up (which is only about 10 years). Adapted from Benaglia et al. (2015)

no need to extrapolate the registry data, as they are available for the entire population at all ages and thus can be used as a bounded estimate of the arrhythmia-related mortality that might be observed in the ICDs cohort and in the AAD population (which at time 0 are on average 60 years old), if we were able to actually follow them up for further 50 years.

This is interesting because it is related to another element of Bayesian modeling that may be helpful in stabilizing the inference. Given the age at entry in the observation is, on average, 60, we may reasonably assume an informative prior distribution on the scale parameters used in the Weibull and Poly-Weibull distributions to imply virtually no probability that individuals can survive longer than other 60 years (i.e. past the time when they would be 120 years old). Benaglia et al. (2015) provide more details on how this has been achieved by setting a Uniform(0, 60) prior on the mean survival time, to then rescale the underlying scale of the Weibull components.

16.5 Conclusions and further developments

Real world evidence is potentially an exciting means to bring external information into analyses based on immature data. The example shown in this chapter has considered a situation

in which cohort data are available for a relatively long time horizon, which nonetheless does not allow a full characterisation of the long-term benefits and costs associated with given interventions. This situation is even more prevalent and perhaps important in many experimental studies investigating cancer drugs, one of the most important and researched areas of clinical and pharmaceutical development.

The Bayesian machinery has the advantage of allowing for a principled integration of data sources, which in cases such as this might allow a substantial "regoralization" of the resulting inference. Of course, this does not come about by magic — in fact, there are several assumptions that need to be encoded in a complex model, such as the one we have discussed here. Judgments of underlying exchangeability across the (sub-)populations that are investigated in different data (e.g. a trial for the drug under investigation and a population registry for some comparators that are already available on the market) are most likely necessary in order to build and successfully run models such as the one described here. However, as in essentially all Bayesian analyzes, these assumptions need to be explicit and the model can be "debugged" in a comprehensive way.

The use of RWE has great potential, particularly when embedded in a wider economic modeling, which is based on the statistical component as the, arguably, most important building block, but that at the same time decisively move forward from the static point of view of finding statistical significance, in favour of a more comprehensive decision-making approach. From the computational point of view, this is also related to innovative methods, such as Hamiltonian Monte Carlo (see Chapter 1), which can be used to produce very efficient estimations in the presence of highly structured models, such as those based on time-to-event and censoring. Research in this area is valuable and in fact ongoing, with recent developments for such models being implemented in software such as `stan` (see Chapter 1).

Bibliography

Ahmed, I., Haramburu, F., Fourrier-Reglat, A., Thiessard, F., Kreft-Jais, C., Miremont-Salame, G., Begaud, B., and Tubert-Bitter, P. (2009). Bayesian pharmacovigilance signal detection methods revisited in a multiple comparison setting. *Statistics in Medicine*, 28(13):1774–1792.

Baio, G. (2012). *Bayesian Methods in Health Economics*. Chapman Hall, CRC, Boca Raton, FL.

Benaglia, T., Jackson, C. H., and Sharples, L. D. (2015). Survival extrapolation in the presence of cause specific hazards. *Statistics in Medicine*, 34(5):796–811.

Demiris, N., Lunn, D., and Sharples, L. D. (2015). Survival extrapolation using the poly-Weibull model. *Statistical Methods in Medical Research*, 24(2):287–301.

Djulbegovic, B., Glasziou, P., Klocksieben, F. A., Reljic, T., VanDenBergh, M., Mhaskar, R., Ioannidis, J. P. A., and Chalmers, I. (2018). Larger effect sizes in nonrandomized studies are associated with higher rates of EMA licensing approval. *Journal of Clinical Epidemiology*, 98:24–32.

Dumouchel, W. (1999). Bayesian data mining in large frequency tables, with an application to the fda spontaneous reporting system. *The American Statistician*, 53(3):177–190.

Hatswell, A., Baio, G., Berlin, J., Irs, A., and Freemantle, N. (2016). Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 19992014. *British Medical Journal Open*, 6(e011666).

Latimer, N. (2011). NICE DSU Technical Support Document 14.

Latimer, N. (2013). Survival Analysis for Economic Evaluations Alongside Clinical Trials—Extrapolation with Patient-Level Data: Inconsistencies, Limitations, and a Practical Guide. *Medical Decision Making*, 33:743–754.

Madigan, D., Ryan, P., Simpson, S., and Zorych, I. (2010). Bayesian methods in pharmacovigilance. In Bernardo, J., Bayarri, M., Berger, J., Dawid, A., Heckerman, D., Smith, A., and West, M., editors, *Bayesian Statistics*. Oxford University Press, 9 edition.

Martina, R., Jenkins, D., Bujkiewicz, S., Dequen, P., and Abrams, K. (2018). The inclusion of real world evidence in clinical development planning. *Trials*, 19(1):468.

Prieto-Merino, D., Quartey, G., Wang, J., and Kim, J. (2011). Why a Bayesian approach to safety analysis in pharmacovigilance is important. *Pharmaceutical Statistics*, 10(6):554–559.

Spiegelhalter, D., Abrams, K., and Myles, J. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley, New York, NY.