Incorporating background mortality into survival extrapolations: determining the accuracy using a simulation study

Author: **Daniël Dante Hordijk** *(s2002310)*

Study programme: **Industrial Engineering and Management** ***(Healthcare Technology & Management)***

University of Twente

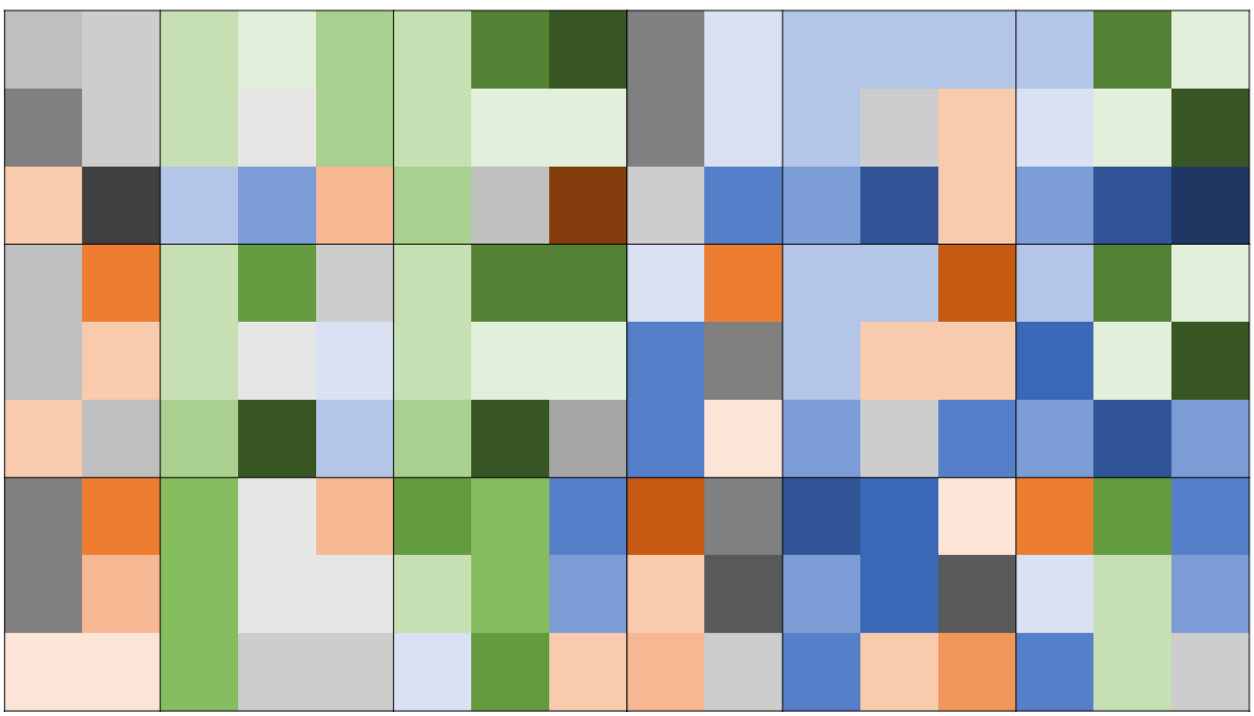
Assignment at: **OPEN Health** *(Rotterdam)*

Supervisors: **dr. X.G.L.V Pouwels** *(University of Twente - 1st supervisor)*

Prof. dr. ir. H. Koffijberg *(University of Twente - 2nd supervisor)*

N. Dunnewind *(OPEN Health)*

A.M.V Gittfried *(OPEN Health)*



Contents

1 Introduction & Background

When companies are developing a new healthcare technology, such as a new drug, diagnostic technique, or surgical procedure, the end goal is to release that technology onto the market. This will almost always involve obtaining approval from the healthcare governing bodies in the countries of their target markets, especially so for new drugs. For these approvals, some governing bodies will want not just an analysis and proof of the health-related effects, but also a health economic evaluation of the new technology when implemented into the respective national healthcare programs based on data gathered during the clinical trials (Coyle et al., 2023; NICE, 2022).

In the United Kingdom (UK) specifically, the National Institute for Health and Care Excellence (NICE) that governs healthcare technology assessments (HTAs) in the UK notes that a large portion of the technologies NICE assesses affect the survival of patients, and thus having an accurate estimation of survival is essential (Latimer, 2011; Latimer, 2013). However, patients are typically only monitored for a limited time during a clinical trial, and some patients will still be alive at the end of this follow-up period. Thus, to get an estimate of the survival of these patients, survival analyses and extrapolations are required (Clark et al., 2003).

Recent developments have shown that current survival extrapolation methods do not adequately account for general population background mortality (GPM), which reflects the mortality of the general population and is affected by factors such as age, sex, calendar year, location, etc. (Rutherford et al., 2020; Verheul et al., 1993). As GPM increases with age, some extrapolations could result in patients having lower mortality than what is observed in the general population if not adjusted correctly for GPM, leading to clinically implausible results (Latimer, 2011). Several researchers have described or compared approaches of incorporating GPM information into survival extrapolations in specific situations, but a general consensus on what method should be used in certain situations is lacking (Guyot et al., 2017; Jackson et al., 2017). Others have recommended assessing the performance of GPM incorporating extrapolation methods in a more generalised simulation study (van Oostrum et al., 2021). This thesis responds to that recommendation.

In the next chapter, a problem description will be given based on the given assignment by OPEN Health and the context of the van Oostrum et al. (2021) study that inspired the assignment. Then, an extensive literature review was conducted, results of which will be reported in Chapter 3. Chapter 4 describes the methods used. Chapter 5 will present the results, which will be further discussed in Chapter 6. Finally, Chapter 7 will give conclusions and recommendations.

2 Problem description

This chapter will start with a brief literature review of some relevant sources cited by van Oostrum et al. and others to get a better understanding of the problem of incorporating GPM into survival analyses as in their research. Then, the assignment as originally presented by OPEN Health will be reviewed and a description of the problem context of this research will be given. Finally, this chapter will be concluded by describing the core problem and defining the research questions.

2.1 Literature review

2.1.1 Survival analyses and extrapolations

Survival analyses are a common technique in health economic evaluations, as a high proportion of technologies assessed affect survival (Latimer, 2011; Latimer, 2013). Survival analyses are used to estimate the time to an event of interest, generally called the survival time. The event can be many things, such as the time from remission until relapse, or from diagnosis until death. Specific survival analysis methods are required because survival data is typically censored, for example because the event of interest does not occur during follow-up for all patients in the trial or some patients are lost to follow-up (Clark et al., 2003).

Most survival analyses will use a Kaplan-Meier curve to visualise and describe survival information gathered during a clinical trial, for example to obtain the median survival. Such a Kaplan-Meier curve gives a non-parametric estimate of the survival probability, or $S(t)$, meaning the probability that a patient survives until time t. Kaplan-Meier curves are used because the specific formula for $S(t)$ accounts for patients that are censored, either because the event of interest did not occur for them or because they were lost to follow-up (Kaplan & Meier, 1958). Another important function in survival analyses is the hazard function, or $h(t)$, which is the instantaneous event rate for someone who has already survived until time t (Clark et al., 2003).

The Kaplan-Meier curve, however, will not give any information about expected survival beyond the trial period. Thus, to extrapolate survival beyond trial observations, a parametric model is often fitted to the patient-level data. For doing so, the hazard function is assumed to follow a certain distribution, such as the Weibull or exponential distribution, and parameters are estimated using maximum likelihood estimation (MLE) (Kleinbaum & Klein, 2005). Once a distribution has been fitted, outcomes such as expected survival can be estimated by extrapolating the survival.

2.1.2 Survival extrapolations incorporating background mortality

When a parametric model is fitted solely to the survival data from a clinical trial to make an extrapolation and estimate the expected survival as described above, the resulting all-cause mortality (ACM) functions disregard the difference between disease-specific mortality (DSM) and general population background mortality (GPM). This can lead to the extrapolated overall survival being higher than the survival seen in the general population since mortality is mostly driven by DSM during the trial period but GPM hazards will increase over time, for example because patients become older during the extrapolation period and the risk of age-related death will increase. In other words, extrapolations based solely on the trial data may produce biased extrapolations as the increasing GPM hazards are ignored, even though they may explain a larger part of mortality than DSM in the long term (van Oostrum et al., 2021). Therefore, recent NICE guidance recommends including GPM in ACM functions, as this currently is not common practice in survival analyses (Rutherford et al., 2020), and other guidance recommends using GPM when DSM is low or a treatment effect is large (Coyle et al., 2023).

Van Oostrum et al. (2021) tested several methods of incorporating GPM hazards into ACM functions, as recommended by NICE, and compared these to an extrapolation without adjustment for GPM on three datasets from oncological case studies. The following four extrapolation methods were compared:

- **Internal additive hazards:** add GPM hazards to the DSM hazard function and fit a distribution over the combined hazard function (Jakobsen et al., 2019; Latimer, 2011),
- **Converging hazards:** patients have a higher initial mortality compared to the general population, but this decreases until the mortality rate converges to that of the general population,
- **Proportional hazards:** a hazard ratio that represents the excess mortality between the patients and the general population is calculated to multiply the GPM hazards with to obtain ACM,
- **External additive hazards:** like the internal additive approach, GPM hazards are added to DSM when making extrapolations, however, GPM is ignored when fitting the survival model (Jackson et al., 2017).

Van Oostrum et al. (2021) compared the four GPM incorporating methods to each other and to extrapolations that did not include GPM hazards. For each of the methods and case studies, multiple parametric distributions were fitted (exponential, Weibull, log-logistic, lognormal and Gompertz).

The findings from van Oostrum et al. (2021) were that GPM adjustments are important, as not adjusting for GPM will often result in survival extrapolations exceeding that of the general population. The one approach they found to have face validity in all instances was the internal additive hazards approach. However, they also found that the approaches have very different outcomes, and some methods could be more applicable in different situations. Finally, they recommended comparing the performance of the four approaches on additional datasets to test the generalisability of their findings beyond the oncological case studies, and to compare the approaches using a simulation study with known DSM and GPM (van Oostrum et al., 2021).

2.2 Given assignment

The assignment given by OPEN Health was a direct response to the call for this research on the generalisability of the findings of van Oostrum et al., which OPEN Health proposed to determine by performing a simulation study. A simulation model would be built that can model patients with differing DSM and GPM, incorporating additional parametric distributions to generate more varied datasets compared to van Oostrum et al. Then, the proposed extrapolation methods by van Oostrum et al. would be applied to the output data of the simulation to compare different scenarios and determine the performance of the methods.

2.3 Problem context

OPEN Health provides several services to pharmaceutical companies, such as providing support in market access strategy and patient engagement. Furthermore, they perform analyses of clinical trials and new health technologies, for example to estimate the effectiveness of a new health technology or for gaining approval from the relevant healthcare governing bodies. Thus, performing survival analyses is a common practice at OPEN Health. The question of whether or not to include GPM in these analyses is typically discussed with clients beforehand, and the answer often depends on the specific requirements for the approval processes of new healthcare technologies.

Herein lied the main problem, as the current guidance for HTAs in, for example, the UK and Canada does recommend considering GPM adjustment in survival analyses (Coyle et al., 2023; Rutherford et al., 2020), guidance on what specific methods to use is lacking (Coyle et al., 2023; Jackson et al., 2017). Van Oostrum et al. (2021) drew some initial conclusions, for example that the internal additive hazards approach was at least face valid for the different case studies used. However, these conclusions only apply to those specific case studies. Thus, more research was needed on the accuracy and applicability of the various GPM adjusting approaches in settings other than the oncological case studies to provide more scientific basis for guidance for HTAs.

2.4 Core problem

To summarize, the core problem at hand was the lack of guidance for selecting a survival extrapolation method that incorporates GPM. Such guidance could be used by OPEN Health to improve the quality of their survival analyses, as it would give a stronger scientific basis when selecting an extrapolation method. There were other problem owners in this case, such as HTA bodies that want to improve their guidelines and other researchers that study survival analysis techniques. Thus, other perspectives that do not necessarily apply to OPEN Health were also considered such that the conclusions and recommendations of the research applied to other problem owners as well. To summarize, the goal of the research was to provide guidance for selecting a survival extrapolation method that incorporates GPM when performing survival analyses for HTAs.

2.5 Research questions

Based on the core problem research questions were defined to solve the following research problem:

More research is needed on the performance of survival extrapolation methods that incorporate GPM because of gaps in literature related to the performance of survival extrapolation methods that incorporate GPM.

As the current research is responding to a recommendation to use a simulation study, the research question corresponding to the research problem is as follows:

What is the performance in terms of accuracy of survival extrapolation methods that incorporate GPM information in scenarios with different patient characteristics and availability of information?

With the following sub-questions:

1. How can we implement the survival extrapolation methods that incorporate GPM into survival analysis packages for R?

To initialise the research and get a better understanding of the extrapolation methods that incorporate GPM, the extrapolation methods compared in this research were implemented into an R script before considering how the simulation would be performed. The selection of extrapolation methods to compare was based on a literature review described in Chapter 3.

2. How do we create simulated datasets which reflect a wide range of relevant patient characteristics?

When the extrapolation methods were implemented and validated, the simulation study was designed. The simulation study should create datasets similar to a clinical trial, but with a known survival. As such, datasets were generated and censored such that the survival was known from the uncensored data, and extrapolations could be performed on the censored data in order to assess the performance of the extrapolation methods. To explore how such datasets could be created using a simulation study a literature review was performed and described in Chapter 3.

3. How accurate are the extrapolation methods?

Finally, to conclude the research, the performance of each of the methods was compared.

3 Literature review

This chapter describes an exploratory literature review aimed at obtaining a better understanding of survival analysis, extrapolation, how GPM can be included, and simulation to ensure the research at hand was performed with the correct techniques and according to the best practices described in literature. Furthermore, gaps in current guidance were identified, as well as the various options available in terms of models, performance criteria, etc. In the next section, the search terms and selection process will be described. Afterwards, the results of the literature review will be presented in two sections, one on survival analysis and extrapolation, and one on simulation studies.

3.1 Search terms

As a broad search for "Survival analysis" on PubMed yielded over 695,000 papers, the query needed to be more specific. Thus, for finding the literature on survival analysis and extrapolation, two searches were performed, namely:

1. "survival analysis" AND "extrapolation"
2. "survival analysis" AND ("background mortality" OR "general population mortality" OR GPM)

Articles that focus on survival analysis techniques in general (i.e. guidance for survival analysis, presents a new technique, etc.) were included, and articles that describe an execution of a survival analysis (i.e. an actual comparison of two interventions) were excluded. Furthermore, studies discussing methods to address measurement errors (i.e., handling wrongly recorded biomarkers, missing data, etc.) in survival analysis and performance comparisons between other factors than the survival model used (i.e. effects of level of censoring, trial population size, etc.) were also excluded, as these were outside of the scope of the research. The first query yielded 455 articles at the time of the search (17th of March 2023), and the second yielded 130 articles on PubMed. The initial articles were assessed based on title to determine whether the articles are general discussions of survival analysis techniques or an execution of a survival analysis. Then, based on abstracts, the articles were screened based on whether they met the inclusion criteria. After screening, the first query yielded 73 relevant articles, and the second a further 10 relevant articles.

3.2 Survival analysis and extrapolation

The general problem context presented in Chapter 2 is also shared by most of the literature found in the search. Within the context of health economic analyses, an estimate of additional costs, resources and health consequences that result from the use of a novel intervention compared to the usual clinical practice is required (Tappenden et al., 2006). Since a high proportion of interventions assessed in HTAs will affect survival, an estimate of survival benefit is essential (Latimer, 2013). For the evaluations, estimates over a lifetime horizon are usually advocated, particularly so for survival, since the evaluations attempt to reflect all differences in costs and outcomes between the two interventions (CADTH, 2017; NICE, 2013). However, since trial data on new interventions is often censored, meaning not all patients will have experienced the event of interest at the end of the trial (Collett, 2003), obtaining a lifetime estimate is not possible using empirical evidence collected during trials and mathematical modelling is required (Eddy, 1985).

The mean survival can be represented by the area under survival curves (Andersson et al., 2013) that plot the proportion of patients that are alive over time (Latimer & Adler, 2022). The mean survival benefit is then represented by the area between the survival curves for the patients that received the intervention and the patients in the control group (who received the usual care or standard treatment) (Collett, 2003; Tappenden et al., 2006). Thus, to obtain an estimate of the lifetime survival, extrapolation is usually performed to obtain a survival curve over the entire lifetime of patients. Generally, parametric models are used to do so, but various other models are available. Since the choice of the model used can have a substantial impact on the survival estimates, and, in turn, cost-effectiveness estimates, the model selection for survival extrapolation is highly important (Bullement et al., 2019; Latimer, 2013; Miners et al., 2005).

3.2.1 Available models

Numerous models are available for selection, and several authors have already performed literature reviews of HTA submissions to get an overview of the models used in research to set up guides and frameworks (Guyot et al., 2011; Jackson et al., 2017; Latimer, 2013; Palmer et al., 2023; Tappenden et al., 2006). Within these reviews, it becomes apparent that there are three main characteristics that differ between models, namely whether proportional hazards (PH) are assumed or not, and the flexibility of the model (Latimer, 2013). The third factor is the use of external data, or data that is not gathered during the trial, which can include data such as GPM or background mortality (Jackson et al., 2017).

The PH assumption refers to the assumption that a *constant* hazard ratio can be applied to a survival curve of one patient group to derive the survival curve of another patient group (Latimer, 2013). The PH assumption can be avoided by fitting two separate models to the two patient groups (Guyot et al., 2011), or having the hazard ratio change over time (Latimer, 2013). Having a hazard ratio that would change over time would also increase the flexibility of the model, which generally refers to the complexity of the model.

In the next sections, the various models available found in the literature review will be discussed separately. Afterwards, methods to include GPM data will be discussed.

3.2.1.1 Parametric models

The simplest of survival models is, unsurprisingly, the most often used method in TAs for NICE (Latimer, 2013), and consists of fitting a parametric model using a certain distribution to the survival data gathered during the trial. The PH assumption is still often used, although there are concerns about the use of the PH assumption in literature (Coyle et al., 2023). When using the PH assumption, a single model is fit to the survival data of the control group (who did not receive the intervention of interest), and the impact of the novel treatment is described by a hazard ratio that can be applied to the survival curve of the control group (Latimer, 2013; Tappenden et al., 2006). However, as mentioned, separate parametric models can be fit to both treatment arms to avoid the PH assumption (Guyot et al., 2011; Latimer, 2011).

When a parametric model is fit, the assumption is made that the hazard function follows a certain probability distribution. The most commonly used distributions are the exponential, Weibull, Gompertz, lognormal and loglogistic distributions (Guyot et al., 2011), and NICE recommends to always compare these models and the generalised gamma distribution in a survival analysis (Latimer, 2011). Other distributions have also been proposed, such as the generalised F distributions (Jackson et al., 2010). Distributions can be fit to the survival data using the maximum likelihood estimation (MLE) method. If the PH assumption is used, the hazard ratio can be included in the likelihood functions (Collett, 2003).

3.2.1.2 Piecewise models

Piecewise or hybrid models involve combining several non-parametric or parametric models (Latimer, 2011), which can be done in several ways. One of the earliest methods proposed was developed by Gelber et al. (1993), referred to as the Gelber method. It involves fitting a parametric model only to the tail of a survival curve and using the estimated parametric model along with estimates from the non-parametric Kaplan-Meier curve to obtain a composite survival-function estimator. The point at which the parametric curve takes over from the Kaplan-Meier can be determined with log-cumulative hazard probability plots. A common criticism of piecewise models is that it requires the analyst to select cut points at which to fit different models, and different decisions of these cut points can affect the overall extrapolation profoundly (Bullement et al., 2019).

3.2.1.3 Cubic spline models

Cubic spline models, or flexible parametric models (FPMs), were originally introduced by Royston and Parmar (2002), mainly to better understand underlying hazard functions and to overcome the PH assumption. The general flexible or spline-based parametric approach attempts to model the logarithm of the baseline hazard function as a natural cubic spline function of log time. Spline functions in general, put simply, are functions that are defined piecewise by polynomials, typically used to smooth or interpolate data. Splines can have any number of subintervals, or k knots, where on the ith knot the spline is defined by the polynomial corresponding to that knot (Ahlberg et al., 2016; Wikipedia, n.d.). Natural cubic splines specifically are constrained to be linear beyond the boundary knots, or $k_{\min}$ and $k_{\max}$.

Royston and Parmar (2002) originally presented methods to smooth log cumulative hazard functions (obtained from typical survival data) into natural cubic spline functions by using full maximum likelihood, including covariates. Since cubic splines are linear beyond the boundary knots, the log hazard at the boundary knot can be used to extrapolate survival to a life-time horizon (Rutherford et al., 2020).

3.2.1.4 Landmark models

Landmark models allow for modelling different responses to treatment, assuming that different responses result in differing survival for patients. For example, a patient group that does not respond to treatment could have a high or increasing hazard, while patients that respond well to treatment can have a low and decreasing hazard. This split into different response groups is done at a defined "landmark" time point, after which separate survival models are fitted to each response group (Rutherford et al., 2020). The models can take any form, but typically parametric models are used (Anderson et al., 1983). Overall survival can then be estimated by weighting the different survival functions by the proportion of patients within that group (Rutherford et al., 2020).

3.2.1.5 Mixture models

Mixture models are similar to landmark models, as they can account for different sub-populations with different survival profiles in a trial. However, mixture models do not explicitly group patients, but rather assign each patient a probability of being in each distribution included in the mixture. Standard parametric models can be used for different mixture components, and standard selection criteria can be used to select the number of mixtures and distributional forms. Extrapolations are weighted for each mixture component and their respective hazard rates (McLachlan et al., 2019; Rutherford et al., 2020).

3.2.1.6 Cure models

Cure models are traditionally used when a proportion of patients will never experience the event of interest, or are, in other words, cured (Boag, 1949). Consequently, their disease-specific hazard rate will reach zero at some point, after which the corresponding cause-specific survival function will reach a plateau at a non-zero value. As the cause-specific survival will then never reach zero, cure models will typically model cause-specific survival alongside other cause mortality, or adopt a relative survival approach. Thus, the model estimates a cure fraction among the population, and estimates survival for uncured and cured patients. By combining the two hazard functions, an overall hazard function can be estimated (Rutherford et al., 2020).

3.2.1.7 Polyhazard models

Polyhazard, or poly-Weibull models are typically used in settings where many competing risks are present. Polyhazard models define an overall hazard function as the sum of several independent risk components, which are described in a Weibull form. Typically, a Bayesian framework is used, where clinical knowledge is used to define the priors (Demiris et al., 2015).

3.2.1.8 Machine learning

Machine learning approaches have also been proposed, such as Random Survival Forests by Ishwaran et al. (2008). These approaches, however, focus on using relationships between a large set of factors into a model, and serve more as a replacement for PH models used in an epidemiological setting (Aivaliotis et al., 2021) rather than for the survival models used in cost-effectiveness analyses.

3.2.2 Incorporation of GPM information

The use of external data, or data that was not gathered during a trial, has been long-established in survival analysis and used in various applications. For example, researchers have proposed using external data to study the loss of life of a patient resulting from a disease (Hakama & Hakulinen, 1977), to put results of survival extrapolations into perspective (Verheul et al., 1993) or to use as a substitute for a control group (Pennington et al., 2018). For this study specifically, however, the type of external data of interest is that which reflects the background, expected, or general population mortality (GPM), terms which are used rather interchangeably in literature.

The expected mortality can be defined as the mortality of a subsample of the general population that is similar to the group of patients at the start of the follow-up period regarding aspects affecting survival, which are typically limited to age, sex and calendar time, and is typically obtained from life table data (Verheul et al., 1993). Using such GPM data has been shown to improve extrapolation performance by several researchers, usually for reasons related to models not accurately capturing the increasing risks of death due to aging (Andersson et al., 2013; van Oostrum et al., 2021). Others found that incorporating background mortality becomes necessary for accurate long-term extrapolation when treatment benefit is large and treatment effect is long, meaning survival is relatively high (Vickers, 2019).

In the next sections, various methods of incorporating such expected mortality data, henceforth referred to as GPM information, found in the literature review will be discussed. Note that most methods are adaptations of models already discussed in Section 3.2.1.

3.2.2.1 Relative survival

Relative survival modelling, also referred to as internal additive hazards (van Oostrum et al., 2021), is a method that derives DSM without requiring specific cause of death information (Andersson et al., 2013). Relative survival modelling is typically used for estimating population cancer survival rates with data obtained from cancer registries, and not from clinical trials, meaning extrapolation is not required (Dickman & Adami, 2006). Recently, however, several authors have adapted the concept to be used for survival extrapolation, such as Rutherford et al. (2020) in recent NICE guidance. It decomposes the ACM hazards of the trial population into two parts:

$$h(t) = h^*(t) + \lambda(t)$$

Where t is the time since diagnosis, $h^*(t)$ is the GPM hazard function, stratified by age, sex, calendar year, and other covariates, and $\lambda(t)$ is the DSM, or excess mortality rate. Then, the corresponding survival function can be rearranged to give the following equation:

$$R(t) = \frac{S(t)}{S^*(t)}$$

Where $R(t)$ is the relative survival as a ratio of the all-cause survival of the trial population and expected GPM survival. $h(t)$ can then be assumed to follow a certain probability distribution, and the separate DSM and GPM hazards can be incorporated in the log-likelihood function of the model during MLE (van Oostrum et al., 2021). More complex estimators for relative survival have been proposed, such as the Pohar-Perme statistic and the standardised relative survival statistic. However, these serve more for comparison of survival between cohorts in, for example, different countries, rather than for survival extrapolation for a single cohort (Perme et al., 2012; Sasieni & Brentnall, 2017).

3.2.2.2 Constant additive hazards

A second approach similar to the relative survival method was described by Jackson et al. (2017). Here, however, the assumption is made that the excess hazards are constant compared to GPM. The excess hazards are estimated from the slope of a linear regression fitted to the logit of the relative survival seen in the latter part of the observed data. The linear regression is typically fit after a time at which the hazard ratio is assumed to behave in a stable manner, for example when invasive diagnostics or procedures are no longer required (Chu et al., 2008; Hwang & Wang, 1999). Extrapolations can then be performed by adding the excess hazards to the GPM hazards (Jackson et al., 2017).

3.2.2.3 External additive hazards

An extension to the constant additive hazards method was used by van Oostrum et al. (2021), where researchers also assume that DSM hazards are always additive to GPM hazards, except not as a constant. Rather, parametric models are fit to the trial data, and GPM hazards are added to the fitted hazards afterwards. GPM hazards are *not* included in MLE for model fitting, differentiating the method from the relative survival method. As such, van Oostrum et al. (2021) refer to this method as external additive hazards, and the relative survival method as internal additive hazards.

3.2.2.4 Converging hazards

The converging hazards method, as described by Jackson et al. (2017), assumes that the disease population may have a higher mortality than GPM initially, but their mortality decreases until the mortality rate converges to GPM after some time. Jackson et al. (2017) proposed setting a time whereafter only GPM hazards affect survival of patient, and survival before is estimated using a parametric model fit to the trial data. Other authors implemented converging hazards by fitting a parametric model to the ACM hazards of the trial population without GPM information, and using the fitted ACM hazards until GPM hazards have become higher (van Oostrum et al., 2021). Thus, the overall hazard function of the trial population is as follows:

$$h(t) = \begin{cases} h^*(t) & if\ h^*(t) > h_{ACM}(t) \\ h_{ACM}(t) & otherwise \end{cases}$$

Jackson et al. (2017) mention that the time at which the hazards converge could be seen as the time at which patients are cured, and thus the converging hazards method has strong similarities to cure models when other cause mortality is included. The converging hazards method does not use a cure fraction, however.

3.2.2.5 Proportional hazards

The proportional hazards assumption has also been proposed for use with external data by Jackson et al. (2017), where the hazard ratio between trial populations and the general population is assumed to be constant. Cause specific mortality can then be obtained by multiplying GPM with a certain hazard ratio. Jackson et al. (2017) state that the hazard ratio should be found in literature.

3.2.2.6 Methods for cubic spline models

The previous methods have mostly focused on "standard" parametric models, although methods to incorporate external data in cubic spline models have also been proposed. Nelson et al. (2007) extended cubic spline models for use with relative survival by including an excess mortality component in the hazard function, also introducing methods to include time-sensitive covariates. Andersson et al. (2013) proposed several extrapolation methods, as the original implementations do not allow for extrapolation, that use assumptions on how DSM compares to GPM as originally suggested by Hakama and Hakulinen (1977).

The *linear trend* method assumes that the hazard function continues linearly beyond the boundary knot, which could also be used with the relative survival models proposed by Nelson et al. (2007). The two other methods proposed are relatively similar to the converging hazards and external additive hazards methods previously described. The *cure* method assumes that after some time the mortality rate of the patient group will return to GPM, like the converging hazards method. Next, the *constant excess hazards* method assumes that excess hazards are constant beyond the boundary knot, which can then be used to estimate the relative survival function (Andersson et al., 2013).

3.2.2.7 Polyhazard models

Benaglia et al. (2015) adapted a polyhazard model for use with datasets from the general population and one from a patient population with some disease of interest to extrapolate the survival of the disease group. Here, the hazards for the disease group are assumed to have two causes, one from the disease of interest, and one for other causes. As causes of death are not observed in the study data, independent Weibull models cannot be fit. Thus, a polyhazard model can be used by assuming that the other-cause survival distribution is the same as that for the general population, and the increase in hazards (or excess hazards) due to the disease is obtained by a cause-specific log hazard ratio between the study and population groups.

3.2.3 Available guidance for model selection

With the available models for survival extrapolations with and without using GPM information described, this section will be devoted to the available guidance on selecting between one of those models to use for survival analysis for an HTA. Most importantly, the NICE Decision Support Unit Technical Support Documents (TSDs) will be discussed, as they are a leading HTA governing body. Furthermore, other literature found in the literature review will be discussed.

TSD 14 presents an algorithm for selecting a model for survival analysis that should be followed when preparing a submission to NICE. TSD 14 focuses mostly on standard parametric models and piecewise models. It is recommended to develop log-cumulative hazard plots that plot the log(-log) of the survival function against log(time) of the two patient groups and selecting different models based on the relationship of the two curves:

- If the curves are straight and parallel, parametric models with a PH assumption should be considered,
- If the curves are straight and not parallel, two separate parametric models should be fit and assessed,
- If the curves are not straight, piecewise or other more flexible models should be considered.

Then, the various models that are under consideration should be fit and assessed based on certain performance criteria, and a selection should be made. For parametric models, TSD 14 recommends that exponential, Weibull, Gompertz, log-logistic lognormal and generalised gamma models should always be considered (henceforth referred to as the "standard parametric models"). To assess the suitability of a model, TSD 14 recommends assessing visual fit of the parametric curve compared to the Kaplan-Meier curve, statistical fit using Akaike's Information Criterion (AIC) or Bayesian Information Criterion (BIC), and clinical plausibility using either external data or an expert opinion (Latimer, 2011). The latter has been mentioned in other articles as a highly important factor to assess suitability of survival models (Bell Gorrod et al., 2019; Williams et al., 2017). More recent NICE guidance (TSD 21) includes descriptions of flexible models for survival analysis intended for use when hazard functions are too complex for the models mentioned in TSD 14, but does not provide a selection algorithm to distinguish between these models (Rutherford et al., 2020).

Other researchers have also discussed procedures to select a survival model. For example, the algorithm described in TSD 14 was criticised by Bagust and Beale (2014) in part for assuming that patient-level data was available to the researchers while manufacturers rarely make patient-level data available, even to those contracted by the manufacturers to review the trial evidence. Bagust and Beale (2014) presented their own methods for survival extrapolation that were later criticized by Latimer (2014) for recommending the exclusion of certain data points, and the recommendation to assume survival follows an exponential model unless other evidence exists. Thus, the methods of Bagust and Beale will not be explained further.

Palmer et al. (2023) presented a selection algorithm for flexible models to address the lack of such an algorithm in TSD 21. The algorithm presents 4 questions that will aid model selection for flexible models, with the most important being how many treatment arms are present, whether flexible models are required and whether a cure fraction can be assumed. Methodologies to answer each of these questions are presented in intermediate steps of the algorithm, where using clinical expert opinion is highly important throughout. Once at the final question of the algorithm, plausible models have been selected and a final comparison can be made based on external evidence, clinical plausibility, hazard plots (including a comparison with GPM hazards), and AIC or BIC statistics.

3.2.3.1 Guidance on use of external data and background mortality

The use of external data is not mentioned in the TSD 14 algorithm for model selection, but *is* recommended for assessing model validity. TSD 14, however, does not discuss incorporating GPM into models directly (Latimer, 2011). TSD 21 does discuss the relative survival model and recommends incorporating external data into the other types of flexible models described in the TSD. For example, for cure models the use of GPM in a relative survival approach is recommended when cause-specific mortality is unknown. Furthermore, TSD 21 recommends that GPM should be used to check if mortality is lower than expected in the general population and incorporated if so (Rutherford et al., 2020). Recent guidance from the Canadian Agency for Drugs and Technologies in Health (CADTH) states that GPM can either guide the shape of the long-term survival curve or be used as an upper limit in extrapolations, and should especially be considered in situations where DSM is low or treatment effect is high (Coyle et al., 2023).

Other authors do not provide much guidance regarding selection between methods to incorporate GPM beyond some inconsistent conclusions and recommending the assessment of clinical validity. For example, for the GPM assumptions for cubic spline models discussed in Section 3.2.2.6, Jakobsen et al. (2019) found that models using the linear trend assumption with relative survival and models using the cure assumption performed well in certain settings, but found no consistent satisfactory performance for any of the assumptions in other settings, only noting that including GPM is more important in younger populations. Andersson et al. (2013) concluded that using the linear trend method with a relative survival approach was sufficiently accurate for use with cubic spline models, although cure models were sometimes found to perform better. van Oostrum et al. (2021) recommend the external additive hazards method for younger populations and converging or proportional hazards models only if their assumptions are clinically plausible. Jackson et al. (2017) simply state that long-term assumptions such as the converging or additive hazards assumption are untestable from data alone and should be justified using clinical expertise. Palmer et al. (2023) note that the flexible models described in their algorithm could be implemented using a relative survival framework to incorporate GPM, but do not mention when this is or is not appropriate. For cure models, however, they always recommend using GPM for the cured population.

3.3 Simulation

To conclude the literature review, literature found related to simulation studies, and specifically what such studies tested and how such studies were executed in the context of survival extrapolation will be discussed. To start, an often used simulation framework should be mentioned, namely that of Morris et al. (2019). It has been used to assess performance of various survival extrapolation methods in, for example, the NICE guidance by Rutherford et al. (2020), and several other articles (Gallacher et al., 2021a, 2021b; Kearns et al., 2021). The framework is designed specifically for comparing statistical methods using Monte Carlo simulation, or simulations that use pseudo-random sampling (Morris et al., 2019), and was used for the rest of this study.

Rutherford et al. (2020) used a simulation study in recent NICE guidance to compare the performance of the flexible survival methods presented in the guidance. They compared the RMST until the end of follow-up and overall mean survival based on extrapolation in a single treatment arm, as they argued that if a survival model extrapolates poorly, it is inappropriate for use in economic modelling regardless of how accurately a treatment effect can be predicted. Benaglia et al. (2015) also used overall mean survival in a simulation study to assess performance of survival extrapolation methods. Some studies have focused specifically on RMST until various times beyond follow-up (Gallacher et al., 2021a, 2021b). Others use a loss of lifetime estimate, which serves as a function of the area between the general population and patient population survival curves up to a specific time point (Jakobsen et al., 2019).

Many different factors related to survival analysis have been compared with simulation. Most often, simulation studies compare extrapolation based on data generated using different survival curves. For example, some studies compared different levels of survival or differently shaped survival curves (Benaglia et al., 2015; Rutherford et al., 2020). Differences between using different cure fractions have also been explored in some studies (Jakobsen et al., 2019; Rutherford et al., 2020). Finally, factors such as differing ages of patients (Jakobsen et al., 2019) or different trial sizes have also been compared (Rutherford et al., 2020). Other studies do not use "theoretical" trials, but rather take real-world trials and use their characteristics to generate data, for example by fitting parametric models to the trial data and using a similar level of censoring as what was seen in the trial (Gallacher et al., 2021a, 2021b).

To generate data, simulation studies usually consider DSM and GPM separately. Rutherford et al. (2020) and Jakobsen et al. (2019) both generate two survival times, one based on their defined DSM functions, and one based on their GPM functions, and then pick the lowest survival time as a patient's ACM survival time. Both simulations also generate an age per patient that is used as input for their GPM functions. Benaglia et al. (2015) generated separate patient and general population datasets to compare effects of models that assume a proportional hazard compared to the general population. Then, to censor the data, simulation studies typically generate another time using a censoring function, for example by generating a value from an exponential function. Each patient is then censored appropriately, either if their censoring time is before their ACM survival time, or their ACM survival time is after the end of follow-up (Benaglia et al., 2015; Gallacher et al., 2021b; Jakobsen et al., 2019; Rutherford et al., 2020).

In Rutherford et al.'s simulation study (2020), a frailty term was used in the survival function for disease specific survival. This frailty term is based on the frailty models described by Hougaard (1995), which assume that variability in survival times originate from two separate sources, the first one being the randomness from a hazard function, the second being a random effect called the frailty. The frailty can be univariate or multivariate, and can serve as a replacement for other covariates (such as lifestyle, smoking, etc.) that are known to affect survival but cannot be explicitly included in the analysis.

4 Methods

In this chapter, the various methods used to obtain results for the research questions defined in Section 2.5 will be described. A simulation study was selected for answering the research questions, specifically as it allowed for many different scenarios to be compared. For designing the simulation study, the framework of Morris et al. (2019) was used (see Section 3.3), meaning this chapter of the report will follow the recommended ADEMP structure. First, the overall **A**ims of the study are defined. Then, the **D**ata-generating mechanisms (DGMs) are described, which are the mechanisms used to generate patient survival data in different scenarios. Next, the **E**stimands of interest will be described, following with a definition of the different (extrapolation) **M**ethods used to obtain these estimands. Finally, the **P**erformance measures used per estimand are given.

4.1 Aims

The overall aim of the simulation study was to assess the performance of several survival extrapolation methods that incorporate GPM information for use in health economic models by comparing their extrapolations to a known overall survival. Since, for use in health economic modelling, the relative treatment effect between treatment arms is not necessarily of interest, only the performance on single treatment arms was assessed. Furthermore, an extrapolation method being able to prove there is a difference in treatment effect is not relevant if the method performs poorly when extrapolating survival. Performance was compared between many different scenarios in order to draw conclusions that could provide guidance for selecting a GPM incorporating method in these scenarios. The different scenarios used will be described in the next section.

4.2 Data generating mechanisms

A DGM denotes how random numbers are used to generate a dataset, and for a simulation study, many DGMs are often used or compared (Morris et al., 2019). One thing to note for DGMs, and more importantly the random number generation, is that these random number generators are not actually random, but pseudo-random, meaning that the chain of values generated can be reproduced using a seed value. The DGMs were implemented in R, which has functions readily available for generating random numbers, and these generators were assumed to be robust.

Many factors can differ between research trials, with arguably the most important being the survival of patients and their characteristics. Another point of interest is the amount of information available to the researcher and thus for extrapolation. For example, different extrapolation methods might be more applicable than others if the age and sex of individual patients is known, rather than knowing solely the mean age and division between sexes. Thus, there are two main dimensions to the scenarios that were varied between the DGMs, namely the patient characteristics and the amount of available information. The patient characteristics affect the survival of the patient, and within survival both DSM and GPM should be considered. Thus, the DGMs were structured such that a time of event was generated for each patient based on a certain DSM and GPM, and then censored the data (i.e., as if only data from a trial was available) to perform the desired extrapolation. In Table 3, an overview of all dimensions used for the DGMs is shown. Each of the dimensions will be discussed separately.

4.2.1 Patient characteristics

Performance of the extrapolation methods between various patient populations with different characteristics were compared. For patient characteristics, there are a vast number of factors that can affect their survival and differ between populations. However, as the number of DGMs could not be exceedingly high due to the runtime of the simulation, only a selection of differences in these factors was compared. The final selection of factors, along with specifics on how the data is generated, was as follows:

4.2.1.1 Age

As the incorporation of GPM information is often recommended for modelling age-related risks of death, comparing patient populations with different ages was highly important. Thus, for each DGM, a random age was assigned to a patient using a normal distribution, rounded down to an integer. Here, three different means were used to represent young ($\mu = 35$), "average" ($\mu = 50$) and older populations ($\mu = 75$). Different standard deviations were also used, however, varying standard deviations were included in a separate dimension of DGMs, referred to as the heterogeneity of the population, that will be described in Section 4.2.1.3.

4.2.1.2 Survival

Next, the survival for each patient needed to be defined. The survival for each patient consisted of two components, as both DSM and GPM hazards should be considered. For GPM survival, the most recently available life tables from the USA obtained from the CDC (Arias & Xu, 2022) were used in the simulation. As each patient also has a generated sex (which is generated differently based on the level of heterogeneity, described in Section 4.2.1.3), the appropriate life table was used for each patient. To generate a time until a GPM event, a uniform random number was generated between 0 and the current survival probability of the patient based on their age and life table. Using this upper limit normalizes the generated survival probability for each patient based on their age, as otherwise a negative time until event could be generated. Then, the newly generated random number was used to interpolate a time until event from the life table. Note that the method to generate a GPM time was the same for all DGMs.

For DSM, both differing levels of survival and underlying probability distributions were compared in order to compare a large variety in survival curves used to generate patient data. To keep the "level of survival" the same over different distributions and to remove the need for human input to determine the parameters, basic parametric models were fit to real-world survival data of different diseases. For this purpose, three Kaplan-Meier curves were selected that reflect a low, medium and high level of survival, which were identified based on a short brainstorm with the OPEN Health supervisors. For low survival, a survival curve of patients with pancreatic cancer was used, with a median survival of 17.0 months (Kuhlmann et al., 2004). For medium survival, a survival curve of patients with myocarditis was used, where 56% of patients were still alive after 5 years (median survival was not reported) (Magnani et al., 2006). For high survival, a survival curve of patients with ulcerative colitis was used, where 59% of patients were still alive after 40 years (median survival was not reported) (Jess et al., 2006).

Each of the Kaplan-Meier curves were digitized, and using the Guyot et al. (2012) algorithm, patient-level data was generated. Then, parametric models were fit to the survival data, using the Weibull, log-normal, log-logistic and Gompertz distribution. These are the standard distributions recommended by Latimer (2011), with the exclusion of the exponential distribution and generalised Gamma distribution. The Weibull distribution is also capable of modelling constant hazards like the exponential distribution does, hence its exclusion. Since the Weibull and lognormal distributions are specialised cases of the generalised Gamma distribution (Latimer, 2011), the generalised Gamma distribution was excluded as well. Then, the parameters of the various fitted survival models were used to generate data for the simulation. The statistical fit of the models to the data was not considered further. The resulting Kaplan-Meier curves and parametric models are shown in Figure 1, Figure 2, and Figure 3. The models and their parameters are shown in Table 1.

Figure 1:
Digitized Kaplan-Meier curve and parametric models used for low survival scenarios based on a pancreatic cancer dataset (Kuhlmann et al., 2004)

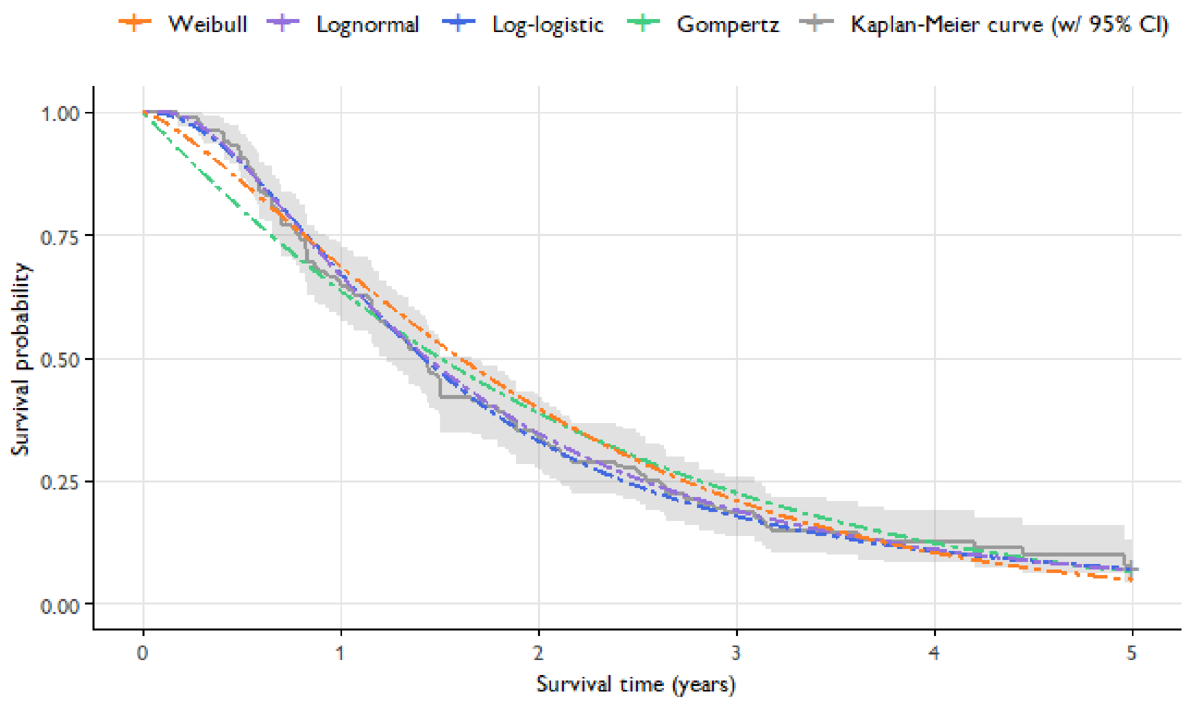


Figure 2:
Digitized Kaplan-Meier curve and parametric models used for medium survival scenarios based on a myocarditis dataset (Magnani et al., 2006)

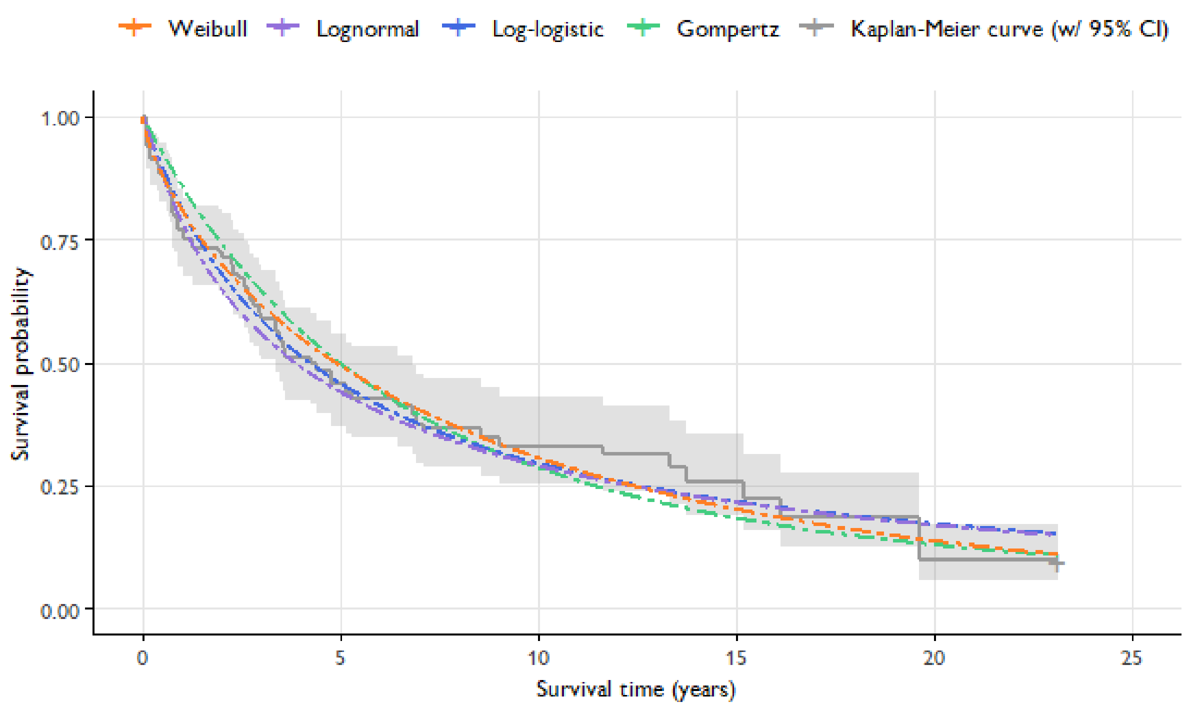


Figure 3:
Digitized Kaplan-Meier curve and parametric models used for high survival scenarios based on an ulcerative colitis dataset (Jess et al., 2006)

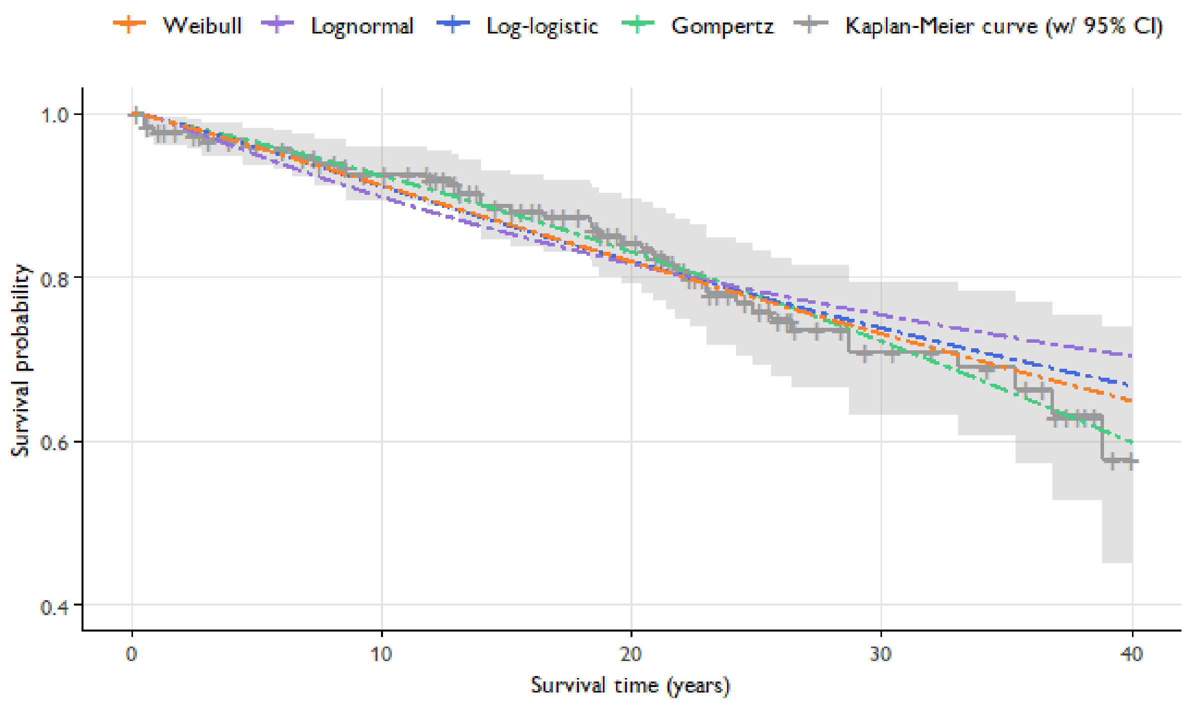


Table 1:
Overview of survival distributions used for simulation

Distribution	Parameter 1	Parameter 2
Low survival (pancreatic cancer)		
Weibull	λ = 2.127	κ = 1.297
Log-logistic	α = 1.418	β = 2.047
Lognormal	μ = 1.437	σ = 2.307
Gompertz	η = 0.094	b = 0.431
Medium survival (myocarditis)		
Weibull	λ = 7.973	κ = 0.745
Log-logistic	α = 4.221	β = 1.012
Lognormal	μ = 3.858	σ = 5.588
Gompertz	η = -0.046	b = 0.156
High survival (ulcerative colitis)		
Weibull	λ = 84.977	κ = 1.112
Log-logistic	α = 72.418	β = 1.172
Lognormal	μ = 109.720	σ = 6.638
Gompertz	η = 0.028	b = 0.007

To generate a time until an all-cause mortality event, the lowest time between the DSM and GPM event times was selected for each patient. The GPM data used assumes no person lives longer than 100 years, meaning the background hazard at 100 years is infinity. As the internal additive hazards method fails if infinite background hazards are supplied, the maximum event time was cut off at 99.999 years.

4.2.1.3 Heterogeneity

In some populations, the differences between patients can be larger than in others. Thus, to get an indication of how different methods perform if a population is more heterogenous, three different levels of heterogeneity were compared by combining several factors. The first factor is the standard deviation of the normal distribution used to generate an age for a patient, with a standard deviation of 3, 6, and 12 for low, medium, and high levels of heterogeneity, respectively.

Secondly, the division over sexes of the patients was considered. Using a random uniform number generator, patients were assigned a sex based on a certain percentage of male versus female patients. In the low heterogeneity scenario, this percentage was either 90% male versus 10% female or 10% male versus 90% female, also assigned randomly with a 50% chance for either sex being the majority. In the medium heterogeneity scenario, this split between sexes was 70% versus 30%, and for the high heterogeneity scenario, the split was an equal 50% versus 50%.

Finally, a frailty term was included for the medium and high heterogeneity scenarios. As described in Section 3.3, frailty models were introduced to be able to model various other factors that affect survival, such as lifestyle and smoking habits, by modelling them as a separate, single random variable. Rutherford et al. (2020), however, applied the concept in reverse in their simulation study by including a secondary random variable in their DGMs for overall DSM survival. This adds an additional source of randomness to the DSM survival times of patients. The DSM survival was then as follows:

$$S_d(t) = S_{d0}(t)^{\exp(Z\beta)}$$

Where $S_{d0}(t)$ are the DSM survival functions as described in the previous section, and Z is the unknown frailty term, distributed normally with a mean of 0 and standard deviation of 1. Different levels of frailty can then be generated by using different values for β. For low heterogeneity, no frailty was used, while for medium and high heterogeneity, $\beta = 0.5$ and $\beta = 2$ were used, respectively.

To summarize, three different levels of heterogeneity were compared, with the following factors:

Table 2:
Scenarios for heterogeneity used in DGMs of simulation

Heterogeneity	σ of age	Division over sexes	β of frailty
Low	3	90% / 10%	*N.A (frailty not included)*
Medium	6	70% / 30%	0.5
High	9	50% / 50%	2

4.2.2 Available information for extrapolation

Besides patient characteristics, another dimension of interest was the available information for extrapolation. Here, four different factors were used for comparison, which are the population size of the trial, the length of follow-up, the level of right censoring, and whether the covariates are known for extrapolation. Thus, after patient data has been generated according to the methods described in the previous section, the data was censored according to these four factors. For each patient, a censoring time was generated from an exponential distribution, which served as the right censoring time. This time, along with the follow-up length, was compared to the ACM event time of each patient. If the ACM event time is after the patients' censoring time or follow-up length, the patient was censored accordingly.

To reduce the total number of DGMs, the factors were combined into a few scenarios as was done for the heterogeneity dimension. Three scenarios were used, which are as follows:

1. Low level of information

This DGM had the lowest amount of information available, with a population size of 100 and a follow-up length of 1 year. For censoring, a rate of 0.3 was used, which should result in approximately 30% of the population being right censored (disregarding their event time). For extrapolation, only the median age of patients and percentage of male versus female patients was known, referred to as having "summary" knowledge available. In high survival scenarios, the generated datasets were found to often have no events occurring within a year of follow-up. As such, scenarios with both low information and high survival were excluded from analysis.

2. Medium level of information

For the medium level of information, a population size of 300 patients was used, with a follow-up length of 3 years. The rate for censoring was 0.067, which means approximately 20% of patients are right censored. For extrapolation, again only summary knowledge was available.

3. High level of information

Finally, the DGM with the highest level of information used a population size of 500 and follow-up length of 10 years. The censoring rate was 0.01, which censors approximately 10% of patients. For extrapolation, the ages and sexes were known for each individual patient and were used for determining the background hazards for the internal additive hazards method, which will be referred to as having "full" information.

4.2.3 DGM overview

To summarize, for each DGM, survival data was generated based on several dimensions of factors. For comparison of results, each level of each dimension was compared to one another, meaning a full-factorial experiment design was used. One combination of scenarios was excluded from analysis, namely those with low information and high survival. Here, patients have a median survival of over 40 years, while only 1 year of follow-up data is recorded, meaning datasets were relatively likely to have zero non-censored events. Not only did models often not converge, but such a situation is unrealistic. In total, 288 DGMs were included. The dimensions, and their levels, are summarized in Table 3.

4.3 Estimands

For estimands, several factors were of interest. Note that per estimand, several performance measures were used that will be described in Section 4.5. The most important estimand used was the overall mean survival, as calculated by the extrapolation methods. Additional estimands used were the restricted mean survival time (RMST), or the fitted survival *during* the trial, cut off when follow-up ends. Note that RMST until the end of follow-up is not an extrapolation, but closer to an interpolation as no estimate is made beyond the available data. However, the estimand served as an indicator of model fit to the short-term or available data, and RMST has also been used in similar simulation studies (Rutherford et al., 2020). Furthermore, survival probability at certain times t based on the level of survival was assessed, with $t = 3$ for low survival, $t = 15$ for medium survival, and $t = 20$ for high survival. This serves as an indicator of clinical validity, as clinicians can typically give such an estimate for a disease. In medium and high information scenarios (with 3 and 10 years of follow-up data, respectively), the survival probability for low survival will not be an extrapolation, as 3 years is still within follow-up. A time t that was an extrapolation for all scenarios would have been preferable, but since the expected disease-specific has already dropped to around 20% after 3 years in low survival scenarios, selecting a later time t increases the risk of the generated survival probability being 0, and thus 3 years was selected for low survival. For medium survival, the expected survival was also around 20% at 15 years, hence why 15 years was selected. For high survival, however, survival would reach 20% long after 40 years, while old populations used an average age of 75 and the maximum age of the simulation was 100. Therefore, 20 years was used as time t for high survival scenarios.

Table 3:
Overview of DGMs for simulation

Dimension *Level*	Definition	# of levels
Age	N ~ (μ, σ_{age})	x 3
Young	μ = 35	
Average	μ = 50	
Old	μ = 75	
Survival	Various parametric models	x 3
Low	Pancreatic cancer survival	
Medium	Myocarditis survival	
High	Ulcerative colitis survival	
Parametric distribution of survival	Weibull, Gompertz, Log-logistic, Lognormal	x 4
Heterogeneity	σ_{age}, division over sexes, frailty	x 3
Low	σ_{age} = 3, 90% / 10%, none	
Medium	σ_{age} = 6, 70% / 30%, β = 0.5	
High	σ_{age} = 12, 50% / 50%, β = 2	
Level of available information	$n_{patients}$, $\lambda_{censoring}$, follow-up length, extrapolation knowledge	x 3
Low	n = 100, λ = 0.3, 1 year, summary	
Medium	n = 300, λ = 0.067, 3 years, summary	
High	n = 500, λ = 0.01, 10 years, full	
	Total number of DGMs:	288*

**Since low information and high survival scenarios are excluded, 288 scenarios are compared rather than the full-factorial 324*

4.4 Extrapolation methods

In this section, the extrapolation methods that were compared will be described, along with details on their implementations. Note that the selection of methods followed from van Oostrum et al. (2021). However, the proportional hazards method was not included as it was unclear how the hazard ratio was obtained in their study. Furthermore, literature related to the method recommends obtaining a hazard ratio from literature, which is not possible in this simulation setting. Other models found in the literature review were not included as they would have required more complicated DGMs that include more factors, which would have increased the number of scenarios tremendously. For example, for testing cure models various cure fractions would have to have been simulated in order to compare performance of the GPM incorporating method when different cure fractions are assumed. For mixture, landmark, polyhazard and machine learning models either multiple patient groups with differing survival or multiple competing risks would have to have been included. Finally, some models were excluded because they require human input during modelling, such as deciding cut points for piecewise models or the number of knots to use for cubic spline models.

For fitting the parametric models, all standard distributions mentioned in NICE guidance were used, which are the exponential, Weibull, log-logistic, log-normal, Gompertz and generalized Gamma distributions (Latimer, 2011). Furthermore, the generalized F distribution was used. For all methods, extrapolations were performed using mid-point Riemann sums with an interval width of 0.1 years, as some methods do not result in a parametric model with a function that can be integrated.

4.4.1 Non-GPM extrapolation

For the first extrapolation approach, a parametric distribution was simply fit to the censored data generated by the DGMs, as described in Section 3.2.1.1. Since the `flexsurv` package in R allows fitting a parametric distribution to survival data, the exact implementation of this method is not discussed further. However, it is important to note that the fitted hazards ($h_{nonGPM}(t)$) of these models were used for some of the other extrapolation methods.

4.4.2 Internal additive hazards

For the internal additive hazards method or relative survival method, `flexsurv` also had a method readily implemented. To use it, a background hazard at the time of event needed to be supplied for each patient in the trial data. To obtain this background hazard from the life table, the assumption was made that hazards are constant throughout a year, as life tables are typically presented with expected survival at a certain age or probability of dying in a year at a certain age. Then, depending on the level of information available, the (censored) event time of the patient was added to either the patient's age or the median age in the trial and a survival probability was obtained from the life table. Again, depending on the level of information available, either a survival probability weighted based on the division over sexes in the trial obtained from the female and male life tables, or a survival probability from the life table corresponding to the patient's sex was obtained.

Since the assumption of constant hazards throughout a year was made, this survival probability could easily be converted to hazards using the parametrization of the exponential distribution (which assumes constant hazards) and solving for the rate:

$$\lambda = -\log S(t)/t$$

Note that $t = 1$, since constant hazards for 1 year were used. Furthermore, it is important that the probability of surviving another year rather than overall survival was used, as otherwise this equation would give the cumulative hazard.

Once the background hazards per patient are calculated, the values were put into the `flexsurv` function and models were fit using the principles described in Section 3.2.2.1.

4.4.3 External additive hazards

For implementing the external additive hazards method, the fitted hazards from the non-GPM models were taken and added to the background hazards in extrapolation. For each point of the Riemann sum, hazards were calculated as follows:

$$h_{ACM}(t) = h_{nonGPM}(t) + \lambda(t)$$

Where $h_{nonGPM}(t)$ are the outputted hazards from the non-GPM models as described in Section 4.4.1, and $\lambda(t)$ are the GPM or background hazards, obtained similarly as in the internal additive hazards method. These hazards were then converted to overall survival and the area under the curve was calculated. For scenarios where patient-level data was available and ages and sexes are known, a weighted average of the hazards was taken based on mean age and division over sexes.

4.4.4 Converging hazards

As with the external additive hazards method, the fitted hazards from the non-GPM models were used to extrapolate for the converging hazards method, except here the background hazards were compared to the non-GPM hazards at each point of the Riemann sum, and the highest hazards were used to obtain the survival curve.

4.5 Performance measures

Two performance measures were used for each estimand. The main measure of interest was the bias of each extrapolation method, or the average deviation from the known survival based on the event times generated by the DGMs and the extrapolated survival using the different extrapolation methods. The known survival was calculated by using the mean survival time of each patient sample, before their event times were censored. The mean was calculated based on the generated patient sample for each replication (with sizes 100, 300 and 500 based on the level of information available), rather than attempting to get the theoretical survival from the underlying survival functions or by pooling all replication data together and calculating the means of these larger datasets. Doing the latter could have resulted in an implicit bias when comparing the extrapolated survival to the known survival. Furthermore, the consistency of each method was assessed. For this purpose, the root mean-square error (RMSE) was used. Initially, coverage was included as a performance measure, but due to the high computation time of obtaining lower and upper limits for the hazards based on parametric models, coverage was excluded from the performance measures.

5 Results

In this chapter, the results from the simulation will be described using the estimands and performance measures as described in Chapter 4. As there were 288 scenarios in total not all results could be analysed in detail, let alone be mentioned in this chapter. Thus, only a selection of results (selected based on a Wilcoxon rank sum test that will be explained further) is shown in this chapter. Complete results are shown in Appendix A, B, and C. Furthermore, this section mostly focuses on summaries and over-arching patterns in results, rather than results for individual scenarios or models. The results are split into results per estimand (see Section 4.3 for a description of the estimands), and reporting for each estimand follows a similar structure. For further interpretation of certain results, specialised datasets were generated, which will be discussed in separate sections throughout this chapter. Note that from this point forward, a "model" denotes a combination of a parametric distribution and method to incorporate GPM (except for non-GPM models, where GPM is not incorporated).

For each estimand, the best performing model in terms of lowest absolute mean bias of the estimand was identified per scenario, meaning it had the lowest absolute mean bias out of all models. Initially, all GPM incorporating methods and parametric distributions used for modelling were considered for each scenario. However, this resulted in relatively "noisy" results, and patterns were difficult to identify. Thus, per scenario, models were filtered based on the parametric distribution used for modelling and the parametric distribution used during data generation. In other words, in scenarios where data was generated using a Weibull distribution, only models that use the Weibull distribution were considered. In this manner, the only difference between the models under consideration was the method in which GPM was incorporated, which was the main point of interest for comparison following from the research question. Moreover, models that used the same underlying parametric distribution used to generate the data should perform well and generate more accurate results than the other distributions from a theoretical perspective regardless of the GPM method. Per estimand, the filtered results will be discussed first and will be used to identify patterns in performance. Then, the unfiltered results will be discussed and compared to the results found in the filtered analysis to draw secondary conclusions.

As only the best performing models were identified, information regarding performance of second best, third best, etc. per scenario is lost. To get an indication of how much better the best performing model performed over the other models, a Wilcoxon rank sum test was employed. Here, per scenario, the biases per replication for each model were taken as samples, and the sample of biases of the best performing model was compared to the other models and tested for statistically significant differences ($\alpha = 0.05$, using a Bonferonni correction for the three or 27 pairs of models compared, depending on whether all parametric distributions or only the parametric distribution used to generate the data were considered). If the best performing model passed the Wilcoxon test for a scenario, it is denoted as a "significant" result.

Per estimand, an overview of the amount of times a model (GPM method and parametric distribution) performed best is shown. Then, per estimand, a visual representation of the best performing models for each individual scenario is shown for one of the four parametric distributions used to generate the data (a subset of 72 scenarios per estimand), as the other dimensions can be ascertained when looking at trial data (i.e., average ages and trial size are known factors), while one cannot easily look at a survival curve and reason it looks like a lognormal distributed curve. The parametric distribution to visually present in this main body of text was selected per estimand based on which distribution had the most significant results using the Wilcoxon test for the filtered results. Visual results for the other distributions used to generate survival data can be found in Appendix A, although they will be discussed in separate sections in this chapter as well.

5.1 Mean survival

5.1.1 Overview

Overall, the external additive hazards GPM incorporating method performed best in the most scenarios in terms of mean survival (142 scenarios out of 288) when filtering for models that used the same parametric distribution as the underlying data. The converging hazards method performed best in 84 scenarios, the internal additive hazards method in 38 scenarios, and non-GPM extrapolations in 19 scenarios. A significantly best method (using the Wilcoxon test) was found in 78% of scenarios overall. Figure 4 charts the number of times each GPM incorporating ranked best for absolute mean bias of mean survival per dimension (keeping all other dimensions the same) on the left, and in total on the right.

Figure 4:

Times each GPM incorporating method ranked best for absolute mean bias of mean survival per dimension and total (models filtered based on parametric distribution)

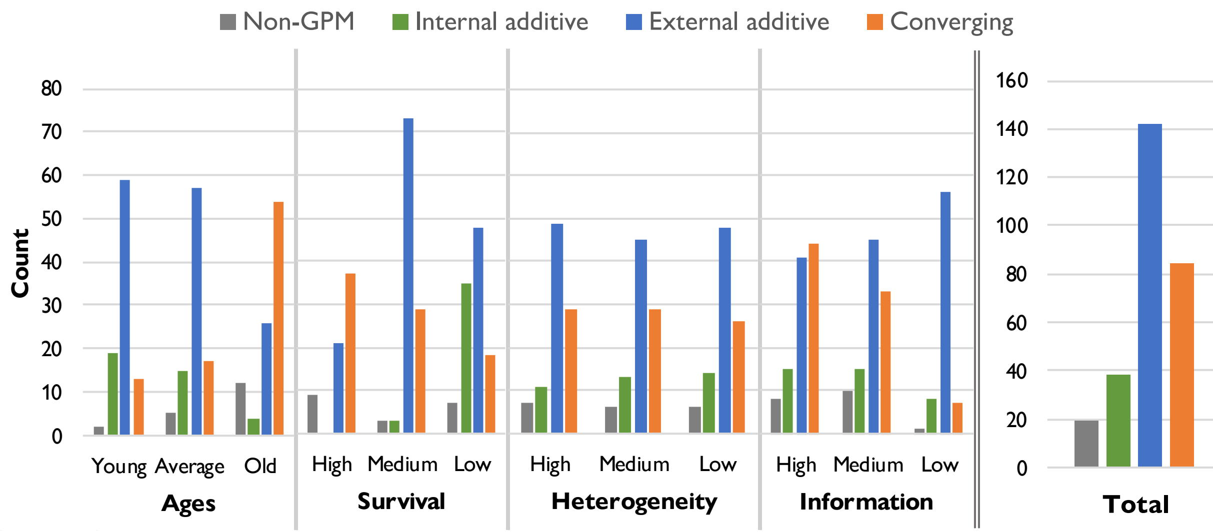
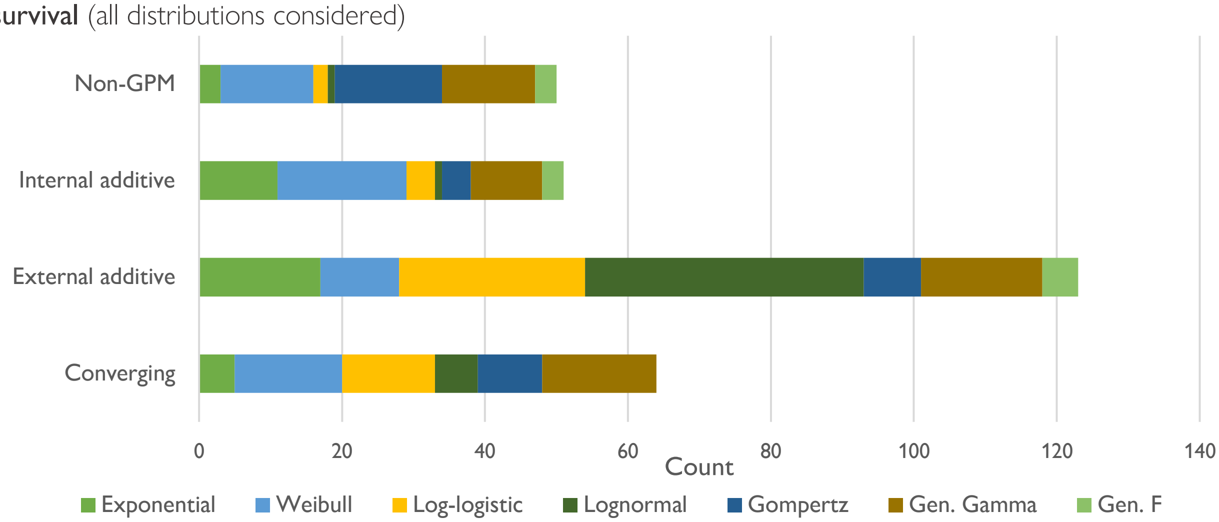


Figure 5:

Times each GPM incorporating method and distribution ranked best for absolute mean bias of mean survival (all distributions considered)



When all parametric distributions were considered in the analysis (regardless of whether the parametric distribution matched that of the underlying data), the external additive hazards method also performed best in the most scenarios in terms of mean survival (123 scenarios out of 288). The converging hazards method performed best in 64 scenarios, the internal additive hazards method in 51 scenarios, and non-GPM extrapolations in 50 scenarios. A significantly best model was found in 36% of scenarios overall. Figure 5 charts the number of times each GPM incorporating method ranked best for mean survival per distribution when all distributions were considered.

5.1.2 Filtered results

5.1.2.1 Log-logistic results

For mean survival, the most significant results were found in the scenarios that used the log-logistic distribution for generating the survival data when filtering for models that used the same distribution as the underlying data (83% of scenarios using the Wilcoxon test). Thus, visualisations of the results per individual scenario will be shown here for the log-logistic scenarios. Visualisations of the results for the other distributions can be found in Appendix A.1 for the filtered results.

The results are presented in three tables, where the first (Table 4) shows the number of times the GPM incorporating method ranked best for the log-logistic distributed scenarios, but more importantly serves as a legend for the other two tables. Table 5 and Table 6 are intended to aid in identifying patterns between dimensions of scenarios, which are used for rows and columns. Differently shaded cells represent the best performing GPM method for the scenario based on the absolute mean bias. Within cells, the mean bias and mean root mean square error (RMSE) is shown, and a black triangle denotes a significant result for the scenario. The pair of tables have swapped main axes for both rows and columns: Table 5 groups by survival and ages, and Table 6 groups by heterogeneity and level of information. The tables presented for other estimands will follow the same structure.

A distinct pattern can be seen in Table 5 can be seen when moving from the bottom left to the top right scenarios, where the **converging hazards** method performed best the most in scenarios with high survival or old ages, although it was mixed with the **external additive hazards** method in scenarios with high survival and young ages; and old ages with either medium or low survival. The **external additive hazards** method performed the best the most often in scenarios with medium survival and young ages; medium survival and average ages; and low survival and average ages. Performance in low survival and young ages was more mixed.

As can be seen in Table 6, the **external additive hazards** method always performed best in scenarios with low information and low heterogeneity, and five out of six times for scenarios with low information and medium heterogeneity. The **converging hazards** method always performed the best in scenarios with medium information and high survival, and with high heterogeneity and old ages. In scenarios with either medium or high information, the **external additive hazards** method generally performed the best, unless survival is high or populations are old, where the **converging hazards** method performed better more often.

Table 4:

Legend for filtered colour-coded table and times a GPM incorporating method ranked best for absolute mean bias of mean survival for log-logistic distributed data. *Amount of times the method ranked best*

	Non-GPM	Internal additive	External additive	Converging
Times method ranked best	*1*	*2*	*39*	*30*

Table 5:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of mean survival per scenario for log-logistic distributed data. Survival and ages as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

		Survival (Log-logistic)							
		High		Medium			Low		
		Information		*Information*			*Information*		
Age	*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium	High
Young	Low	▲ *1.042, 6.347*	▲ *-0.949, 1.361*	▲ *0.381, 3.233*	▲ *0.052, 0.903*	▲ *-0.09, 0.341*	*0.059, 0.435*	*-0.01, 0.113*	*-0.016, 0.063*
Young	Medium	▲ *0.851, 6.063*	▲ *-1.205, 1.558*	▲ *-0.084, 3.122*	▲ *-0.07, 0.888*	▲ *-0.063, 0.365*	*0.025, 0.532*	▲ *0.036, 0.144*	▲ *0.007, 0.078*
Young	High	▲ *0.618, 2.848*	▲ *-1.436, 1.65*	*-0.015, 2.67*	▲ *-0.369, 1.02*	▲ *-0.104, 0.522*	*0.216, 1.42*	▲ *0.078, 0.541*	▲ *0.092, 0.308*
Average	Low	▲ *1.808, 3.54*	▲ *-2.106, 2.106*	▲ *0.121, 2.388*	▲ *-0.067, 0.625*	▲ *-0.271, 0.319*	*0.024, 0.396*	*-0.002, 0.101*	*-0.006, 0.056*
Average	Medium	▲ *1.569, 3.397*	▲ *1.974, 1.976*	*-0.148, 2.209*	▲ *-0.155, 0.627*	▲ *-0.305, 0.354*	*-0.038, 0.449*	▲ *0.013, 0.127*	*-0.025, 0.066*
Average	High	▲ *1.08, 1.865*	▲ *1.412, 1.423*	*0.522, 2.081*	▲ *-0.384, 0.752*	▲ *-0.454, 0.535*	▲ *-0.313, 1.027*	▲ *-0.032, 0.423*	▲ *-0.112, 0.242*
Old	Low	▲ *0.648, 0.698*	▲ *0.06, 0.12*	▲ *-0.353, 0.848*	▲ *-0.543, 0.546*	▲ *0.35, 0.35*	▲ *0.021, 0.289*	▲ *0.081, 0.097*	▲ *0.049, 0.056*
Old	Medium	▲ *0.318, 0.504*	▲ *-0.274, 0.29*	▲ *-0.508, 0.853*	▲ *-0.636, 0.639*	▲ *0.248, 0.249*	▲ *-0.066, 0.293*	▲ *-0.115, 0.123*	▲ *0.086, 0.089*
Old	High	▲ *-0.161, 0.454*	▲ *-0.637, 0.638*	▲ *0.523, 0.874*	▲ *0.48, 0.505*	▲ *-0.015, 0.137*	*0.39, 0.601*	▲ *0.392, 0.401*	▲ *0.132, 0.148*

Table 6:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of mean survival per scenario for log-logistic distributed data. Information and heterogeneity as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

Heterogeneity	Age	Information Low: Survival (Log-logistic) Medium	Information Low: Survival (Log-logistic) Low	Information Medium: Survival (Log-logistic) High	Information Medium: Survival (Log-logistic) Medium	Information Medium: Survival (Log-logistic) Low	Information High: Survival (Log-logistic) High	Information High: Survival (Log-logistic) Medium	Information High: Survival (Log-logistic) Low
Low	Young	0.381, 3.233 ▲	0.059, 0.435	1.042, 6.347 ▲	0.052, 0.903 ▲	-0.01, 0.113	-0.949, 1.361 ▲	-0.09, 0.341 ▲	-0.016, 0.063
Low	Average	0.121, 2.388 ▲	0.024, 0.396	1.808, 3.54 ▲	-0.067, 0.625 ▲	-0.002, 0.101	-2.106, 2.106 ▲	-0.271, 0.319 ▲	-0.006, 0.056
Low	Old	-0.353, 0.848 ▲	0.021, 0.289 ▲	0.648, 0.698 ▲	-0.543, 0.546 ▲	0.081, 0.097 ▲	0.06, 0.12 ▲	0.35, 0.35 ▲	0.049, 0.056 ▲
Medium	Young	-0.084, 3.122 ▲	0.025, 0.532	0.851, 6.063 ▲	-0.07, 0.888 ▲	0.036, 0.144 ▲	-1.205, 1.558 ▲	-0.063, 0.365 ▲	0.007, 0.078 ▲
Medium	Average	-0.148, 2.209 ▲	-0.038, 0.449	1.569, 3.397 ▲	-0.155, 0.627 ▲	0.013, 0.127 ▲	1.974, 1.976 ▲	-0.305, 0.354 ▲	-0.025, 0.066 ▲
Medium	Old	-0.508, 0.853 ▲	-0.066, 0.293 ▲	0.318, 0.504 ▲	-0.636, 0.639 ▲	-0.115, 0.123 ▲	-0.274, 0.29 ▲	0.248, 0.249 ▲	0.086, 0.089 ▲
High	Young	-0.015, 2.67	0.216, 1.42	0.618, 2.848 ▲	-0.369, 1.02 ▲	0.078, 0.541 ▲	-1.436, 1.65 ▲	-0.104, 0.522 ▲	0.092, 0.308 ▲
High	Average	0.522, 2.081	-0.313, 1.027 ▲	1.08, 1.865 ▲	-0.384, 0.752 ▲	-0.032, 0.423 ▲	1.412, 1.423 ▲	-0.454, 0.535 ▲	-0.112, 0.242 ▲
High	Old	0.523, 0.874 ▲	0.39, 0.601	-0.161, 0.454 ▲	0.48, 0.505 ▲	0.392, 0.401 ▲	-0.637, 0.638 ▲	-0.015, 0.137 ▲	0.132, 0.148 ▲

5.1.2.2 Results for other parametric distributions

Overall, not many patterns in performance for mean survival seen for the log-logistic distributed data in Table 5 and Table 6 were visible across the distributions used to generate survival data for the filtered results (see Appendix A.1). Generally speaking, looking at patterns seen in Table 5, the **converging hazards** method performed best more often in scenarios with either high survival or old populations across distributions, even more so in scenarios with high information. In scenarios with medium survival and young ages; medium survival and average ages; and low survival and average ages, the **external additive hazards** method performed best the most in all but the Weibull distributed scenarios, where it was mixed with either the **converging hazards** method or the **internal additive hazards** method (see Table 23). Compared to patterns identified in Table 6, the **external additive hazards** method always performed best in scenarios with low information and low heterogeneity across distributions, and other patterns did not hold across distributions.

Across distributions, the **internal additive hazards** method was rarely selected as the best performing method for mean survival, except for the Weibull distributed scenarios (see Table 23), where in low survival scenarios the **internal additive hazards** method performed the best the most amount of times. Across other distributions, the only times the **internal additive hazards** performed best in a scenario was in low survival scenarios as well.

5.1.3 Unfiltered results

5.1.3.1 Log-logistic results

For results of mean survival where all models were considered, regardless of whether the parametric distribution used for modelling matched the distribution used to generate the survival data, tables with a similar structure to Tables 4, 5 and 6 will be presented and compared to the filtered results. Thus, the unfiltered results of mean survival for the log-logistic distributed scenarios are shown here (since the log-logistic distributed scenarios had the most significant results for the filtered results). Visualisations of results for the other distributions can be found in Appendix B.1, although they will be discussed in this section as well. For the unfiltered results, multiple models using different parametric distributions were under consideration for each GPM incorporating method. As such, Tables 7, 8 and 9 now differ in brightness of colour based on the best performing parametric distribution used for modelling. Table 7 serves as a legend for Tables 8 and 9, and shows the number of times each model (parametric distribution and GPM incorporating method) ranked best for scenarios that used the log-logistic distribution to generate survival data.

None of the patterns seen in the filtered results for mean survival (Tables 5 or 6) showed as explicitly in the unfiltered results (Tables 8 and 9). The scenarios with medium survival and young ages; medium survival and average ages; and low survival and average ages still showed a slight preference for the **external additive hazards** method, albeit more mixed (see Table 8), as was seen in the filtered results (Table 5). In scenarios with low survival, young ages, there was a stronger preference in the unfiltered results for the **external additive hazards** method than in the filtered results. Furthermore, in scenarios with low information, there was a preference for the **external additive hazards** method regardless of heterogeneity, where in the filtered results this was only true for low or medium heterogeneity (see Table 6).

5.1.3.2 Results for other parametric distributions

Comparing the unfiltered results for mean survival of the log-logistic scenarios to scenarios that used other parametric distributions to generate the data, the patterns described in the previous section held for the Weibull distributed scenarios, less strongly for lognormal distributed scenarios and not at all for Gompertz distributed scenarios (see Table 80). Compared to the patterns seen in the filtered results across distributions, the **converging hazards** method still performed better more often scenarios with high survival or old populations in the unfiltered results, although performance within scenarios with high survival or old ages was more mixed with all the other methods. Furthermore, when moving from high to low survival or from older to younger ages, it becomes more rare for the **converging hazards** method to have performed best.

The **internal additive hazards** method was selected as the best performing method for mean survival more often in scenarios with medium survival in scenarios that used other distributions to generate the data, compared to it only being selected in low survival scenarios for filtered results. The **internal additive hazards** method was still mostly absent from scenarios with high survival, however. Across distributions, **non-GPM** extrapolations mostly performed best in scenarios with high survival, and were relatively absent in scenarios with medium or low survival. Another noteworthy result was how for the Gompertz distributed data, models that used the Gompertz distribution only ranked highest in 8 scenarios, while generalised gamma ranked highest in 21 scenarios and Weibull ranked highest in 19 scenarios (see Table 79). For other distributions, the distribution used to generate the data was selected the most often as the highest ranking distribution for extrapolation.

Table 7:

Legend for colour-coded table and times a model ranked best for absolute mean bias of mean survival for log-logistic distributed data. *Amount of times the model ranked best*

	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F	Total
Non-GPM	*2*	*1*	*1*	*0*	*5*	*4*	*2*	*15*
Internal additive	*2*	*4*	*2*	*0*	*0*	*0*	*1*	*9*
External additive	*6*	*0*	*14*	*4*	*3*	*3*	*1*	*31*
Converging	*1*	*3*	*5*	*2*	*1*	*5*	*0*	*17*
Total	*11*	*8*	*22*	*6*	*9*	*12*	*4*	

Table 8:

Colour-coded table for best ranking model for absolute mean bias of mean survival per scenario for log-logistic distributed data. Survival and ages as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

		Survival (Log-logistic)							
		High		Medium			Low		
		Information		*Information*			*Information*		
Age	*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium	High
Young	Low	*1.042, 6.347*	*-0.48, 0.844* ▲	*0.381, 3.233* ▲	*0.052, 0.903* ▲	*-0.042, 0.621*	*0.059, 0.435*	*-0.01, 0.113*	*-0.016, 0.063*
Young	Medium	*0.454, 2.1* ▲	*0.525, 7.918* ▲	*-0.084, 3.122*	*-0.07, 0.888* ▲	*-0.061, 0.551* ▲	*0.002, 0.336*	*0.036, 0.144*	*0.001, 0.082*
Young	High	*0.618, 2.848* ▲	*0.08, 1.066* ▲	*-0.015, 2.67*	*-0.024, 1.74* ▲	*-0.104, 0.522*	*0.208, 1.515*	*0.078, 0.541*	*0.024, 0.345*
Average	Low	*-0.456, 1.228* ▲	*0.06, 3.948* ▲	*0.121, 2.388* ▲	*-0.067, 0.625* ▲	*0.093, 0.354*	*0.024, 0.396*	*-0.002, 0.101*	*-0.006, 0.056*
Average	Medium	*-0.184, 14.968*	*0.269, 3.915* ▲	*-0.148, 2.209* ▲	*0.04, 2.064*	*0.039, 0.339*	*0.031, 0.303*	*0.013, 0.127*	*-0.012, 0.065*
Average	High	*-0.029, 4.214*	*-0.241, 0.58* ▲	*0.235, 3.827*	*-0.258, 1.075* ▲	*0.203, 0.379* ▲	*0.164, 1.14*	*0.013, 0.452*	*0.008, 0.233*
Old	Low	*0.165, 0.929*	*-0.028, 0.315*	*-0.123, 0.889*	*0.1, 0.616*	*0.028, 0.075*	*0.021, 0.289* ▲	*0.028, 0.082*	*0.004, 0.034*
Old	Medium	*0.011, 0.589*	*-0.141, 0.534*	*-0.082, 2.266*	*0.035, 0.577* ▲	*0.013, 0.08*	*-0.016, 0.206* ▲	*-0.015, 0.069* ▲	*-0.027, 0.037* ▲
Old	High	*-0.096, 0.547* ▲	*0.011, 0.384*	*0.223, 1.445*	*-0.036, 0.377*	*-0.015, 0.137*	*-0.238, 0.497* ▲	*0.081, 0.263* ▲	*-0.014, 0.076*

Table 9:

Colour-coded table for best ranking model for absolute mean bias of mean survival per scenario for log-logistic distributed data. Information and heterogeneity as main axes. *(Bias, RMSE),* ▲ *= Significant result*

		Information							
		Low		Medium			High		
		Survival (Log-logistic)		*Survival (Log-logistic)*			*Survival (Log-logistic)*		
Heterogeneity	**Age**	Medium	Low	High	Medium	Low	High	Medium	Low
Low	Young	▲ 0.381, 3.233	0.059, 0.435	1.042, 6.347	▲ 0.052, 0.903	-0.01, 0.113	▲ -0.48, 0.844	-0.042, 0.621	-0.016, 0.063
Low	Average	▲ 0.121, 2.388	0.024, 0.396	▲ -0.456, 1.228	▲ -0.067, 0.625	-0.002, 0.101	▲ 0.06, 3.948	0.093, 0.354	-0.006, 0.056
Low	Old	-0.123, 0.889	▲ 0.021, 0.289	0.165, 0.929	0.1, 0.616	0.028, 0.082	-0.028, 0.315	0.028, 0.075	0.004, 0.034
Medium	Young	-0.084, 3.122	0.002, 0.336	▲ 0.454, 2.1	▲ -0.07, 0.888	0.036, 0.144	▲ 0.525, 7.918	▲ -0.061, 0.551	0.001, 0.082
Medium	Average	▲ -0.148, 2.209	0.031, 0.303	-0.184, 14.968	0.04, 2.064	0.013, 0.127	▲ 0.269, 3.915	0.039, 0.339	-0.012, 0.065
Medium	Old	-0.082, 2.266	▲ -0.016, 0.206	0.011, 0.589	▲ 0.035, 0.577	▲ -0.015, 0.069	-0.141, 0.534	0.013, 0.08	▲ -0.027, 0.037
High	Young	-0.015, 2.67	0.208, 1.515	▲ 0.618, 2.848	▲ -0.024, 1.74	0.078, 0.541	▲ 0.08, 1.066	-0.104, 0.522	0.024, 0.345
High	Average	0.235, 3.827	0.164, 1.14	-0.029, 4.214	▲ -0.258, 1.075	0.013, 0.452	▲ -0.241, 0.58	▲ 0.203, 0.379	0.008, 0.233
High	Old	0.223, 1.445	▲ -0.238, 0.497	▲ -0.096, 0.547	-0.036, 0.377	▲ 0.081, 0.263	0.011, 0.384	-0.015, 0.137	-0.014, 0.076

5.1.4 Exploring performance of the external additive hazards method

To explore possible reasons why the external additive hazards method outperformed the internal additive hazards method (the latter of which is currently most often recommended in literature) more often, a set of models was generated for two scenarios with different levels of survival that had relatively large differences in mean bias of extrapolated mean survival between the internal additive hazards and external additive hazards method and where an external additive hazards model performed relatively well.

The first was the scenario with young ages, high survival generated using the Gompertz distribution, medium heterogeneity and high information, where external additive hazards models combined had a mean bias of mean survival of 0.47, and the internal additive hazards models combined had a mean bias of 21.13. Following from the unfiltered results from the simulation for this scenario, the external additive hazards models that performed the best used the generalised gamma distribution (with a mean bias of 0.015, the lowest out of all the models for this scenario) and the internal additive hazards models that performed the best used the Weibull distribution (with a mean bias of 13.54). Figure 6 charts the expected GPM and DSM survival curves for this scenario with the generated known Kaplan-Meier curve, and the extrapolated survival over time for the generalised gamma external additive hazards model and the Weibull internal additive hazards model (the extrapolated non-GPM generalised gamma survival is shown as its hazards are used for the external additive hazards extrapolation).

The second is the scenario with average ages, medium survival generated using the log-logistic distribution, low heterogeneity and ;pw information, where external additive hazards models combined had a mean bias of -0.17, and internal additive hazards models combined had a mean bias of 4.17. Based on the unfiltered results of the simulation, the external additive hazards models that performed the best used the log-logistic distribution (with a mean bias of 0.12, the lowest out of all the models for this scenario. The best performing model did not have a significant result using the Wilcoxon test), and the internal additive hazards models that performed best used the Weibull distribution (with a mean bias of -2.14, which was the fourth lowest mean bias for this scenario). Figure 7 charts the expected GPM and DSM survival curves for this scenario with the generated known Kaplan-Meier curve, and the extrapolated survival over time for log-logistic external additive hazards model and the Weibull internal additive hazards model (the extrapolated non-GPM log-logistic survival is shown as its hazards are used for the external additive hazards extrapolation).

Figure 6:

Survival curves for the best performing external additive hazards and internal additive hazards model in terms of mean survival for scenario with young ages, high survival, medium heterogeneity, high information, Gompertz distribution

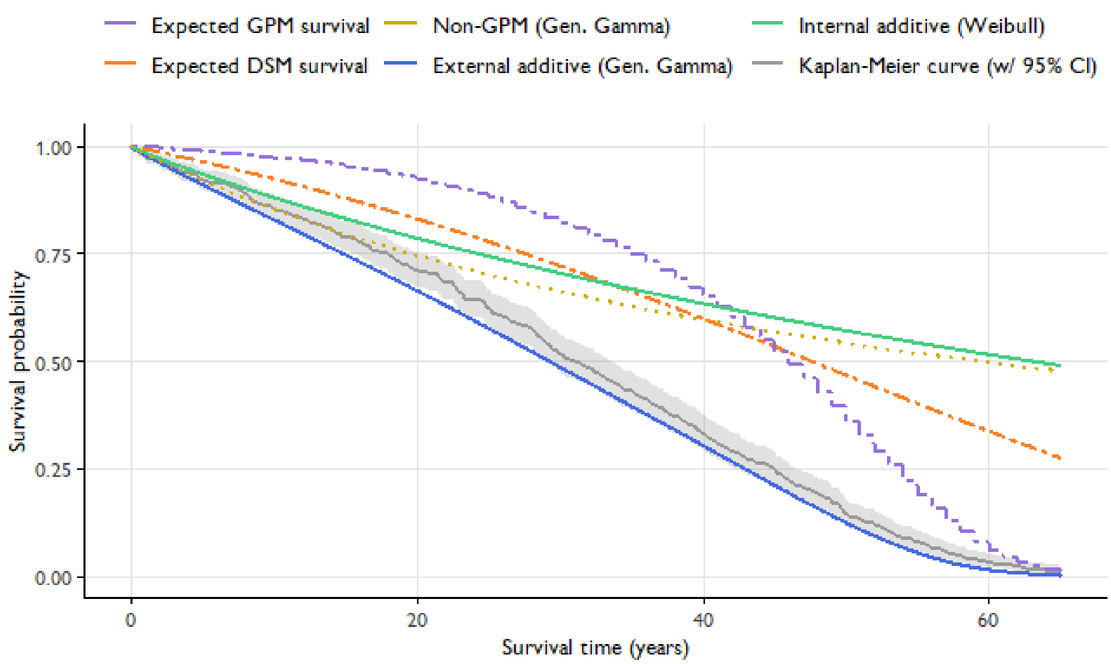
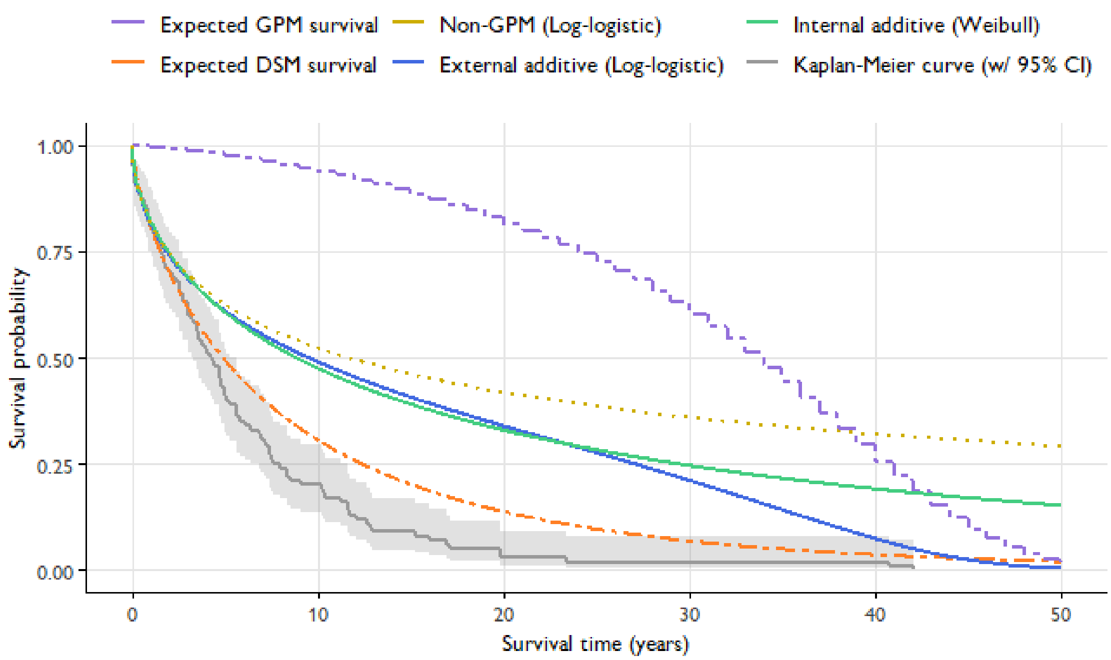


Figure 7:

Survival curves for the best performing external additive hazards and internal additive hazards model in terms of mean survival for scenario with average ages, medium Weibull survival, low heterogeneity and low information



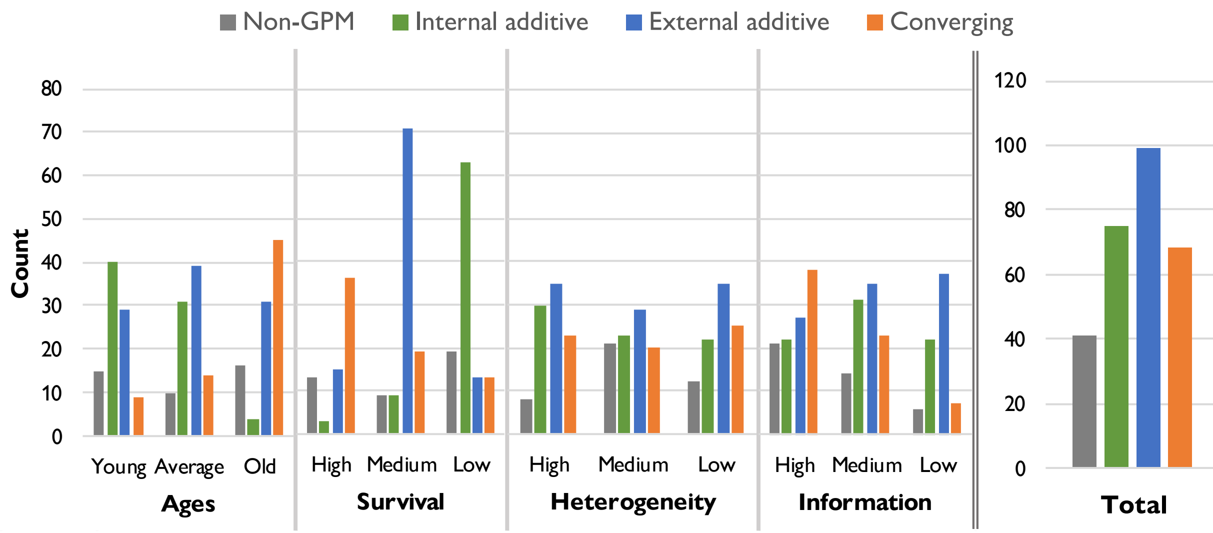
5.2 Survival probability at time *t*

5.2.1 Overview

For survival probability at time t (t = 3 for low survival, t = 15 for medium survival, t = 20 for high survival), the external additive hazards GPM incorporating method performed best in terms of absolute mean bias in the most scenarios when filtering for models that used the same parametric distribution as the distribution used to generate the survival data (99 out of 288 scenarios). The internal additive method ranked the highest in 75 scenarios, the converging hazards method in 68 scenarios and the non-GPM extrapolations in 41 scenarios. A significant result was found in 53% of scenarios (using the Wilcoxon test). Note that for low survival, and either medium or high information scenarios, the estimated probabilities were not extrapolations, but interpolations, as the time t is either at the end or within follow-up. Figure 8 charts the number of times each GPM incorporating ranked best for absolute mean bias of survival probability at time t per dimension (keeping all other dimensions the same) on the left, and in total on the right.

Figure 8:

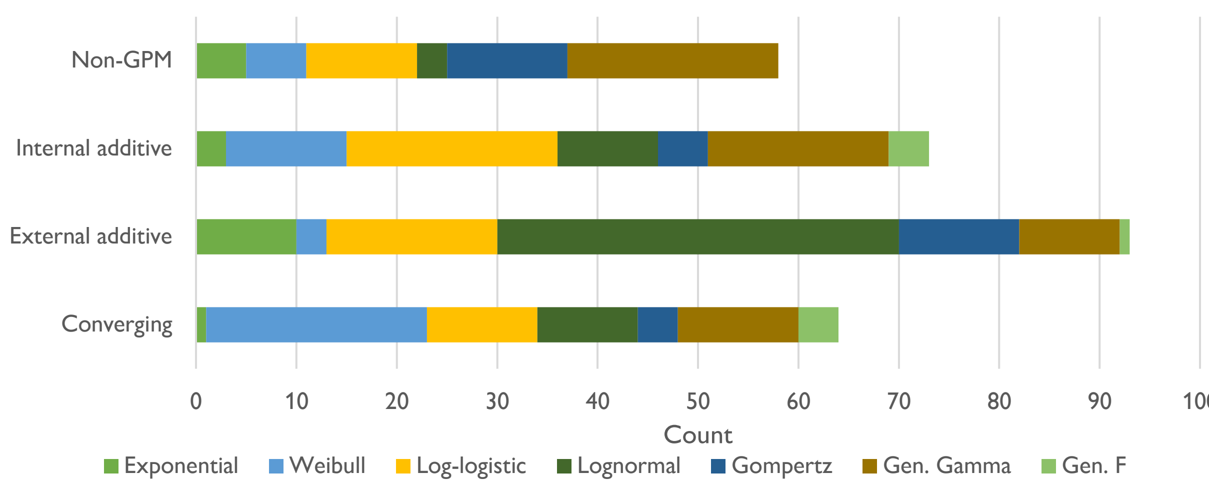
Times each GPM incorporating method ranked best for absolute mean bias of survival probability at time *t* per dimension and total (models filtered based on parametric distribution)



The external additive hazards method also performed best in the most scenarios for survival probability at time *t* when all parametric distributions were considered for analysis (93 out of 288 scenarios). Next, the internal additive hazards method performed the best in 73 scenarios, the converging hazards method in 64 scenarios and the non-GPM extrapolations in 58 scenarios. A significantly best model was found in 14% of the scenarios overall using the Wilcoxon test. In Figure 9 the number of scenarios each method ranked best for survival probability at time *t* per distribution when all distributions were considered is shown.

Figure 9:

Times each GPM incorporating method and distribution ranked best for absolute mean bias of survival probability at time *t* (all distributions considered)



5.2.2 Filtered results

5.2.2.1 Lognormal results

The scenarios with lognormal distributed survival data had the most significant results for survival probability at time *t* (60% of scenarios) when filtering for models that also used the lognormal distribution, thus, its results are presented in Tables 10, 11 and 12. Visualisations of the results for the other distributions can be found in Appendix A.2 for the filtered results. The results are presented in a similar manner to Tables 4, 5 and 6 (see Section 5.1.2 for an explanation). For scenarios that were not an extrapolation (low survival and medium or high information), a * is noted in the column of the results tables.

Table 10:

Legend for filtered colour-coded table and times a GPM incorporating method ranked best for absolute mean bias of survival probability at time *t* for lognormal distributed data. *Amount of times the method ranked best*

	Non-GPM	Internal additive	External additive	Converging
Times method ranked best	6	15	37	13

Table 11:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of survival probability at time t per scenario for lognormal distributed data. Survival and ages as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

		Survival (Lognormal)							
		High		Medium			Low		
		Information		Information			Information		
Age	Heterogeneity	Medium	High	Low	Medium	High	Low	Medium*	High*
Young	Low	▲ 0.008, 0.125	-0.027, 0.03	-0.003, 0.078	-0.002, 0.025	0.003, 0.011	-0.002, 0.059	-0.004, 0.009	-0.004, 0.009
Young	Medium	NA	▲ -0.024, 0.027	-0.001, 0.077	-0.002, 0.026	0.004, 0.011	-0.008, 0.059	-0.004, 0.009	-0.004, 0.009
Young	High	▲ -0.003, 0.044	▲ -0.013, 0.02	-0.001, 0.064	▲ 0.003, 0.024	▲ -0.002, 0.01	-0.001, 0.041	-0.001, 0.008	▲ -0.001, 0.009
Average	Low	▲ -0.027, 0.069	▲ -0.06, 0.06	▲ 0.002, 0.073	▲ -0.003, 0.024	▲ -0.011, 0.014	0.001, 0.06	-0.002, 0.009	▲ -0.002, 0.008
Average	Medium	▲ -0.021, 0.065	▲ -0.055, 0.055	▲ 0.002, 0.071	▲ -0.004, 0.024	▲ -0.012, 0.014	-0.001, 0.058	▲ -0.001, 0.009	▲ -0.001, 0.009
Average	High	▲ 0, 0.044	▲ -0.033, 0.037	▲ 0.004, 0.058	▲ 0.004, 0.023	▲ -0.009, 0.014	0, 0.043	-0.002, 0.008	▲ 0.003, 0.01
Old	Low	▲ 0.02, 0.021	▲ 0.008, 0.011	▲ -0.002, 0.027	▲ -0.009, 0.014	▲ -0.021, 0.021	0.017, 0.061	0.007, 0.012	0, 0.009
Old	Medium	▲ -0.007, 0.018	▲ -0.019, 0.021	▲ -0.006, 0.028	▲ -0.014, 0.017	▲ -0.026, 0.026	▲ 0.005, 0.054	0.005, 0.011	-0.004, 0.009
Old	High	▲ -0.032, 0.035	▲ -0.043, 0.043	▲ -0.016, 0.031	▲ -0.025, 0.027	▲ 0.02, 0.021	▲ -0.013, 0.04	0.004, 0.009	-0.021, 0.021

Table 12:

Filtered colour-coded table for best ranking method for absolute mean bias of survival probability at time t per scenario for lognormal distributed data. Information and heterogeneity as main axes. *(Bias, RMSE),* ▲ = *Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

		Information							
		Low		Medium			High		
		Survival (Lognormal)		Survival (Lognormal)			Survival (Lognormal)		
Heterogeneity	Age	Medium	Low	High	Medium	Low*	High	Medium	Low*
Low	Young	-0.003, 0.078	-0.002, 0.059	0.008, 0.125 ▲	-0.002, 0.025	-0.004, 0.009	-0.027, 0.03	0.003, 0.011	-0.004, 0.009
Low	Average	0.002, 0.073 ▲	0.001, 0.06	-0.027, 0.069 ▲	-0.003, 0.024 ▲	-0.002, 0.009	-0.06, 0.06 ▲	-0.011, 0.014 ▲	-0.002, 0.008 ▲
Low	Old	-0.002, 0.027 ▲	0.017, 0.061	0.02, 0.021 ▲	-0.009, 0.014 ▲	0.007, 0.012	0.008, 0.011 ▲	-0.021, 0.021 ▲	0, 0.009
Medium	Young	-0.001, 0.077	-0.008, 0.059	NA	-0.002, 0.026	-0.004, 0.009	-0.024, 0.027 ▲	0.004, 0.011	-0.004, 0.009
Medium	Average	0.002, 0.071 ▲	-0.001, 0.058	-0.021, 0.065 ▲	-0.004, 0.024 ▲	-0.001, 0.009 ▲	-0.055, 0.055 ▲	-0.012, 0.014 ▲	-0.001, 0.009 ▲
Medium	Old	-0.006, 0.028 ▲	0.005, 0.054 ▲	-0.007, 0.018 ▲	-0.014, 0.017 ▲	0.005, 0.011	-0.019, 0.021 ▲	-0.026, 0.026 ▲	-0.004, 0.009
High	Young	-0.001, 0.064	-0.001, 0.041	-0.003, 0.044 ▲	0.003, 0.024 ▲	-0.001, 0.008	-0.013, 0.02 ▲	-0.002, 0.01 ▲	-0.001, 0.009 ▲
High	Average	0.004, 0.058 ▲	0, 0.043	0, 0.044 ▲	0.004, 0.023 ▲	-0.002, 0.008	-0.033, 0.037 ▲	-0.009, 0.014 ▲	0.003, 0.01 ▲
High	Old	-0.016, 0.031 ▲	-0.013, 0.04 ▲	-0.032, 0.035 ▲	-0.025, 0.027 ▲	0.004, 0.009	-0.043, 0.043 ▲	0.02, 0.021 ▲	-0.021, 0.021

As can be seen in Table 11, the **external additive hazards** method performed best for nearly all the scenarios that had medium survival in scenarios that used lognormal distributed survival data for survival probability at time *t*. In scenarios with high survival and young or average populations, the **external additive hazards** method also performed the best the most often. The **converging hazards** method performed the best for all scenarios with high survival and old ages. For scenarios with low survival and young or average ages, the **internal additive hazards** method performed the best most often. Performance in scenarios with low survival and old ages was mixed. Table 12 shows no distinct patterns in performance between the methods when comparing information and heterogeneity as main axes for scenarios that used the lognormal distribution for survival data.

5.2.2.2 Results for other parametric distributions

There were a few similarities in patterns in performance for survival probability at time *t* across the distributions used to generate the survival data. Over all distributions, the **converging hazards** method is only outperformed once by a **non-GPM** extrapolation out of all the scenarios with high survival and old ages. The results for scenarios with low survival were similar throughout the distributions, where the **internal additive hazards** method performed best the most often for young and average ages, and performance was mixed between the methods for old ages. For scenarios with medium survival, the **external additive hazards** method performed the best the most often for scenarios that used log-logistic (Table 38) and Gompertz (Table 44) distributions, but not for the scenarios that used the Weibull distribution, where performance was mixed between the methods (see Table 35). It was relatively rare for the **internal additive hazards** method to have performed the best in either medium or high survival scenarios across distributions.

5.2.3 Unfiltered results

5.2.3.1 Lognormal results

The lognormal results when all parametric distributions used for modelling were under consideration will be presented in Tables 13, 14 and 15 in order to compare to the filtered lognormal results (which had the most significant results) described in the previous section. For the unfiltered results, a significant result was found in 7% of the scenarios that used the lognormal distribution to generate the survival data. Visualisations for the scenarios with other parametric distributions can be found in Appendix B.2. The tables shown here follow a similar structure to those described in Section 5.1.3 for the unfiltered results.

The unfiltered results for lognormal distributed scenarios for survival probability at time *t* were relatively similar to the filtered results (Tables 11 and 12) in scenarios with medium and low survival. Table 14 shows that in medium survival scenarios, models that used the **external additive hazards** method still performed the best the most often, but the performance was more mixed with other models across the age, information and heterogeneity dimensions. In low survival scenarios with young or average ages, the **internal additive hazards** method performed the best most often. In low survival scenarios with old ages performance was relatively mixed between the methods (see Table 14), as was seen in Table 11. However, the scenarios with high survival no longer showed a preference for the **external additive hazards** method in scenarios with young or average ages where performance was now mixed between the methods. In scenarios with high survival and old ages, the **converging hazards** method was now outperformed by **non-GPM** extrapolations in the high information scenarios. With heterogeneity and information used as main axes the results looked similar (see Table 15 and Table 12), as there were no distinct patterns in performance between the methods.

5.2.3.2 Results for other parametric distributions

Across the scenarios using different parametric distributions to generate survival data, results looked relatively similar for high and low survival scenarios for survival probability at time *t* when all parametric distributions used for modelling were considered in analysis (see Appendix B.2). In scenarios with high survival and young or average ages performance was mixed between the methods across distributions. For scenarios with high survival and old ages the **converging hazards** method nearly always performed the best in medium information scenarios, and the **non-GPM** extrapolations always performed the best in high information scenarios across distributions. It was still relatively rare for the **internal additive hazards** method to perform the best in high and medium survival scenarios for the unfiltered results, as was seen in the filtered results across distributions. In low survival scenarios, the **internal additive hazards** method still performed the best the most often in scenarios with young or average ages across distributions when compared to the filtered results.

For scenarios with medium survival, however, there no longer was a preference for the **external additive hazards** method in the results for scenarios that used the Weibull (Table 71) and Gompertz (Table 80) distributions for survival data. Finally, for scenarios that used the Weibull distribution for generating survival data, Weibull models did not perform the best in the most scenarios (13 scenarios out of 72), but rather the generalised gamma models were selected the most often (25 scenarios).

Table 13:

Legend for colour-coded table and times a model ranked best for absolute mean bias of survival probability at time *t* for lognormal distributed data. *Amount of times the model ranked best*

	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F	Total
Non-GPM	*2*	*1*	*5*	*3*	*2*	*4*	*0*	***17***
Internal additive	*0*	*1*	*4*	*7*	*0*	*3*	*4*	***19***
External additive	*2*	*0*	*0*	*12*	*4*	*4*	*0*	***22***
Converging	*0*	*5*	*3*	*2*	*1*	*1*	*2*	***14***
Total	***4***	***7***	***12***	***24***	***7***	***12***	***6***	

Table 14:

Colour-coded table for best ranking model for absolute mean bias of survival probability at time *t* per scenario for lognormal distributed data. Survival and ages as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

Survival (Lognormal)

		High		Medium			Low		
		Information		*Information*			*Information*		
Age	*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium*	High*
Young	Low	*0.008, 0.125*	*-0.002, 0.023*	*-0.003, 0.078*	*0.002, 0.046*	*0.001, 0.011*	*-0.002, 0.059*	*-0.002, 0.006*	*-0.004, 0.009*
Young	Medium	*0.272, 0.272*	*0.001, 0.022*	*-0.001, 0.077*	*0.001, 0.046*	*0.002, 0.01*	*-0.008, 0.059*	*-0.002, 0.006*	*-0.004, 0.009*
Young	High	▲ *-0.003, 0.044*	*0.001, 0.019*	*-0.001, 0.064*	*0.003, 0.024*	*-0.002, 0.01*	*-0.001, 0.061*	*-0.001, 0.006*	*-0.001, 0.009*
Average	Low	*-0.004, 0.132*	*0.016, 0.032*	*0.002, 0.073*	*0.003, 0.024*	*0.004, 0.012*	*0.001, 0.06*	*0, 0.006*	*-0.002, 0.008*
Average	Medium	*0, 0.12*	*0.018, 0.032*	*0.002, 0.071*	*-0.003, 0.024*	*0.006, 0.013*	*-0.001, 0.058*	*0.001, 0.006*	*-0.001, 0.009*
Average	High	▲ *0, 0.044*	*-0.012, 0.021*	*0.004, 0.058*	*0.001, 0.034*	*-0.009, 0.014*	*0, 0.043*	*-0.001, 0.006*	*0.001, 0.008*
Old	Low	*-0.002, 0.028*	▲ *0.005, 0.027*	*-0.002, 0.027*	▲ *-0.002, 0.026*	*0.005, 0.009*	*-0.004, 0.117*	*-0.003, 0.01*	*0, 0.009*
Old	Medium	*-0.007, 0.018*	*-0.004, 0.028*	*0.005, 0.05*	*0.001, 0.025*	*0.001, 0.008*	*-0.002, 0.109*	*-0.005, 0.011*	*-0.002, 0.009*
Old	High	*0.029, 0.035*	*0.025, 0.045*	*0.01, 0.06*	▲ *-0.001, 0.018*	*0.002, 0.01*	*0, 0.04*	*0, 0.006*	*0.007, 0.013*

Table 15:

Colour-coded table for best ranking model for absolute mean bias of survival probability at time t per scenario for lognormal distributed data. Information and heterogeneity as main axes. *(Bias, RMSE), ▲ = Significant result, * denotes scenarios that are not extrapolations*

		Information: Low		Information: Medium			Information: High		
		Survival (Lognormal)		*Survival (Lognormal)*			*Survival (Lognormal)*		
Heterogeneity	Age	Medium	Low	High	Medium	Low*	High	Medium	Low*
Low	Young	*-0.003, 0.078*	*-0.002, 0.059*	*0.008, 0.125*	*0.002, 0.046*	*-0.002, 0.006*	*-0.002, 0.023*	*0.001, 0.011*	*-0.004, 0.009*
Low	Average	*0.002, 0.073*	*0.001, 0.06*	*-0.004, 0.132*	*0.003, 0.024*	*0, 0.006*	*0.016, 0.032*	*0.004, 0.012*	*-0.002, 0.008*
Low	Old	*-0.002, 0.027*	*-0.004, 0.117*	*-0.002, 0.028*	▲ *-0.002, 0.026*	*-0.003, 0.01*	▲ *0.005, 0.027*	*0.005, 0.009*	*0, 0.009*
Medium	Young	*-0.001, 0.077*	*-0.008, 0.059*	*0.272, 0.272*	*0.001, 0.046*	*-0.002, 0.006*	*0.001, 0.022*	*0.002, 0.01*	*-0.004, 0.009*
Medium	Average	*0.002, 0.071*	*-0.001, 0.058*	*0, 0.12*	*-0.003, 0.024*	*0.001, 0.006*	*0.018, 0.032*	*0.006, 0.013*	*-0.001, 0.009*
Medium	Old	*0.005, 0.05*	*-0.002, 0.109*	*-0.007, 0.018*	*0.001, 0.025*	*-0.005, 0.011*	*-0.004, 0.028*	*0.001, 0.008*	*-0.002, 0.009*
High	Young	*-0.001, 0.064*	*-0.001, 0.061*	▲ *-0.003, 0.044*	*0.003, 0.024*	*-0.001, 0.006*	*0.001, 0.019*	*-0.002, 0.01*	*-0.001, 0.009*
High	Average	*0.004, 0.058*	*0, 0.043*	▲ *0, 0.044*	*0.001, 0.034*	*-0.001, 0.006*	*-0.012, 0.021*	*-0.009, 0.014*	*0.001, 0.008*
High	Old	*0.01, 0.06*	*0, 0.04*	*0.029, 0.035*	▲ *-0.001, 0.018*	*0, 0.006*	*0.025, 0.045*	*0.002, 0.01*	*0.007, 0.013*

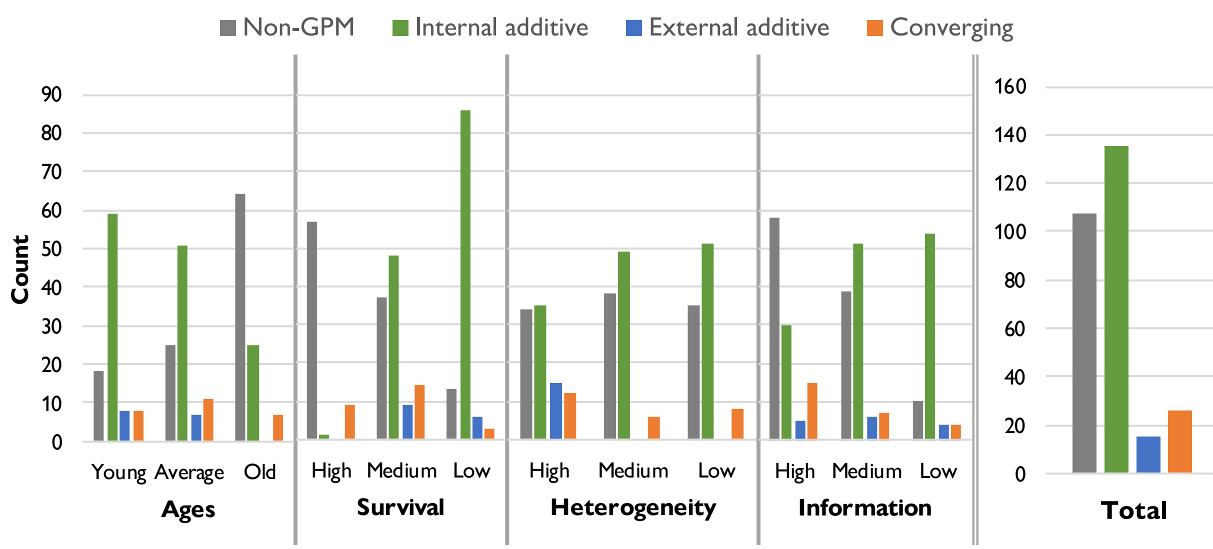
5.3 RMST

5.3.1 Overview

Overall, for absolute mean bias RMST (calculated until the end of follow-up, which was 1 year for low information, 3 years for medium information, and 10 years for high information scenarios), the internal additive hazards method performed best the most often when filtering for models that used the same parametric distribution as the generated survival data (135 out of 288 scenarios). Next, the non-GPM extrapolations performed the best in 107 scenarios, the converging hazards method in 26 scenarios, and the external additive hazards method in 15 scenarios. A significant result was found in 57% of scenarios overall when models were filtered based on the parametric distribution used for modelling and generating the data. Figure 10 charts the number of times each GPM incorporating ranked best for absolute mean bias of RMST per dimension (keeping all other dimensions the same) on the left, and in total on the right.

Figure 10:

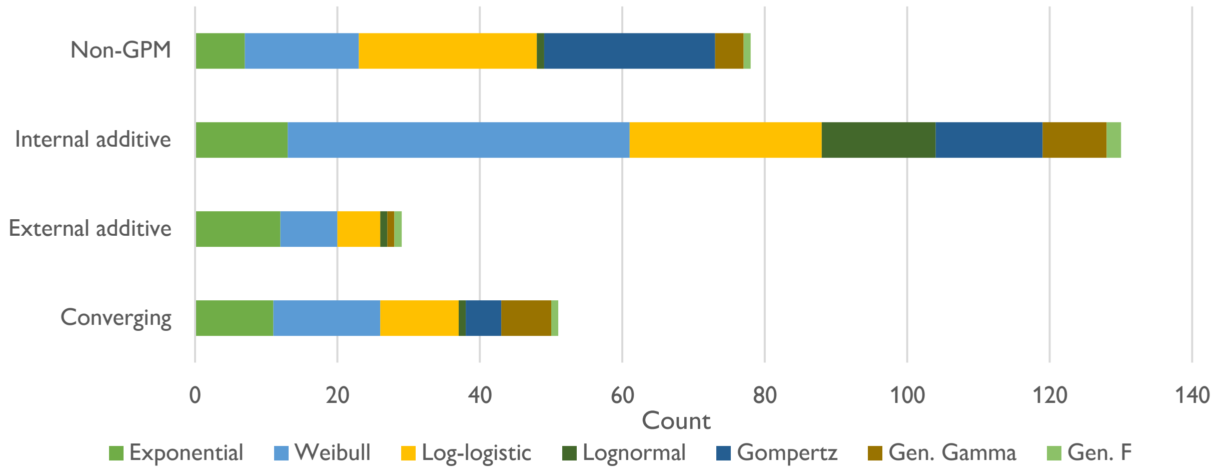
Times each GPM incorporating method ranked best for absolute mean bias of RMST per dimension and **total** (models filtered based on parametric distribution)



When all models using all distributions were considered for analysis, the internal additive hazards method also performed best the most for RMST (130 out of 288 scenarios). The non-GPM extrapolations performed best in 75 scenarios, the converging hazards method in 51 scenarios, and the external additive hazards method in 29 scenarios. In 20% of scenarios a significant result was found using the Wilcoxon test for the unfiltered results. In Figure 11 the number of scenarios where each GPM incorporating method ranked best per distribution is shown for RMST when all distributions are considered.

Figure 11:

Times each GPM incorporating method and distribution ranked best for absolute mean bias of RMST (all distributions considered)



5.3.2 Exploring performance of the non-GPM models

Before continuing onto the scenario-specific results for RMST, the performance of non-GPM *inter*polations for RMST should be discussed, as there were 78 scenarios in the unfiltered results where non-GPM interpolations had a lower absolute mean bias, which was more often than both the converging hazards method (51 scenarios) and the external additive hazards method (21 scenarios). There were a combined 104 out of 576 scenarios where non-GPM extrapolations had a lower absolute mean bias than models that incorporate GPM for the other two estimands. However, for overall mean survival and survival probability at time *t*, non-GPM extrapolations were selected as the best the least often compared to the other methods.

Across all three estimands, non-GPM extrapolations performed the best in 186 out of 864 scenarios across estimands in results that were not filtered for matching distributions used to generate survival data and used for modelling, yet only 55 (30%) of these results were significant using the Wilcoxon test. The converging hazards method performed the best in 179 out of 864 scenarios across estimands, yet only 24 (13%) of these results were significant using the Wilcoxon test. Within the **in**significant results for non-GPM extrapolations, the converging hazards method as second best in 90 out of 131 scenarios across estimands, while within **in**significant results for converging hazards extrapolations, the non-GPM extrapolations were second best in 27 out of 155 scenarios across estimands.

5.3.3 Filtered results

5.3.3.1 Lognormal results

For the filtered results where only models that used the same parametric distribution as the distribution used to generate the survival data were under consideration, the results for the scenarios that used the lognormal distribution had the most significant results (72% using the Wilcoxon test), and thus its results will be shown here. Visualisations for the other distributions can be found in Appendix A.3. The results are presented in a similar manner to Tables 4, 5 and 6 (see Section 5.1.2 for an explanation) in Tables 16, 17 and 18.

Overall, there were two distinct patterns that can be seen in Tables 17 and 18 for the scenarios that used the lognormal distribution to generate the survival data. Table 17 shows that **non-GPM** extrapolations nearly always performed best in high survival scenarios, and as survival becomes lower and ages become younger, the **internal additive hazards** method performed best more often in the lognormal distributed scenarios. Secondly, as can be seen in Table 18, the **internal additive hazards** method always performed best in scenarios with low information. Then, as information becomes higher, **non-GPM** extrapolations performed better more often.

Table 16:

Legend for filtered colour-coded table and times a GPM incorporating method ranked best for absolute mean bias of RMST for lognormal distributed data. *Amount of times the method ranked best*

	Non-GPM	Internal additive	External additive	Converging
Times method ranked best	*26*	*42*	*0*	*3*

Table 17:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of RMST per scenario for lognormal distributed data. Survival and ages as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

		Survival (Lognormal)							
		High		Medium			Low		
		Information		*Information*			*Information*		
Age	*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium	High
Young	Low	▲ *-0.001, 0.004*	*-0.016, 0.026*	▲ *-0.011, 0.011*	▲ *-0.021, 0.023*	▲ *-0.006, 0.034*	▲ *-0.016, 0.016*	▲ *-0.037, 0.037*	▲ *-0.041, 0.041*
Young	Medium	*NA*	*-0.017, 0.027*	▲ *-0.012, 0.012*	▲ *-0.022, 0.024*	▲ *-0.002, 0.034*	▲ *-0.017, 0.017*	▲ *-0.035, 0.035*	▲ *-0.037, 0.038*
Young	High	*0.004, 0.009*	*-0.035, 0.042*	*-0.008, 0.012*	▲ *-0.005, 0.017*	*-0.02, 0.042*	*-0.02, 0.02*	▲ *-0.028, 0.029*	▲ *-0.008, 0.034*
Average	Low	▲ *-0.003, 0.006*	▲ *-0.041, 0.045*	▲ *-0.01, 0.011*	▲ *-0.012, 0.017*	*-0.064, 0.066*	▲ *-0.015, 0.015*	▲ *-0.032, 0.032*	▲ *-0.028, 0.029*
Average	Medium	▲ *-0.004, 0.006*	▲ *-0.044, 0.047*	▲ *-0.01, 0.011*	▲ *-0.014, 0.019*	*-0.068, 0.07*	▲ *-0.016, 0.016*	▲ *-0.029, 0.03*	▲ *-0.019, 0.023*
Average	High	*-0.008, 0.011*	*-0.073, 0.075*	▲ *-0.008, 0.012*	▲ *0.001, 0.017*	*-0.05, 0.057*	▲ *-0.019, 0.02*	▲ *-0.023, 0.025*	▲ *0.049, 0.055*
Old	Low	▲ *-0.019, 0.02*	▲ *-0.211, 0.211*	▲ *0.002, 0.007*	*-0.04, 0.04*	*-0.11, 0.11*	▲ *-0.003, 0.007*	▲ *0.019, 0.021*	▲ *-0.016, 0.026*
Old	Medium	▲ *-0.021, 0.022*	▲ *-0.197, 0.197*	▲ *0.001, 0.007*	*-0.043, 0.043*	*-0.106, 0.106*	▲ *-0.005, 0.008*	▲ *0.022, 0.023*	*-0.019, 0.028*
Old	High	*-0.029, 0.029*	▲ *-0.187, 0.187*	▲ *0, 0.011*	*-0.041, 0.042*	▲ *-0.12, 0.12*	▲ *-0.012, 0.015*	▲ *0.03, 0.033*	*-0.065, 0.066*

Table 18:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of RMST per scenario for lognormal distributed data. Information and heterogeneity as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

		Information: Low		Information: Medium			Information: High		
		Survival (Lognormal)		*Survival (Lognormal)*			*Survival (Lognormal)*		
Heterogeneity	**Age**	Medium	Low	High	Medium	Low	High	Medium	Low
Low	Young	-0.011, 0.011 ▲	-0.016, 0.016 ▲	-0.001, 0.004 ▲	-0.021, 0.023 ▲	-0.037, 0.037 ▲	-0.016, 0.026	-0.006, 0.034 ▲	-0.041, 0.041 ▲
Low	Average	-0.01, 0.011 ▲	-0.015, 0.015 ▲	-0.003, 0.006 ▲	-0.012, 0.017 ▲	-0.032, 0.032 ▲	-0.041, 0.045 ▲	-0.064, 0.066	-0.028, 0.029 ▲
Low	Old	0.002, 0.007 ▲	-0.003, 0.007 ▲	-0.019, 0.02 ▲	-0.04, 0.04	0.019, 0.021 ▲	-0.211, 0.211 ▲	-0.11, 0.11	-0.016, 0.026 ▲
Medium	Young	-0.012, 0.012 ▲	-0.017, 0.017 ▲	NA	-0.022, 0.024 ▲	-0.035, 0.035 ▲	-0.017, 0.027	-0.002, 0.034 ▲	-0.037, 0.038 ▲
Medium	Average	-0.01, 0.011 ▲	-0.016, 0.016 ▲	-0.004, 0.006 ▲	-0.014, 0.019 ▲	-0.029, 0.03 ▲	-0.044, 0.047 ▲	-0.068, 0.07	-0.019, 0.023 ▲
Medium	Old	0.001, 0.007 ▲	-0.005, 0.008 ▲	-0.021, 0.022 ▲	-0.043, 0.043	0.022, 0.023 ▲	-0.197, 0.197 ▲	-0.106, 0.106	-0.019, 0.028
High	Young	-0.008, 0.012	-0.02, 0.02	0.004, 0.009	-0.005, 0.017 ▲	-0.028, 0.029 ▲	-0.035, 0.042	-0.02, 0.042	-0.008, 0.034 ▲
High	Average	-0.008, 0.012 ▲	-0.019, 0.02 ▲	-0.008, 0.011	0.001, 0.017 ▲	-0.023, 0.025 ▲	-0.073, 0.075	-0.05, 0.057	0.049, 0.055 ▲
High	Old	0, 0.011 ▲	-0.012, 0.015 ▲	-0.029, 0.029	-0.041, 0.042	0.03, 0.033 ▲	-0.187, 0.187 ▲	-0.12, 0.12 ▲	-0.065, 0.066

5.3.3.2 Results for other parametric distributions

Both patterns described in the previous sections for scenarios that use lognormal distributed survival data generally held across distributions (see Appendix A.3), with the largest difference being that the **internal additive hazards** method no longer always outperformed the other methods in scenarios with low information, although it still performed the best the most in those scenarios across distributions. Furthermore, in scenarios with old ages and medium survival, there was a stronger preference for the **non-GPM** extrapolations, which performed best in 27 out of 36 scenarios with those characteristics across distributions. The **external additive hazards** method only performed the best in 15 scenarios out of 72 for RMST when filtering models for the parametric distribution used to generate the survival data. 12 out of those 15 scenarios were in the Weibull distributed scenarios (see Table 48), where the **external additive hazards** method only performed best in scenarios with high heterogeneity, except for scenarios with high survival or old ages. The other 3 scenarios where the **external additive hazards** method were in the log-logistic distributed scenarios (see Table 51), where it also only ever performed best in scenarios with high heterogeneity and medium survival.

5.3.4 Unfiltered results

5.3.4.1 Lognormal results

For comparison with the filtered lognormal results (which had the most significant results) described in the previous section, the results for the lognormal distributed scenarios when all parametric distributions used for modelling were included in the analysis will be shown here. In the lognormal distributed scenarios, a significant result was found in 26% of scenarios when all parametric distributions are considered. Visualisations for the scenarios with other parametric distributions can be found in Appendix B.3. Tables 19, 20 and 21 follow a similar structure to those described in Section 5.1.3 for the unfiltered results.

Compared to the filtered results for lognormal distributed scenarios, similar patterns can be seen in both Tables 17 and 20, and Tables 18 and 21. Table 20 shows that the **non-GPM** extrapolations nearly always performed best in high survival scenarios when models using all parametric distributions were considered. Once again, the **internal additive hazards** method performed best more often as survival becomes lower and ages become younger in the lognormal distributed scenarios. Furthermore, when comparing Tables 18 and 22 the **internal additive hazards** method nearly always performed best in scenarios with low information in both tables. However, for the unfiltered results (Table 21), performance was more mixed between all GPM methods in scenarios with medium or high information, where in the filtered results (Table 18) the mix was mostly between the **internal additive hazards** method and **non-GPM** extrapolations.

Table 19:

Legend for colour-coded table and times a model ranked best for absolute mean bias of RMST for lognormal distributed data. *Amount of times the model ranked best*

	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F	Total
Non-GPM	*2*	*6*	*5*	*0*	*5*	*2*	*0*	*20*
Internal additive	*2*	*15*	*9*	*4*	*4*	*3*	*0*	*37*
External additive	*2*	*4*	*0*	*0*	*0*	*0*	*0*	*6*
Converging	*0*	*4*	*5*	*0*	*0*	*0*	*0*	*9*
Total	*6*	*29*	*19*	*4*	*9*	*5*	*0*	

5.3.4.2 Results for other parametric distributions

Results across the scenarios with different parametric distributions for RMST when all parametric distributions used for modelling were considered in analysis looked relatively similar (see Appendix B.3). The patterns described in the previous section for lognormal distributed scenarios for the **internal additive hazards** method held across distributions, where it performed best more often as survival becomes lower and ages become younger. Compared to the filtered results across distributions in scenarios where the **non-GPM** extrapolations generally performed best the most (high survival scenarios, and scenarios with medium survival and old ages), the **converging hazards** method performed best more often in the unfiltered results (for example, compare the filtered and unfiltered results for Weibull distributed scenarios in Tables 47 and 83).

Table 20:

Colour-coded table for best ranking model for absolute mean bias of of RMST per scenario for lognormal distributed data. Survival and ages as main axes. *(Bias, RMSE),* ▲ *= Significant result (Wilcoxon)*

Age	Heterogeneity	Survival (Lognormal): High — Information: Medium	High — High	Medium — Low	Medium — Medium	Medium — High	Low — Low	Low — Medium	Low — High
Young	Low	*0, 0.004*	*0.003, 0.018*	*-0.002, 0.016*	*0.002, 0.013*	*-0.005, 0.032*	*-0.012, 0.013*	*-0.009, 0.013* ▲	*-0.024, 0.024* ▲
Young	Medium	*-0.002, 0.004*	*0.004, 0.019*	*-0.004, 0.016*	*0.003, 0.014*	*-0.002, 0.032*	*-0.013, 0.013*	*-0.005, 0.012*	*-0.011, 0.017* ▲
Young	High	*0.001, 0.008*	*-0.005, 0.03*	*0, 0.009*	*-0.005, 0.017*	*-0.014, 0.04*	*0, 0.01*	*-0.011, 0.016* ▲	*-0.008, 0.034*
Average	Low	*-0.001, 0.005*	*-0.003, 0.023*	*-0.004, 0.007* ▲	*0.001, 0.012* ▲	*-0.025, 0.039* ▲	*-0.012, 0.012* ▲	*-0.005, 0.012* ▲	*-0.014, 0.016* ▲
Average	Medium	*-0.001, 0.005*	*-0.003, 0.023*	*-0.004, 0.007* ▲	*0.001, 0.012*	*-0.023, 0.041* ▲	*-0.012, 0.012* ▲	*0, 0.012* ▲	*-0.001, 0.015*
Average	High	*0, 0.008*	*0.002, 0.031*	*0.001, 0.009*	*-0.001, 0.017*	*-0.003, 0.043*	*0.001, 0.01*	*-0.007, 0.014* ▲	*0.036, 0.046* ▲
Old	Low	*-0.007, 0.009*	*-0.026, 0.03*	*0.002, 0.007*	*-0.001, 0.014*	*-0.009, 0.03*	*-0.002, 0.006* ▲	*0.009, 0.014* ▲	*0.002, 0.024*
Old	Medium	*-0.008, 0.01*	*-0.027, 0.031*	*0.001, 0.007*	*-0.001, 0.014*	*-0.004, 0.03*	*-0.002, 0.011*	*0, 0.017*	*-0.006, 0.027*
Old	High	*0, 0.018*	*-0.022, 0.055*	*0, 0.01*	*-0.001, 0.022* ▲	*0.009, 0.035*	*-0.003, 0.01*	*0.006, 0.019*	*-0.022, 0.038*

Finally, models using the Weibull distribution ranked highest a relatively high amount of times for RMST compared to other estimands where performance across distributions used for modelling was more evenly spread. In total, models that used the Weibull distribution were selected as the best in 87 out of 288 scenarios, and were selected as the best the most often in scenarios where either the Gompertz, log-logistic or lognormal distribution was used to generate survival data. However, in scenarios where the Weibull distribution was used for generating data, models that used the log-logistic distribution performed the best in 18 scenarios, and models that used the Weibull distribution performed best in 16 scenarios.

Table 21:

Colour-coded table for best ranking model for absolute mean bias of RMST per scenario for lognormal distributed data. Information and heterogeneity as main axes. *(Bias, RMSE),* ▲ *= Significant result (Wilcoxon)*

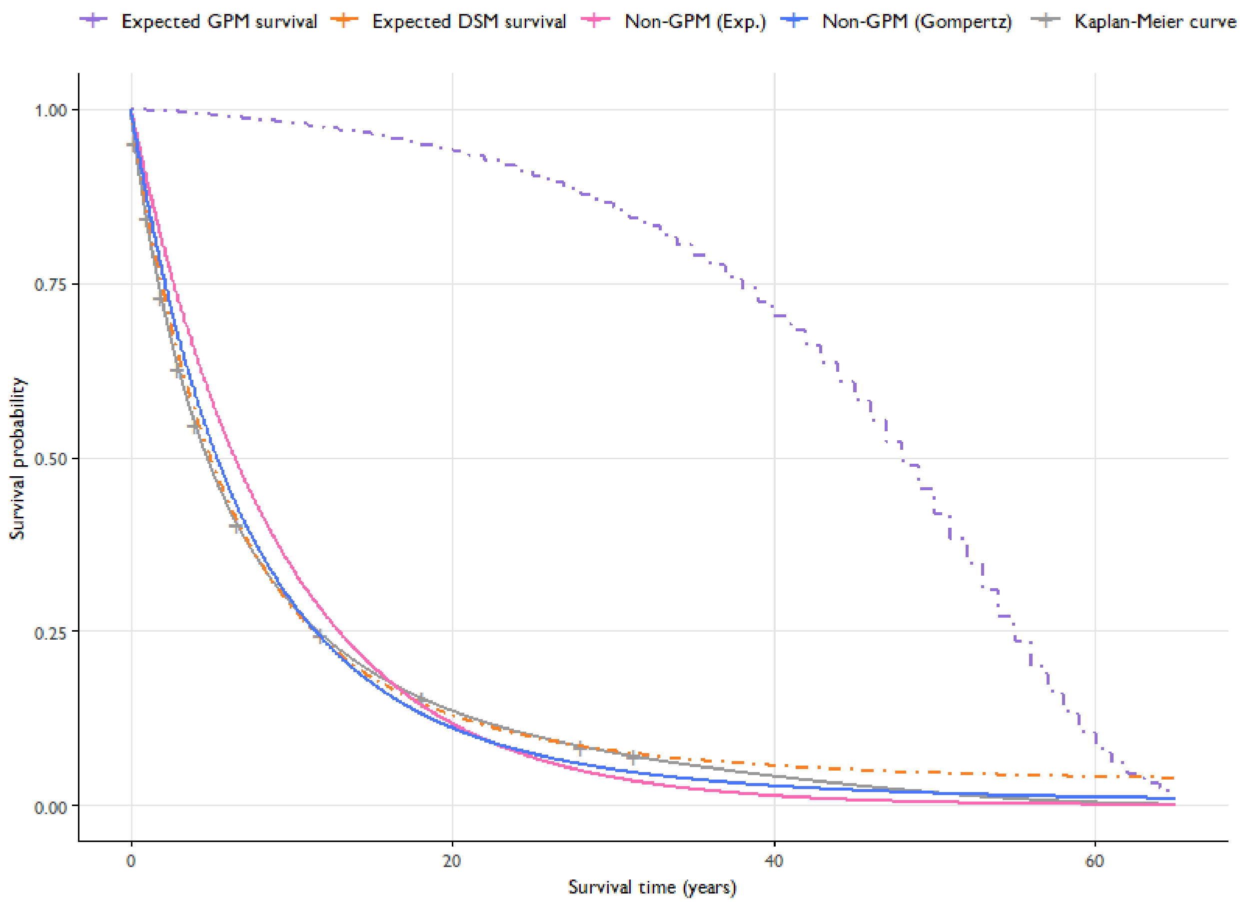
			Information							
			Low		Medium			High		
			Survival (Lognormal)		*Survival (Lognormal)*			*Survival (Lognormal)*		
Heterogeneity	Age		Medium	Low	High	Medium	Low	High	Medium	Low
Low	Age	Young	-0.002, 0.016	-0.012, 0.013	0, 0.004	0.002, 0.013	-0.009, 0.013 ▲	0.003, 0.018	-0.005, 0.032	-0.024, 0.024 ▲
Low	Age	Average	-0.004, 0.007 ▲	-0.012, 0.012 ▲	-0.001, 0.005	0.001, 0.012 ▲	-0.005, 0.012 ▲	-0.003, 0.023	-0.025, 0.039 ▲	-0.014, 0.016 ▲
Low	Age	Old	0.002, 0.007	-0.002, 0.006 ▲	-0.007, 0.009	-0.001, 0.014	0.009, 0.014 ▲	-0.026, 0.03	-0.009, 0.03	0.002, 0.024
Medium	Age	Young	-0.004, 0.016	-0.013, 0.013	-0.002, 0.004	0.003, 0.014	-0.005, 0.012	0.004, 0.019	-0.002, 0.032	-0.011, 0.017 ▲
Medium	Age	Average	-0.004, 0.007 ▲	-0.012, 0.012 ▲	-0.001, 0.005	0.001, 0.012	0, 0.012 ▲	-0.003, 0.023	-0.023, 0.041 ▲	-0.001, 0.015
Medium	Age	Old	0.001, 0.007	-0.002, 0.011	-0.008, 0.01	-0.001, 0.014	0, 0.017	-0.027, 0.031	-0.004, 0.03	-0.006, 0.027
High	Age	Young	0, 0.009	0, 0.01	0.001, 0.008	-0.005, 0.017	-0.011, 0.016 ▲	-0.005, 0.03	-0.014, 0.04	-0.008, 0.034
High	Age	Average	0.001, 0.009	0.001, 0.01	0, 0.008	-0.001, 0.017	-0.007, 0.014 ▲	0.002, 0.031	-0.003, 0.043	0.036, 0.046 ▲
High	Age	Old	0, 0.01	-0.003, 0.01	0, 0.018	-0.001, 0.022 ▲	0.006, 0.019	-0.022, 0.055	0.009, 0.035	-0.022, 0.038

5.3.5 Exploring discrepancies in parametric distribution for simulation and for modelling

To explore what occurred in scenarios where the parametric distribution of the best performing model did not correspond with the parametric distribution used to generate the data, a scenario that used the Gompertz distribution and where models using other distributions performed best for all three estimands was selected to explore further. This was the scenario with young ages, medium survival, medium heterogeneity and high information. However, deviating from the high information scenario, a trial size of 1,000,000 patients, a very low level of right censoring and follow-up length of 100 was used to determine whether Gompertz models would outperform the models using other distributions when the dataset becomes larger. In Figure 12, the survival curve for the best performing model overall (the non-GPM exponential model), and the best performing model using the Gompertz distribution for the bias of mean survival (which was also the model without GPM adjustment) is shown alongside the KM-curve and the expected DSM and GPM survival.

Figure 12:

Survival curves for the best performing model overall and for best performing model using the Gompertz distribution for mean survival for scenario with young ages, medium Gompertz survival, medium heterogeneity and high information (trial size increased to 1,000,000 and almost no censoring)



5.4 Summary

To summarize, in results where models were filtered based on the parametric distribution used for generating the survival data and for modelling, the methods that incorporate GPM outperformed non-GPM extrapolations in 264 out of 288 scenarios for mean survival, 242 scenarios for survival probability at time *t*, and 176 scenarios for RMST (an average of 79% of scenarios across estimands). For results where all parametric distributions were considered, the methods that incorporate GPM outperformed non-GPM extrapolations in 238 out of 288 scenarios for mean survival, 230 scenarios for survival probability at time *t*, and 210 scenarios for RMST (an average of 78% of scenarios across estimands).

6 Discussion

The results show that incorporation of GPM information into survival extrapolations is important in certain situations, as there was a large number of tested scenarios where an extrapolation that incorporated GPM outperformed a non-GPM extrapolation (an average of 79% of scenarios across estimands when models were filtered based on the parametric distributions used for generating survival data, and an average of 78% of scenarios across estimands when all models were considered for analysis). However, between the GPM incorporating methods, patterns between which method performed best were not as clear as was hoped for at the start of this research. This chapter will start with theoretical interpretations for the identified patterns in performance and further interpretation of certain results that contradict current knowledge from literature. Next, the limitations and the strengths of the study will be discussed and related to previous literature found. Finally, to conclude this chapter, the identified patterns in performance will be discussed and used to draw recommendations for guidance for HTA submissions and recommendations for future research will be made.

6.1 Interpretation of results

In this section the results of the study will be interpreted further and possible reasons for certain results being contradictory to current knowledge or otherwise unexpected will be identified.

6.1.1 Performance of the external additive hazards method

In general, the external additive hazards method extrapolated the best in the most scenarios for overall mean survival and survival probability at time *t* in both the filtered and unfiltered results (RMST was an interpolation, as it was calculated until the end of follow-up). However, in literature, most authors that compared GPM incorporating methods recommend using an internal additive hazards approach (more commonly referred to as relative survival models) (Andersson et al., 2013; Palmer et al., 2023; Rutherford et al., 2020). To explore possible reasons why, two scenarios were selected for further exploration, as described in Section 5.1.4.

In Figure 6, the internal additive hazards model follows the known Kaplan-Meier curve relatively closely until at some point it starts overestimating the survival after around 7 – 10, while the external additive hazards model follows the Kaplan-Meier curve more closely until the end of the data (when patients have reached age 100). Figure 7 shows that both models extrapolated rather poorly. However, the external additive hazards model starts adjusting for GPM more strongly after around 20 years when comparing its extrapolation to its non-GPM counterpart. This indicates that the internal additive hazards method does not account enough for GPM at the tail-end of the data when GPM becomes higher, while the external additive hazards method does. This could be due to the internal additive hazards method only capturing GPM hazards of patients during the trial, meaning GPM hazards will be the same for all patients that are censored at the end of follow-up. Thus, the internal additive hazards method could lead to a bias in scenarios where many patients are censored at the end of follow-up, as the effect of GPM hazards is underestimated if the GPM hazards are low at that time.

For example, for the dataset shown in Figure 6 (using high survival and young ages), 384 out of 500 patients were still alive at the end of follow-up (10 years), where expected GPM survival is still above 95%. For the dataset shown in Figure 7 (using medium survival and average ages), 71 out of 100 patients were still alive at the end of follow-up (1 year), where expected GPM survival is also still above 95%. Since the external additive hazards method uses the entire life table to calculate GPM hazards after fitting a model, rather than only using a background hazard calculated for each patient before fitting a model, this bias is not present. This would also explain why the internal additive hazards method generally only performed best in low survival scenarios across estimands, as in those scenarios, patients are less likely to still be alive at the end of follow-up and GPM has less of an effect on overall ACM survival.

6.1.2 Performance of non-GPM extrapolations

Next, the performance of non-GPM extrapolations should be discussed, as although there seems to be a consensus in the literature that incorporating GPM information should increase accuracy of a survival model, there were several scenarios in which a non-GPM extrapolation had a lower absolute mean bias for an estimand than a model that did incorporate GPM information. Across estimands and survival distributions used to generate survival data, non-GPM extrapolations generally only performed the best in scenarios with high survival or old ages, where the other method that often performed the best was the converging hazards method. Since the converging hazards method takes the lowest hazards from either the GPM hazards or the fitted hazards from the non-GPM models, the non-GPM extrapolations and converging hazards extrapolations should look relatively similar in scenarios where DSM survival is high, and GPM hazards are the most important cause of mortality.

This is confirmed not only by the converging hazards often being selected as the best performing method in scenarios similar to scenarios where non-GPM extrapolations were selected as the best, but also by the relatively low number of significant results for the non-GPM extrapolations and converging hazards extrapolations. Furthermore, in many scenarios where the non-GPM extrapolations performed best, but results were *not* significant using the Wilcoxon test, models using the converging hazards method were often second best (as described in Section 5.3.2), meaning there was no significant difference between the extrapolations that did not use GPM information and the converging hazards extrapolations. Thus, the converging hazards method seems to be an adequate substitute in situations where non-GPM extrapolations performed the best, but in situations where the converging hazards method performed the best, non-GPM extrapolations are less likely to be an adequate substitute.

6.1.3 Discrepancies in parametric distribution for simulation and for modelling

Within the results that consider models using all parametric distributions when selecting the best performing model, there were several sets of scenarios where the distribution used for generating the survival data is not the same as the distribution most often selected as the best performing model (regardless of the GPM incorporating method). This occurred for scenarios that use the Gompertz distribution to generate survival data for the overall mean survival estimand (generalised gamma is selected most often) and for scenarios that use the Weibull distribution to generate survival data for the survival probability at time *t* estimand (generalised gamma is selected most often). For RMST, models that used the Weibull distribution were selected most often in scenarios that used the Gompertz, log-logistic or lognormal distributions to generate survival data, and models that used the log-logistic distribution were selected most often in scenarios that used the Weibull distribution to generate survival data.

First, it should be noted that the Weibull distribution is a specialised case of the generalised gamma distribution (Latimer, 2011), which may explain the discrepancy for Weibull distributed scenarios for the survival probability at time *t* estimand (where models using the generalised Gamma distribution performed best in more scenarios than models using the Weibull distribution). For the other distributions, a scenario where the best performing model did not use the same parametric distribution as the one used to generate data for all three estimands was selected to explore further, as described in Section 5.3.5. For this scenario, a large dataset of 1,000,000 patients was generated. Figure 12 shows that the KM-curve matches the shape of the expected DSM survival until around 30 years where it starts to decrease faster than the expected DSM survival, likely due to GPM having a stronger effect after a certain period of time. The Gompertz model actually followed the KM-curve relatively closely, but underestimates survival from around 10 to 30 years. The non-GPM exponential model was the best performing model for bias of overall mean survival, but had a worse visual fit than the Gompertz model. As such, it seems safe to assume that discrepancies in the distribution used to generate survival data and the distribution selected for modelling were due to GPM altering the shape of the expected DSM survival in smaller datasets, and using traditional selection techniques for a parametric distribution (which includes assessing visual fit) would result in selecting a distribution that matches the underlying data.

6.2 Limitations

There are various limitations found throughout execution of this research that should be considered for the conclusions and that could possibly be addressed in future research. First, it is possible that more patterns in performance between methods across scenarios would have become apparent if more analysis had been done on the methods that were only slightly outperformed by the best performing method. For example, more analysis on the pool of best methods (meaning the methods that are among the highest ranking using a pairwise Wilcoxon signed rank test) could have been conducted, as it is possible other GPM methods could have been within that pool, and more patterns could emerge if also considering the second, third, etc. best performing methods per scenario where a significant result was not found. Furthermore, when all parametric distributions are under consideration, there could be scenarios where the pool of best methods all use the same GPM incorporating method that were not found using the current analysis, or the pool of best methods actually contains a majority of models using another GPM incorporating method. An attempt to rectify these issues was made by filtering the models under consideration for using the same parametric distribution as was used for generating the survival data in that scenario, but using a more sophisticated selection process could have improved the overall analysis.

The estimands used could also be improved. RMST is an often-used metric in survival analysis, and was also used in another simulation study on survival extrapolations, where RMST until the end of follow-up was assessed (Rutherford et al., 2020), and thus, RMST was implemented similarly for this study. In doing so, the estimand did not explicitly help to answer the main research question, which was to assess extrapolation performance while RMST until the end of follow-up can be estimated without extrapolation. Furthermore, the RMST would also not be of interest in real-world situations for a lot of scenarios tested in this study. For example, for a trial with the characteristics of the high survival and medium information scenarios, looking at RMST has very little value when the expected survival (based on the underlying survival distribution using the ulcerative colitis data) is still over 90% after 3 years. This could be addressed by changing the estimand to a RMST at twice the follow-up length of the trial, for instance. However, the RMST as it currently stands did prove to be useful as an indicator of short-term fit of the survival models to the data. For survival probability at time t in low survival ($t = 3$) scenarios, the estimand was also not extrapolated in the scenarios with medium or high information (where follow-up is 3 and 10 years, respectively).

Finally, some minor limitations and assumptions used in the coding of this research should be addressed. For example, ages are limited to 100 due to the lifetables assuming infinite hazards at age 100, thus making this a hard assumption to avoid. Due to these infinite hazards, however, the max event time of a patient had to be limited to 99.999, as certain R function used for the internal additive hazards method would fail if infinite hazards were supplied. However, this would only affect modelling if the event time was uncensored, which can only occur if a patient is older than 99 for low information scenarios, older than 97 for medium information scenarios or older than 90 for high information scenarios. When 100,000,000 patients were generated using functions for the old ages and high heterogeneity scenario, only 4% of patients were older than 90, and thus this assumption is assumed to not have affected results significantly.

The generalised F distribution was also considered in analysis, but in a majority of the scenarios (217 out of 288), generalised F models failed to converge in at least one replication due to the standard optimisation function used by the `flexsurv` package failing to find an optimal result. Since changing the optimisation function resulted in other models often failing to converge, the decision was made to keep using the standard optimisation function (the Broyden-Fletcher-Goldfarb-Shanno algorithm).

6.3 Strengths

In this section, strengths of the study will be discussed. Overall, this study seems to be the first simulation study for survival extrapolation where such a large number of scenarios is compared, as although all the dimensions of scenarios that were used in this study have been compared in other simulation studies individually, none have compared all dimensions in a full-factorial manner. Furthermore, although the analysis methods used (such as the colour-coded tables and Wilcoxon test) could be improved further (as discussed in the limitations), they are relatively novel methods that could be employed for similar simulation studies in the future.

6.4 Comparison with literature

Before making recommendations for guidance, the results of this study will be briefly compared to previous literature that is not official HTA guidance. Firstly, as has already been discussed in Section 6.1.1, most articles that discuss GPM incorporating methods recommend using an internal additive hazards approach, yet this study found many scenarios where an external additive hazards approach performed better than an internal additive hazards approach. Furthermore, Jackson et al. (2017) stated that the long-term assumptions used to incorporate GPM information (most of which were included in this study) cannot be tested from data alone, while there certainly are a few patterns in performance resulting from this study that prove otherwise (which will be discussed in Section 6.5).

Otherwise, this study supports the general consensus that incorporating GPM into survival extrapolations improves survival extrapolations, especially considering the interpretation for scenarios where non-GPM extrapolations performed best (see Section 6.1.2). Furthermore, the results of the simulation somewhat support a conclusion drawn by van Oostrum et al. (2021), as they recommend using the external additive hazards method in young populations. In this study, the external additive hazards method generally performed the best most often in scenarios with young ages *and* medium survival for overall mean survival (for both filtered and unfiltered results). In scenarios with young ages and low or high survival, however, the preference for the external additive hazards method was less strong, although it still performed the best the most often.

6.5 Recommendations for guidance for selecting a GPM method

Current guidance for survival extrapolations for HTA submissions states that incorporating GPM information should at least be used for assessing clinical plausibility of the extrapolation (Latimer, 2011), and more recent NICE guidance recommends to always incorporate GPM information, where using clinical expertise to select a GPM incorporating method is recommended (Rutherford et al., 2020). The simulation results show that incorporating GPM information generally improves survival extrapolations, providing support for the guidance stating that GPM information should always be incorporated in parametric survival models. In scenarios where non-GPM extrapolations did outperform GPM incorporating models in terms of absolute mean bias, the results were often not significant using the Wilcoxon test and the converging hazards method was often the second best performing method (see Section 6.1.2).

Furthermore, the results show that the current recommendations for incorporating GPM information into survival extrapolations using the standard parametric models, which mostly recommend using the internal additive hazards approach (referred to as relative survival in the guidance) (Rutherford et al., 2020), could be extended to explicitly include the external additive hazards and converging hazards methods. For example, for mean survival, the internal additive hazards method was outperformed by the external additive hazards method and converging hazards method more often (see Figures 4 and 5). Currently, the converging hazards and external additive hazards are only vaguely mentioned throughout the guidance, and mostly for other types of models (cubic spline models, mixture, landmark, piecewise and cure models). The distinction between the three methods could be more explicit, and concise information as to how to extrapolate or fit a standard parametric model using the assumptions as was done for the relative survival approach is lacking (Rutherford et al., 2020).

Before continuing onto patterns in the performance of the methods, it should be noted that there are technically two decisions to be made for the types of models tested in this research, namely the way in which GPM information is incorporated, but also what parametric distribution to use for the model. Current guidance already has a clear selection process for choosing a parametric distribution based on trial data until the end of its follow-up. Given the differences in results for the RMST and overall mean survival estimand, it is arguable that the GPM incorporating method should be selected first and then a parametric distribution should be selected using the already published guidance afterwards.

For example, current guidance recommends assessing visual fit using AIC or BIC tests to assess statistical fit of the models under consideration to the (short-term) trial data to select a parametric distribution (Latimer, 2011). RMST is also known at the end of a trial, and thus, performance of a model on RMST in this study could be seen as a test on fit to the short-term trial data. Since the results for RMST show that external additive hazards models were selected the least often as the best performing method, while for overall mean survival it was the selected the most, selecting a GPM incorporating method using the current guidance could lead to inaccurate results as the internal additive hazards models might have a higher AIC or BIC.

First, for overall mean survival and survival probability at time *t* estimands, the internal additive hazards method generally did not perform well in scenarios with high survival in both the filtered (see Appendix A.1 and A.2) and unfiltered results (see Appendix B.1 and B.2). This is likely due to the relatively high number of patients that are still alive at the end of follow-up, where GPM can still be low (see Section 6.1.1). Thus, for trials with high survival, using either a converging or external additive hazards adjustment is more appropriate. Although the converging hazards method had a lower bias over the high survival scenarios (0.54, see Table 97) than the external additive hazards method (-2.52, see Table 96) overall, there were high survival scenarios in which the external additive hazards models significantly outperformed converging hazards models, and thus both methods should be considered.

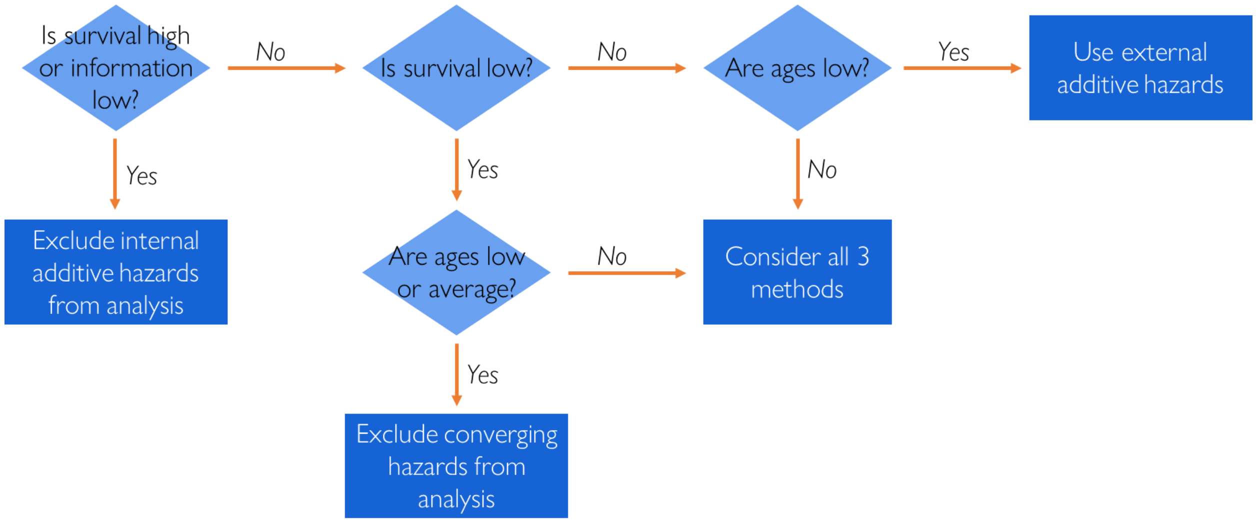
In low survival, the internal additive hazards method did generally perform better for survival probability at time *t*, but this was likely due to the medium and high information scenarios not being an extrapolation as results for survival probability at time *t* look similar to the results for RMST in scenarios with low survival. For mean survival, there were a few low survival scenarios where the internal additive hazards performed the best, and thus the method should be included in analysis for low survival diseases. Furthermore, the converging hazards generally performed better in scenarios with either old ages or high or medium survival, and was rarely the best performing method in scenarios with low survival and young or average ages (and never significantly so) for both the overall mean survival and survival probability at time *t* estimands. Thus, the converging hazards method can be excluded from analysis in trials with low survival and young or average ages.

Next, in low information scenarios, external additive hazards models were often among the best for overall mean survival, often significantly so (see Appendix A.1 and Appendix B.1), which is likely due to a similar reason to what is described in Section 6.1.1, as more patients are likely to still be alive at the end of follow-up if the follow-up length is short. Thus, the internal additive hazards method performed poorly in these scenarios. Since there were still a few low information scenarios where converging hazards models performed the best in both the filtered and unfiltered results, the method cannot be excluded entirely in low information scenarios and both the external additive method and converging hazards method should be considered.

The patterns described in the previous paragraphs are summarised in a flowchart in Figure 13. For definitions of the scenarios, refer to Table 3. More specific patterns were found in the results, but they involve combinations of three dimensions of scenarios. If a trial happens to match three out of the four discernible dimensions (level of survival, ages, heterogeneity and level of information), a researcher could look at the colour coded tables presented in Appendix A and Appendix B. For scenarios not mentioned in Chapter 5 or in this section, using clinical expertise to select between a GPM incorporating method still seems to be the only sufficient way to select a GPM incorporating method.

Figure 13:

Flowchart for selecting a GPM incorporating method for parametric survival extrapolations. *(Refer to Table 3 for definitions of scenarios)*



6.6 Recommendations for future research

In future research more complex DGMs could be tested, as for generating data, only an additive hazards assumption was used in this study, where the lowest out of a generated DSM and GPM time was selected for the ACM event time per patient. Although results sometimes showed preference for using a converging hazards assumption, the robustness of these methods over more complex data generation mechanisms should be assessed. This could be assessed by selecting a base-case scenario for the other dimensions of scenarios used in this study, and only comparing different assumptions about the relationship between DSM and GPM. Using more complex DGMs would also allow the possibility of more complex models to be compared, for example by incorporating a cure fraction and comparing incorporation of GPM information into cure models using different assumptions. However, as the mechanism used in this research has been used in multiple prior simulation studies (Rutherford et al., 2020) and (Jakobsen et al., 2019), comparing more complex survival functions was outside of the scope. Furthermore, the bias that could have resulted from using an additive hazards assumption in data generation was not addressed, which could for example have been alleviated by explicitly using a converging hazards methods during data generation and weighting the results. However, considering that the results still showed scenarios where the converging hazards method performed best, this bias may be assumed to be small.

Furthermore, the low performance of the internal additive hazards method in datasets with a relatively low GPM and high level of censoring that was identified in Section 6.1.1 could be explored further, as the consensus in literature seems to be that the internal additive hazards method usually is most appropriate the best while the results prove otherwise for the scenarios considered in this study. Another possible extension of the simulation could be to vary the factors for scenarios continuously and attempt to fit a statistical model to the results to find patterns between the methods on a continuous scale.

7 Conclusions & recommendations

To conclude, and to answer the main research question of the study *("What is the performance in terms of accuracy of survival extrapolation methods that incorporate GPM information in scenarios with different patient characteristics and availability of information?")*, methods that incorporate GPM information outperformed non-GPM extrapolations in a majority of scenarios across estimands overall. Thus, the study strongly supports the current NICE guidance and showed that GPM information should always be incorporated into survival extrapolations, as even in scenarios where non-GPM extrapolations outperformed extrapolations that did use GPM information, the converging hazards method was generally an adequate substitute.

The results of the study show that the method used to incorporate GPM information should be selected before choosing what parametric model to use, and that all standard parametric distributions mentioned by NICE should be considered. Based on the performance of the internal additive hazards method being rather low, yet it being the only method currently explicitly mentioned in NICE guidance for parametric models, the research would recommend including the converging hazards and external additive hazards methods explicitly in HTA guidance for standard parametric models.

Furthermore, the study shows that there are certain GPM methods that can be excluded from analysis in certain situations. In situations with high survival (median survival of over 40 years, or similar survival to ulcerative colitis), or with low information (trial size of around 100 patients per treatment arm or less, around 30% of patients being right censored and follow-up time of 1 year), the internal additive hazards method can be excluded. In situations with a young patient population (average age of 35) and medium survival (median survival of over 5 years, or survival similar to myocarditis), the external additive hazards method can be used unless there are explicit doubts of its clinical validity. In situations with low survival (median survival of 17 months, or similar survival to pancreatic cancer), and young (average age of 35) or average ages (average age of 50), the converging hazards method can be excluded from analysis. These findings are summarised in Figure 13. For other situations simulated in this study, clinical expertise should be used to select a GPM incorporating method, as is in line with current NICE guidance.

In future research, performance of the GPM incorporating methods used in this study, as well as other models, should be explored using more complex relationships between GPM and DSM, for example by using a converging hazards assumption in the data generating mechanisms rather than an additive hazards assumption. Moreover, the relationship between the performance of the internal additive hazards method and datasets that have a low GPM and high level of censoring should be explored further. Finally, a similar simulation study could be performed with the intention of fitting a statistical model to the results from the onset.

8 Bibliography

Ahlberg, J. H., Nilson, E. N., & Walsh, J. L. (2016). The Theory of Splines and Their Applications: Mathematics in Science and Engineering: A Series of Monographs and Textbooks, Vol. 38 (Vol. 38). Elsevier.

Aivaliotis, G., Palczewski, J., Atkinson, R., Cade, J. E., & Morris, M. A. (2021). A comparison of time to event analysis methods, using weight status and breast cancer as a case study. Sci Rep, 11(1), 14058. https://doi.org/10.1038/s41598-021-92944-z

Anderson, J. R., Cain, K. C., & Gelber, R. D. (1983). Analysis of Survival by Tumor Response. Journal of Clinical Oncology, 1(11), 710-719.

Andersson, T. M., Dickman, P. W., Eloranta, S., Lambe, M., & Lambert, P. C. (2013). Estimating the loss in expectation of life due to cancer using flexible parametric survival models. Stat Med, 32(30), 5286-5300. https://doi.org/10.1002/sim.5943

Arias, E., & Xu, J. (2022). United States life tables, 2020. National Vital Statistics Report, 71(1). https://doi.org/https://dx.doi.org/10.15620/cdc:118055

Bagust, A., & Beale, S. (2014). Survival analysis and extrapolation modeling of time-to-event clinical trial data for economic evaluation: an alternative approach. Med Decis Making, 34(3), 343-351. https://doi.org/10.1177/0272989X13497998

Bell Gorrod, H., Kearns, B., Stevens, J., Thokala, P., Labeit, A., Latimer, N., Tyas, D., & Sowdani, A. (2019). A Review of Survival Analysis Methods Used in NICE Technology Appraisals of Cancer Treatments: Consistency, Limitations, and Areas for Improvement. Med Decis Making, 39(8), 899-909. https://doi.org/10.1177/0272989X19881967

Benaglia, T., Jackson, C. H., & Sharples, L. D. (2015). Survival extrapolation in the presence of cause specific hazards. Stat Med, 34(5), 796-811. https://doi.org/10.1002/sim.6375

Boag, J. W. (1949). Maximum Likelihood Estimates of the Proportion of Patients Cured by Cancer Therapy. Journal of the Royal Statistical Society, 11(1), 15-53.

Bullement, A., Latimer, N. R., & Bell Gorrod, H. (2019). Survival Extrapolation in Cancer Immunotherapy: A Validation-Based Case Study. Value Health, 22(3), 276-283. https://doi.org/10.1016/j.jval.2018.10.007

CADTH. (2017). Guidelines for the economic evaluation of health technologies: Canada.

Chu, P. C., Wang, J. D., Hwang, J. S., & Chang, Y. Y. (2008). Estimation of life expectancy and the expected years of life lost in patients with major cancers: extrapolation of survival curves under high-censored rates. Value Health, 11(7), 1102-1109. https://doi.org/10.1111/j.1524-4733.2008.00350.x

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003). Survival analysis part I: basic concepts and first analyses. Br J Cancer, 89(2), 232-238. https://doi.org/10.1038/sj.bjc.6601118

Collett, D. (2003). Modelling survival data in medical research (2 ed.). CRC Press.

Coyle, D., Haines, A., & Lee, K. (2023). Extrapolating Clinical Evidence Within Economic Evaluations. Canadian Journal of Health Technologies, 3(5).

Demiris, N., Lunn, D., & Sharples, L. D. (2015). Survival extrapolation using the poly-Weibull model. Stat Methods Med Res, 24(2), 287-301. https://doi.org/10.1177/0962280211419645

Dickman, P. W., & Adami, H. O. (2006). Interpreting trends in cancer patient survival. J Intern Med, 260(2), 103-117. https://doi.org/10.1111/j.1365-2796.2006.01677.x

Eddy, D. (1985). Technology assessment: the role of mathematical modelling. Assessing medical technologies, 144-153.

Gallacher, D., Kimani, P., & Stallard, N. (2021a). Extrapolating Parametric Survival Models in Health Technology Assessment Using Model Averaging: A Simulation Study. Med Decis Making, 41(4), 476-484. https://doi.org/10.1177/0272989X21992297

Gallacher, D., Kimani, P., & Stallard, N. (2021b). Extrapolating Parametric Survival Models in Health Technology Assessment: A Simulation Study. Med Decis Making, 41(1), 37-50. https://doi.org/10.1177/0272989X20973201

Gelber, R. D., Goldhirsch, A., & Cole, B. F. (1993). Parametric extrapolation of survival estimates with applications to quality of life evaluation of treatments. Controlled Clinical Trials, 14(6), 485-499. https://doi.org/https://doi.org/10.1016/0197-2456(93)90029-D

Guyot, P., Ades, A. E., Beasley, M., Lueza, B., Pignon, J. P., & Welton, N. J. (2017). Extrapolation of Survival Curves from Cancer Trials Using External Information. Med Decis Making, 37(4), 353-366. https://doi.org/10.1177/0272989X16670604

Guyot, P., Ades, A. E., Ouwens, M. J. N. M., & Welton, N. J. (2012). Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Medical Research Methodology, 12(1), 9. https://doi.org/10.1186/1471-2288-12-9

Guyot, P., Welton, N. J., Ouwens, M. J., & Ades, A. E. (2011). Survival time outcomes in randomized, controlled trials and meta-analyses: the parallel universes of efficacy and cost-effectiveness. Value Health, 14(5), 640-646. https://doi.org/10.1016/j.jval.2011.01.008

Hakama, M., & Hakulinen, T. (1977). Estimating the expectation of life in cancer survival studies with incomplete follow-up information. Journal of Chronic Diseases, 30(9), 585-597.

Hougaard, P. (1995). Frailty Models for Survival Data. Lifetime Data Analysis, 1, 255-273.

Hwang, J.-S., & Wang, J.-D. (1999). Monte Carlo estimation of extrapolation of quality-adjusted survival for follow-up studies. Statistics in medicine, 18(13), 1627-1640. https://doi.org/10.1002/(sici)1097-0258(19990715)18:13<1627::Aid-sim159>3.0.Co;2-d

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. The Annals of Applied Statistics, 2(3). https://doi.org/10.1214/08-aoas169

Jackson, C., Stevens, J., Ren, S., Latimer, N., Bojke, L., Manca, A., & Sharples, L. (2017). Extrapolating Survival from Randomized Trials Using External Data: A Review of Methods. Med Decis Making, 37(4), 377-390. https://doi.org/10.1177/0272989X16639900

Jackson, C. H., Sharples, L. D., & Thompson, S. G. (2010). Survival models in health economic evaluations: balancing fit and parsimony to improve prediction. Int J Biostat, 6(1), Article 34. https://doi.org/10.2202/1557-4679.1269

Jakobsen, L. H., Andersson, T. M., Biccler, J. L., El-Galaly, T. C., & Bogsted, M. (2019). Estimating the loss of lifetime function using flexible parametric relative survival models. BMC Med Res Methodol, 19(1), 23. https://doi.org/10.1186/s12874-019-0661-8

Jess, T., Loftus, E. V., Jr., Harmsen, W. S., Zinsmeister, A. R., Tremaine, W. J., Melton, L. J., 3rd, Munkholm, P., & Sandborn, W. J. (2006). Survival and cause specific mortality in patients with inflammatory bowel disease: a long term outcome study in Olmsted County, Minnesota, 1940-2004. Gut, 55(9), 1248-1254. https://doi.org/10.1136/gut.2005.079350

Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association, 53(282), 457-481. https://doi.org/10.1080/01621459.1958.10501452

Kearns, B., Stevenson, M. D., Triantafyllopoulos, K., & Manca, A. (2021). Comparing current and emerging practice models for the extrapolation of survival data: a simulation study and case-study. BMC Med Res Methodol, 21(1), 263. https://doi.org/10.1186/s12874-021-01460-1

Kleinbaum, D. G., & Klein, M. (2005). Survival Analysis: A Self-Learning Text. Springer. https://books.google.nl/books?id=GNhzxRkFnJ0C

Kuhlmann, K. F. D., de Castro, S. M. M., Wesseling, J. G., ten Kate, F. J. W., Offerhaus, G. J. A., Busch, O. R. C., van Gulik, T. M., Obertop, H., & Gouma, D. J. (2004). Surgical treatment of pancreatic adenocarcinoma. European Journal of Cancer, 40(4), 549-558. https://doi.org/10.1016/j.ejca.2003.10.026

Latimer, N. (2011). NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. http://www.nicedsu.org.uk

Latimer, N. R. (2013). Survival analysis for economic evaluations alongside clinical trials--extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. Med Decis Making, 33(6), 743-754. https://doi.org/10.1177/0272989X12472398

Latimer, N. R. (2014). Response to "survival analysis and extrapolation modeling of time-to-event clinical trial data for economic evaluation: an alternative approach" by Bagust and Beale. Med Decis Making, 34(3), 279-282. https://doi.org/10.1177/0272989X13511302

Latimer, N. R., & Adler, A. I. (2022). Extrapolation beyond the end of trials to estimate long term survival and cost effectiveness. BMJ Medicine, 1(1). https://doi.org/10.1136/bmjmed-2021-000094

Magnani, J. W., Danik, H. J., Dec, G. W., Jr., & DiSalvo, T. G. (2006). Survival in biopsy-proven myocarditis: a long-term retrospective analysis of the histopathologic, clinical, and hemodynamic predictors. Am Heart J, 151(2), 463-470. https://doi.org/10.1016/j.ahj.2005.03.037

McLachlan, G. J., Lee, S. X., & Rathnayake, S. I. (2019). Finite Mixture Models. Annual Review of Statistics and Its Application, 6(1), 355-378. https://doi.org/10.1146/annurev-statistics-031017-100325

Miners, A. H., Garau, M., Fidan, D., & Fischer, A. J. (2005). Comparing estimates of cost effectiveness submitted to the National Institute for Clinical Excellence (NICE) by different organisations: retrospective study. BMJ, 330(7482), 65. https://doi.org/10.1136/bmj.38285.482350.82

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. Stat Med, 38(11), 2074-2102. https://doi.org/10.1002/sim.8086

Nelson, C. P., Lambert, P. C., Squire, I. B., & Jones, D. R. (2007). Flexible parametric models for relative survival, with application in coronary heart disease. Stat Med, 26(30), 5486-5498. https://doi.org/10.1002/sim.3064

NICE. (2013). Guide to the methods of technology appraisal 2013.

NICE. (2022). NICE health technology evaluations: the manual. https://www.nice.org.uk/process/pmg36/

Palmer, S., Borget, I., Friede, T., Husereau, D., Karnon, J., Kearns, B., Medin, E., Peterse, E. F. P., Klijn, S. L., Verburg-Baltussen, E. J. M., Fenwick, E., & Borrill, J. (2023). A Guide to Selecting Flexible Survival Models to Inform Economic Evaluations of Cancer Immunotherapies. Value Health, 26(2), 185-192. https://doi.org/10.1016/j.jval.2022.07.009

Pennington, M., Grieve, R., der Meulen, J. V., & Hawkins, N. (2018). Value of External Data in the Extrapolation of Survival Data: A Study Using the NJR Data Set. Value Health, 21(7), 822-829. https://doi.org/10.1016/j.jval.2017.12.023

Perme, M. P., Stare, J., & Esteve, J. (2012). On estimation in relative survival. Biometrics, 68(1), 113-120. https://doi.org/10.1111/j.1541-0420.2011.01640.x

Royston, P., & Parmar, M. K. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Stat Med, 21(15), 2175-2197. https://doi.org/10.1002/sim.1203

Rutherford, M., Lambert, P., Sweeting, M., Pennington, R., Crowther, M., Abrams, K., & Latimer, N. (2020). NICE DSU Technical Support Document 21. Flexible Methods for Survival Analysis. http://www.nicedsu.org.uk

Sasieni, P., & Brentnall, A. R. (2017). On standardized relative survival. Biometrics, 73(2), 473-482. https://doi.org/10.1111/biom.12578

Tappenden, P., Chilcott, J., Ward, S., Eggington, S., Hind, D., & Hummel, S. (2006). Methodological issues in the economic analysis of cancer treatments. Eur J Cancer, 42(17), 2867-2875. https://doi.org/10.1016/j.ejca.2006.08.010

van Oostrum, I., Ouwens, M., Remiro-Azocar, A., Baio, G., Postma, M. J., Buskens, E., & Heeg, B. (2021). Comparison of Parametric Survival Extrapolation Approaches Incorporating General Population Mortality for Adequate Health Technology Assessment of New Oncology Drugs. Value Health, 24(9), 1294-1301. https://doi.org/10.1016/j.jval.2021.03.008

Verheul, H., Dekker, E., Bossuyt, P., Moulijn, A., & Dunning, A. (1993). Background mortality in clinical survival studies. The Lancet, 341, 872-875.

Vickers, A. (2019). An Evaluation of Survival Curve Extrapolation Techniques Using Long-Term Observational Cancer Data. Med Decis Making, 39(8), 926-938. https://doi.org/10.1177/0272989X19875950

Wikipedia. (n.d., 03/03/2023). Spline (mathematics). Wikimedia Foundation. Retrieved 05/04/2023 from https://en.wikipedia.org/wiki/Spline_(mathematics)

Williams, C., Lewsey, J. D., Mackay, D. F., & Briggs, A. H. (2017). Estimation of Survival Probabilities for Use in Cost-effectiveness Analyses: A Comparison of a Multi-state Modeling Survival Analysis Approach with Partitioned Survival and Markov Decision-Analytic Modeling. Med Decis Making, 37(4), 427-439. https://doi.org/10.1177/0272989X16670617

Appendix A: Filtered colour-coded tables for best ranking methods per scenario

This Appendix shows the results of the simulation by colour-coding the tested methods and showing the best performing method for each scenario based on the absolute mean bias for the estimand. Separate tables are shown for each estimand and underlying survival distribution used to generate data separately. For each distribution used to generate survival data, only models that used the same distribution were considered. The DGM dimensions (survival, age, heterogeneity, level of information) are used as rows and columns for the tables, tables are shown in pairs with flipped "minor" "and major" dimension of rows and columns. Cells show the mean bias and mean RMSE for the method over 2,500 replications. A triangle denotes whether the result was significant based on a Wilcoxon signed-rank test (α = 0.05). Each pair of tables is accompanied by another that mostly serves as a legend, but also shows the overall number of times a method ranked best and the percentage of that result which was significant for each survival distribution used to generate the data.

A.1: Overall mean survival

Weibull distributed data

Table 22:

Legend for filtered colour-coded table and times a GPM incorporating method ranked best for absolute mean bias of mean survival for Weibull distributed data. *Amount of times the method ranked best*

	Non-GPM	Internal additive	External additive	Converging
Times method ranked best	*8*	*23*	*14*	*23*

Table 23:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of mean survival per scenario for Weibull distributed data. Survival and ages as main axes. *(Bias, RMSE),* ▲ = *Significant result (Wilcoxon)*

		Survival (Weibull)							
		High		Medium			Low		
		Information		*Information*			*Information*		
Age	*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium	High
Young	Low	*NA*	*-1.413, 1.811* ▲	*0.556, 3.7* ▲	*-0.011, 1.081* ▲	*-0.153, 0.339* ▲	*0.035, 0.376*	*-0.041, 0.057* ▲	*-0.046, 0.046* ▲
Young	Medium	*NA*	*-1.762, 2.09* ▲	*-0.325, 3.627* ▲	*0.05, 1.3*	*0.117, 0.407* ▲	*-0.305, 0.485*	*-0.19, 0.192* ▲	*-0.058, 0.058* ▲
Young	High	*0.091, 5.364* ▲	*1.023, 1.356* ▲	*-3.353, 4.742*	*-1.288, 1.804* ▲	*-0.019, 0.512* ▲	*-3.31, 3.347*	*-2.14, 2.142*	*-0.626, 0.64* ▲
Average	Low	*NA*	*1.964, 1.996* ▲	*0.054, 2.756* ▲	*-0.188, 0.817* ▲	*-0.377, 0.398* ▲	*0.034, 0.358*	*-0.034, 0.052* ▲	*-0.036, 0.036* ▲
Average	Medium	*NA*	*1.644, 1.706* ▲	*-0.447, 2.72* ▲	*0.309, 1* ▲	*0.329, 0.395* ▲	*-0.29, 0.472*	*-0.169, 0.171* ▲	*-0.039, 0.039* ▲
Average	High	*-0.608, 2.378* ▲	*0.806, 0.932* ▲	*-1.391, 3.261*	*-0.076, 1.093*	*0.428, 0.505* ▲	*-2.371, 2.451*	*-1.286, 1.297* ▲	*0.102, 0.249* ▲
Old	Low	*0.426, 0.683* ▲	*0.057, 0.121* ▲	*-0.651, 1.105* ▲	*0.572, 0.63* ▲	*0.15, 0.153* ▲	*-0.057, 0.301*	*-0.042, 0.052*	*-0.053, 0.053*
Old	Medium	*0.027, 0.554* ▲	*-0.275, 0.287* ▲	*0.418, 1.402* ▲	*0.365, 0.473* ▲	*0.061, 0.093* ▲	*-0.105, 0.384*	*-0.012, 0.065* ▲	*-0.055, 0.055*
Old	High	*0.403, 1.053* ▲	*0.176, 0.251* ▲	*0.147, 1.417* ▲	*0.012, 0.315* ▲	*-0.097, 0.133* ▲	*-0.539, 0.823* ▲	*0.193, 0.316* ▲	*0.016, 0.076* ▲

Table 24:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of mean survival per scenario for Weibull distributed data. Information and heterogeneity as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

			Information							
			Low		Medium			High		
			Survival (Weibull)		*Survival (Weibull)*			*Survival (Weibull)*		
			Medium	Low	High	Medium	Low	High	Medium	Low
Heterogeneity Low	Age	Young	*0.556, 3.7* ▲	*0.035, 0.376*	*NA*	*-0.011, 1.081* ▲	*-0.041, 0.057* ▲	*-1.413, 1.811* ▲	*-0.153, 0.339* ▲	*-0.046, 0.046* ▲
		Average	*0.054, 2.756* ▲	*0.034, 0.358*	*NA*	*-0.188, 0.817* ▲	*-0.034, 0.052* ▲	*1.964, 1.996* ▲	*-0.377, 0.398* ▲	*-0.036, 0.036* ▲
		Old	*-0.651, 1.105* ▲	*-0.057, 0.301*	*0.426, 0.683* ▲	*0.572, 0.63* ▲	*-0.042, 0.052*	*0.057, 0.121* ▲	*0.15, 0.153* ▲	*-0.053, 0.053*
Medium	Age	Young	*-0.325, 3.627* ▲	*-0.305, 0.485*	*NA*	*0.05, 1.3*	*-0.19, 0.192* ▲	*-1.762, 2.09* ▲	*0.117, 0.407* ▲	*-0.058, 0.058* ▲
		Average	*-0.447, 2.72* ▲	*-0.29, 0.472*	*NA*	*0.309, 1* ▲	*-0.169, 0.171* ▲	*1.644, 1.706* ▲	*0.329, 0.395* ▲	*-0.039, 0.039* ▲
		Old	*0.418, 1.402* ▲	*-0.105, 0.384*	*0.027, 0.554* ▲	*0.365, 0.473* ▲	*-0.012, 0.065* ▲	*-0.275, 0.287* ▲	*0.061, 0.093* ▲	*-0.055, 0.055*
High	Age	Young	*-3.353, 4.742*	*-3.31, 3.347*	*0.091, 5.364* ▲	*-1.288, 1.804* ▲	*-2.14, 2.142*	*1.023, 1.356* ▲	*-0.019, 0.512* ▲	*-0.626, 0.64* ▲
		Average	*-1.391, 3.261*	*-2.371, 2.451*	*-0.608, 2.378* ▲	*-0.076, 1.093*	*-1.286, 1.297* ▲	*0.806, 0.932* ▲	*0.428, 0.505* ▲	*0.102, 0.249* ▲
		Old	*0.147, 1.417* ▲	*-0.539, 0.823* ▲	*0.403, 1.053* ▲	*0.012, 0.315* ▲	*0.193, 0.316* ▲	*0.176, 0.251* ▲	*-0.097, 0.133* ▲	*0.016, 0.076* ▲

Log-logistic distributed data

Table 25:

Legend for filtered colour-coded table and times a GPM incorporating method for absolute mean bias of mean survival for log-logistic distributed data. *Amount of times the method ranked best*

	Non-GPM	Internal additive	External additive	Converging
Times method ranked best	*1*	*2*	*39*	*30*

Table 26:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of mean survival per scenario for log-logistic distributed data. Survival and ages as main axes. *(Bias, RMSE),* ▲ = *Significant result (Wilcoxon)*

Survival (Log-logistic)

Age	*Heterogeneity*	High – *Information* Medium	High – *Information* High	Medium – *Information* Low	Medium – *Information* Medium	Medium – *Information* High	Low – *Information* Low	Low – *Information* Medium	Low – *Information* High
Young	Low	*1.042, 6.347* ▲	*-0.949, 1.361* ▲	*0.381, 3.233* ▲	*0.052, 0.903* ▲	*-0.09, 0.341* ▲	*0.059, 0.435*	*-0.01, 0.113*	*-0.016, 0.063*
Young	Medium	*0.851, 6.063* ▲	*-1.205, 1.558* ▲	*-0.084, 3.122* ▲	*-0.07, 0.888* ▲	*-0.063, 0.365* ▲	*0.025, 0.532*	*0.036, 0.144* ▲	*0.007, 0.078* ▲
Young	High	*0.618, 2.848* ▲	*-1.436, 1.65* ▲	*-0.015, 2.67*	*-0.369, 1.02* ▲	*-0.104, 0.522* ▲	*0.216, 1.42*	*0.078, 0.541* ▲	*0.092, 0.308* ▲
Average	Low	*1.808, 3.54* ▲	*-2.106, 2.106* ▲	*0.121, 2.388* ▲	*-0.067, 0.625* ▲	*-0.271, 0.319* ▲	*0.024, 0.396*	*-0.002, 0.101*	*-0.006, 0.056*
Average	Medium	*1.569, 3.397* ▲	*1.974, 1.976* ▲	*-0.148, 2.209* ▲	*-0.155, 0.627* ▲	*-0.305, 0.354* ▲	*-0.038, 0.449*	*0.013, 0.127* ▲	*-0.025, 0.066* ▲
Average	High	*1.08, 1.865* ▲	*1.412, 1.423* ▲	*0.522, 2.081*	*-0.384, 0.752* ▲	*-0.454, 0.535* ▲	*-0.313, 1.027* ▲	*-0.032, 0.423* ▲	*-0.112, 0.242* ▲
Old	Low	*0.648, 0.698* ▲	*0.06, 0.12* ▲	*-0.353, 0.848* ▲	*-0.543, 0.546* ▲	*0.35, 0.35* ▲	*0.021, 0.289* ▲	*0.081, 0.097* ▲	*0.049, 0.056* ▲
Old	Medium	*0.318, 0.504* ▲	*-0.274, 0.29* ▲	*-0.508, 0.853* ▲	*-0.636, 0.639* ▲	*0.248, 0.249* ▲	*-0.066, 0.293* ▲	*-0.115, 0.123* ▲	*0.086, 0.089* ▲
Old	High	*-0.161, 0.454* ▲	*-0.637, 0.638* ▲	*0.523, 0.874* ▲	*0.48, 0.505* ▲	*-0.015, 0.137* ▲	*0.39, 0.601*	*0.392, 0.401* ▲	*0.132, 0.148* ▲

Table 27:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of mean survival per scenario for log-logistic distributed data. Information and heterogeneity as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

		Information							
		Low		Medium			High		
		Survival (Log-logistic)		Survival (Log-logistic)			Survival (Log-logistic)		
Heterogeneity	Age	Medium	Low	High	Medium	Low	High	Medium	Low
Low	Young	▲ 0.381, 3.233	0.059, 0.435	▲ 1.042, 6.347	▲ 0.052, 0.903	-0.01, 0.113	▲ -0.949, 1.361	▲ -0.09, 0.341	-0.016, 0.063
Low	Average	▲ 0.121, 2.388	0.024, 0.396	▲ 1.808, 3.54	▲ -0.067, 0.625	-0.002, 0.101	▲ -2.106, 2.106	▲ -0.271, 0.319	-0.006, 0.056
Low	Old	▲ -0.353, 0.848	▲ 0.021, 0.289	▲ 0.648, 0.698	▲ -0.543, 0.546	▲ 0.081, 0.097	▲ 0.06, 0.12	▲ 0.35, 0.35	▲ 0.049, 0.056
Medium	Young	▲ -0.084, 3.122	0.025, 0.532	▲ 0.851, 6.063	▲ -0.07, 0.888	▲ 0.036, 0.144	▲ -1.205, 1.558	▲ -0.063, 0.365	▲ 0.007, 0.078
Medium	Average	▲ -0.148, 2.209	-0.038, 0.449	▲ 1.569, 3.397	▲ -0.155, 0.627	▲ 0.013, 0.127	▲ 1.974, 1.976	▲ -0.305, 0.354	▲ -0.025, 0.066
Medium	Old	▲ -0.508, 0.853	▲ -0.066, 0.293	▲ 0.318, 0.504	▲ -0.636, 0.639	▲ -0.115, 0.123	▲ -0.274, 0.29	▲ 0.248, 0.249	▲ 0.086, 0.089
High	Young	-0.015, 2.67	0.216, 1.42	▲ 0.618, 2.848	▲ -0.369, 1.02	▲ 0.078, 0.541	▲ -1.436, 1.65	▲ -0.104, 0.522	▲ 0.092, 0.308
High	Average	0.522, 2.081	▲ -0.313, 1.027	▲ 1.08, 1.865	▲ -0.384, 0.752	▲ -0.032, 0.423	▲ 1.412, 1.423	▲ -0.454, 0.535	▲ -0.112, 0.242
High	Old	▲ 0.523, 0.874	0.39, 0.601	▲ -0.161, 0.454	▲ 0.48, 0.505	▲ 0.392, 0.401	▲ -0.637, 0.638	▲ -0.015, 0.137	▲ 0.132, 0.148

Lognormal distributed data

Table 28:

Legend for filtered colour-coded table and times a GPM incorporating method ranked best for absolute mean bias of mean survival for lognormal distributed data. *Amount of times the method ranked best*

	Non-GPM	Internal additive	External additive	Converging
Times method ranked best	*1*	*7*	*48*	*15*

Table 29:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of mean survival per scenario for lognormal distributed data. Survival and ages as main axes. *(Bias, RMSE)*, ▲ = *Significant result (Wilcoxon)*

		Survival (Lognormal)							
		High		Medium			Low		
		Information		*Information*			*Information*		
Age	*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium	High
Young	Low	▲ *-0.924, 3.828*	▲ *-0.462, 0.967*	▲ *0.107, 2.786*	▲ *0.022, 0.86*	▲ *-0.121, 0.346*	*0.037, 0.372*	*-0.032, 0.076*	▲ *-0.037, 0.039*
Young	Medium	*NA*	▲ *-0.546, 1.098*	▲ *0.158, 2.766*	▲ *0.043, 0.879*	▲ *-0.113, 0.348*	*0.045, 0.496*	*-0.021, 0.118*	▲ *-0.026, 0.048*
Young	High	▲ *0.075, 2.113*	▲ *-0.413, 1.221*	▲ *0.195, 2.584*	▲ *0.298, 1.021*	▲ *0.051, 0.503*	▲ *0.188, 1.527*	▲ *0.09, 0.601*	▲ *-0.029, 0.288*
Average	Low	▲ *-0.612, 1.949*	▲ *-1.633, 1.635*	▲ *0.082, 2.1*	▲ *-0.087, 0.603*	▲ *-0.321, 0.352*	*0.071, 0.385*	*-0.017, 0.068*	▲ *-0.023, 0.03*
Average	Medium	▲ *-0.777, 1.924*	▲ *-1.792, 1.795*	▲ *0.058, 2.04*	▲ *-0.124, 0.629*	▲ *-0.347, 0.382*	*0.056, 0.466*	*0, 0.108*	*0.007, 0.04*
Average	High	▲ *-0.631, 1.5*	▲ *-1.658, 1.684*	▲ *0.072, 1.749*	▲ *0.065, 0.697*	▲ *-0.369, 0.492*	▲ *0.051, 1.124*	▲ *-0.016, 0.429*	▲ *-0.183, 0.264*
Old	Low	▲ *0.838, 0.843*	▲ *-0.098, 0.138*	▲ *-0.344, 0.766*	▲ *-0.527, 0.531*	▲ *0.249, 0.249*	▲ *0.115, 0.352*	*0.052, 0.083*	*0.001, 0.028*
Old	Medium	▲ *0.531, 0.582*	▲ *-0.426, 0.429*	▲ *-0.37, 0.771*	▲ *-0.597, 0.6*	▲ *0.159, 0.164*	▲ *0.056, 0.345*	*0.096, 0.121*	▲ *0.022, 0.038*
Old	High	▲ *0.081, 0.443*	▲ *-0.8, 0.8*	▲ *-0.486, 0.758*	▲ *0.596, 0.607*	▲ *-0.076, 0.146*	▲ *-0.226, 0.508*	▲ *-0.406, 0.412*	▲ *0.078, 0.11*

Table 30:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of mean survival per scenario for lognormal distributed data. Information and heterogeneity as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

			Information							
			Low		Medium			High		
			Survival (Lognormal)		*Survival (Lognormal)*			*Survival (Lognormal)*		
Heterogeneity			Medium	Low	High	Medium	Low	High	Medium	Low
Low	Age	Young	*0.107, 2.786* ▲	*0.037, 0.372*	*-0.924, 3.828* ▲	*0.022, 0.86* ▲	*-0.032, 0.076*	*-0.462, 0.967* ▲	*-0.121, 0.346* ▲	*-0.037, 0.039* ▲
		Average	*0.082, 2.1* ▲	*0.071, 0.385*	*-0.612, 1.949* ▲	*-0.087, 0.603* ▲	*-0.017, 0.068*	*-1.633, 1.635* ▲	*-0.321, 0.352* ▲	*-0.023, 0.03* ▲
		Old	*-0.344, 0.766* ▲	*0.115, 0.352* ▲	*0.838, 0.843* ▲	*-0.527, 0.531* ▲	*0.052, 0.083*	*-0.098, 0.138* ▲	*0.249, 0.249* ▲	*0.001, 0.028*
Medium	Age	Young	*0.158, 2.766* ▲	*0.045, 0.496*	*NA*	*0.043, 0.879* ▲	*-0.021, 0.118*	*-0.546, 1.098* ▲	*-0.113, 0.348* ▲	*-0.026, 0.048* ▲
		Average	*0.058, 2.04* ▲	*0.056, 0.466*	*-0.777, 1.924* ▲	*-0.124, 0.629* ▲	*0, 0.108*	*-1.792, 1.795* ▲	*-0.347, 0.382* ▲	*0.007, 0.04*
		Old	*-0.37, 0.771* ▲	*0.056, 0.345* ▲	*0.531, 0.582* ▲	*-0.597, 0.6*	*0.096, 0.121*	*-0.426, 0.429* ▲	*0.159, 0.164* ▲	*0.022, 0.038* ▲
High	Age	Young	*0.195, 2.584* ▲	*0.188, 1.527* ▲	*0.075, 2.113* ▲	*0.298, 1.021* ▲	*0.09, 0.601* ▲	*-0.413, 1.221* ▲	*0.051, 0.503* ▲	*-0.029, 0.288* ▲
		Average	*0.072, 1.749* ▲	*0.051, 1.124* ▲	*-0.631, 1.5* ▲	*0.065, 0.697* ▲	*-0.016, 0.429* ▲	*-1.658, 1.684* ▲	*-0.369, 0.492* ▲	*-0.183, 0.264* ▲
		Old	*-0.486, 0.758* ▲	*-0.226, 0.508* ▲	*0.081, 0.443* ▲	*0.596, 0.607* ▲	*-0.406, 0.412* ▲	*-0.8, 0.8* ▲	*-0.076, 0.146* ▲	*0.078, 0.11* ▲

Gompertz distributed data

Table 31:

Legend for filtered colour-coded table and times a GPM incorporating method ranked best for absolute mean bias of mean survival for Gompertz distributed data. *Amount of times the method ranked best*

	Non-GPM	Internal additive	External additive	Converging
Times method ranked best	*9*	*6*	*41*	*16*

Table 32:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of mean survival per scenario for Gompertz distributed data. Survival and ages as main axes. *(Bias, RMSE),* ▲ *= Significant result (Wilcoxon)*

		Survival (Gompertz)							
		High		Medium			Low		
		Information		*Information*			*Information*		
Age	*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium	High
Young	Low	*-4.44, 14.143* ▲	*0.498, 5.327* ▲	*7.272, 13.268*	*1.573, 4.517* ▲	*-0.184, 0.724* ▲	*4.48, 4.964*	*-0.007, 0.101*	*-0.042, 0.042* ▲
Young	Medium	*-3.34, 14.042* ▲	*-0.49, 5.246* ▲	*7.21, 13.496*	*2.748, 5.062* ▲	*0.65, 0.952* ▲	*4.462, 5.177*	*0.151, 0.368*	*-0.017, 0.046* ▲
Young	High	*2.048, 8.417* ▲	*1.697, 2.191* ▲	*14.071, 14.834*	*9.887, 9.887* ▲	*3.248, 3.248* ▲	*14.477, 14.5*	*8.179, 8.179* ▲	*2.154, 2.154* ▲
Average	Low	*3.328, 15.746* ▲	*-0.382, 2.601* ▲	*4.077, 9.096*	*0.46, 2.713* ▲	*0.299, 0.548* ▲	*2.741, 3.225*	*0.001, 0.096*	*-0.032, 0.032* ▲
Average	Medium	*3.137, 15.385* ▲	*-0.55, 2.561* ▲	*3.462, 8.701*	*1.052, 2.958* ▲	*-0.181, 0.442* ▲	*2.909, 3.613*	*0.05, 0.286*	*-0.001, 0.034* ▲
Average	High	*-0.124, 4.596* ▲	*-1.281, 1.498* ▲	*8.472, 9.123* ▲	*5.44, 5.442* ▲	*0.876, 0.905* ▲	*9.103, 9.124* ▲	*4.674, 4.674* ▲	*0.684, 0.687* ▲
Old	Low	*-0.944, 1.644* ▲	*0.006, 0.325* ▲	*-0.179, 2.658* ▲	*0.219, 0.906* ▲	*0.026, 0.089* ▲	*0.595, 1.1*	*-0.026, 0.081*	*-0.049, 0.049*
Old	Medium	*1.006, 3.708* ▲	*-0.156, 0.373* ▲	*-0.194, 2.58* ▲	*0.254, 0.873* ▲	*0.001, 0.089* ▲	*0.494, 1.13* ▲	*0.005, 0.163*	*-0.044, 0.045*
Old	High	*-0.563, 1.031* ▲	*0.179, 0.342* ▲	*1.133, 1.765* ▲	*-0.113, 0.401* ▲	*-0.282, 0.284* ▲	*1.94, 1.986* ▲	*0.355, 0.389* ▲	*-0.224, 0.225* ▲

Table 33:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of mean survival per scenario for Gompertz distributed data. Information and heterogeneity as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

		Information							
		Low		Medium			High		
		Survival (Gompertz)		*Survival (Gompertz)*			*Survival (Gompertz)*		
Heterogeneity	**Age**	Medium	Low	High	Medium	Low	High	Medium	Low
Low	Young	*7.272, 13.268*	*4.48, 4.964*	*-4.44, 14.143* ▲	*1.573, 4.517* ▲	*-0.007, 0.101*	*0.498, 5.327* ▲	*-0.184, 0.724* ▲	*-0.042, 0.042* ▲
Low	Average	*4.077, 9.096*	*2.741, 3.225*	*3.328, 15.746* ▲	*0.46, 2.713* ▲	*0.001, 0.096*	*-0.382, 2.601* ▲	*0.299, 0.548* ▲	*-0.032, 0.032* ▲
Low	Old	*-0.179, 2.658* ▲	*0.595, 1.1*	*-0.944, 1.644* ▲	*0.219, 0.906* ▲	*-0.026, 0.081*	*0.006, 0.325* ▲	*0.026, 0.089* ▲	*-0.049, 0.049*
Medium	Young	*7.21, 13.496*	*4.462, 5.177*	*-3.34, 14.042* ▲	*2.748, 5.062* ▲	*0.151, 0.368*	*-0.49, 5.246* ▲	*0.65, 0.952* ▲	*-0.017, 0.046* ▲
Medium	Average	*3.462, 8.701*	*2.909, 3.613*	*3.137, 15.385* ▲	*1.052, 2.958* ▲	*0.05, 0.286*	*-0.55, 2.561* ▲	*-0.181, 0.442* ▲	*-0.001, 0.034* ▲
Medium	Old	*-0.194, 2.58* ▲	*0.494, 1.13* ▲	*1.006, 3.708* ▲	*0.254, 0.873* ▲	*0.005, 0.163*	*-0.156, 0.373* ▲	*0.001, 0.089* ▲	*-0.044, 0.045*
High	Young	*14.071, 14.834*	*14.477, 14.5*	*2.048, 8.417* ▲	*9.887, 9.887* ▲	*8.179, 8.179* ▲	*1.697, 2.191* ▲	*3.248, 3.248* ▲	*2.154, 2.154* ▲
High	Average	*8.472, 9.123* ▲	*9.103, 9.124* ▲	*-0.124, 4.596* ▲	*5.44, 5.442* ▲	*4.674, 4.674* ▲	*-1.281, 1.498* ▲	*0.876, 0.905* ▲	*0.684, 0.687* ▲
High	Old	*1.133, 1.765* ▲	*1.94, 1.986* ▲	*-0.563, 1.031* ▲	*-0.113, 0.401* ▲	*0.355, 0.389* ▲	*0.179, 0.342* ▲	*-0.282, 0.284* ▲	*-0.224, 0.225* ▲

A.2: Survival probability at time *t*

Weibull distributed data

Table 34:

Legend for filtered colour-coded table and times a GPM incorporating method ranked best for absolute mean bias of survival probability at time t for Weibull distributed data. *Amount of times the method ranked best*

	Non-GPM	Internal additive	External additive	Converging
Times method ranked best	*8*	*25*	*12*	*23*

Table 35:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of survival probability at time t per scenario for Weibull distributed data. Survival and ages as main axes. *(Bias, RMSE),* ▲ *= Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

		Survival (Weibull)							
		High		Medium			Low		
		Information		*Information*			*Information*		
Age	*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium*	High*
Young	Low	*NA*	*0.013, 0.028*	*0.001, 0.115*	*-0.004, 0.037*	*0.002, 0.012*	*-0.006, 0.083*	▲ *-0.006, 0.009*	*-0.006, 0.01*
Young	Medium	*NA*	*0.012, 0.029*	*-0.01, 0.116*	*-0.006, 0.037*	*-0.003, 0.012*	*-0.051, 0.086*	▲ *-0.012, 0.013*	▲ *0.009, 0.012*
Young	High	▲ *-0.061, 0.095*	*-0.006, 0.02*	*-0.108, 0.125*	▲ *-0.053, 0.056*	▲ *-0.003, 0.011*	*-0.075, 0.082*	▲ *-0.006, 0.008*	▲ *0.033, 0.033*
Average	Low	*NA*	*0.041, 0.047*	▲ *-0.002, 0.103*	▲ *-0.008, 0.035*	*0.008, 0.013*	*-0.001, 0.08*	▲ *-0.004, 0.008*	▲ *-0.004, 0.009*
Average	Medium	*NA*	*0.038, 0.045*	*0.002, 0.111*	*0.001, 0.036*	*0.002, 0.012*	*-0.051, 0.085*	▲ *-0.01, 0.011*	▲ *0.006, 0.011*
Average	High	*-0.02, 0.086*	*0.027, 0.032*	*-0.076, 0.105*	▲ *-0.025, 0.037*	*-0.002, 0.011*	*-0.074, 0.081*	▲ *-0.003, 0.007*	▲ *0.026, 0.026*
Old	Low	▲ *0.013, 0.019*	▲ *0.008, 0.011*	▲ *-0.011, 0.035*	▲ *-0.02, 0.021*	▲ *0.026, 0.026*	*0.004, 0.075*	*-0.006, 0.009*	*-0.007, 0.01*
Old	Medium	▲ *-0.014, 0.021*	▲ *-0.018, 0.02*	▲ *-0.021, 0.037*	▲ *-0.029, 0.029*	▲ *0.018, 0.018*	*-0.027, 0.078*	▲ *0.011, 0.012*	*0.005, 0.01*
Old	High	▲ *-0.039, 0.04*	*-0.039, 0.039*	▲ *-0.002, 0.044*	▲ *0.011, 0.02*	▲ *0.007, 0.011*	▲ *-0.037, 0.058*	*-0.008, 0.01*	*0.009, 0.012*

Table 36:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of survival probability at time t per scenario for Weibull distributed data. Information and heterogeneity as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

Heterogeneity	Age	Information: Low Survival (Weibull) Medium	Low Survival (Weibull) Low	Medium Survival (Weibull) High	Medium Survival (Weibull) Medium	Medium Survival (Weibull) Low*	High Survival (Weibull) High	High Survival (Weibull) Medium	High Survival (Weibull) Low*
Low	Young	*0.001, 0.115*	*-0.006, 0.083*	*NA*	*-0.004, 0.037*	*-0.006, 0.009* ▲	*0.013, 0.028*	*0.002, 0.012*	*-0.006, 0.01*
Low	Average	*-0.002, 0.103* ▲	*-0.001, 0.08*	*NA*	*-0.008, 0.035* ▲	*-0.004, 0.008* ▲	*0.041, 0.047*	*0.008, 0.013*	*-0.004, 0.009* ▲
Low	Old	*-0.011, 0.035* ▲	*0.004, 0.075*	*0.013, 0.019* ▲	*-0.02, 0.021* ▲	*-0.006, 0.009*	*0.008, 0.011* ▲	*0.026, 0.026* ▲	*-0.007, 0.01*
Medium	Young	*-0.01, 0.116*	*-0.051, 0.086*	*NA*	*-0.006, 0.037*	*-0.012, 0.013* ▲	*0.012, 0.029*	*-0.003, 0.012*	*0.009, 0.012* ▲
Medium	Average	*0.002, 0.111*	*-0.051, 0.085*	*NA*	*0.001, 0.036*	*-0.01, 0.011* ▲	*0.038, 0.045*	*0.002, 0.012*	*0.006, 0.011* ▲
Medium	Old	*-0.021, 0.037* ▲	*-0.027, 0.078*	*-0.014, 0.021* ▲	*-0.029, 0.029* ▲	*0.011, 0.012* ▲	*-0.018, 0.02* ▲	*0.018, 0.018* ▲	*0.005, 0.01*
High	Young	*-0.108, 0.125*	*-0.075, 0.082*	*-0.061, 0.095* ▲	*-0.053, 0.056* ▲	*-0.006, 0.008* ▲	*-0.006, 0.02*	*-0.003, 0.011* ▲	*0.033, 0.033* ▲
High	Average	*-0.076, 0.105*	*-0.074, 0.081*	*-0.02, 0.086*	*-0.025, 0.037* ▲	*-0.003, 0.007* ▲	*0.027, 0.032*	*-0.002, 0.011*	*0.026, 0.026* ▲
High	Old	*-0.002, 0.044* ▲	*-0.037, 0.058* ▲	*-0.039, 0.04* ▲	*0.011, 0.02* ▲	*-0.008, 0.01*	*-0.039, 0.039*	*0.007, 0.011* ▲	*0.009, 0.012*

Log-logistic distributed data

Table 37:

Legend for filtered colour-coded table and times a GPM incorporating method ranked best for absolute mean bias of survival probability at time t for log-logistic distributed data. *Amount of times the method ranked best*

	Non-GPM	Internal additive	External additive	Converging
Times method ranked best	*11*	*21*	*20*	*20*

Table 38:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of survival probability at time t per scenario for log-logistic distributed data. Survival and ages as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

Survival (Log-logistic)

		Survival: High		Survival: Medium			Survival: Low		
		Information		*Information*			*Information*		
Age	*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium*	High*
Young	Low	*-0.034, 0.137*	*0.018, 0.028*	*0.005, 0.09*	*-0.001, 0.027*	*0.004, 0.011*	*0, 0.057*	*-0.004, 0.01*	*-0.004, 0.009*
Young	Medium	*-0.032, 0.132*	*0.018, 0.028*	*0.001, 0.089*	*0.003, 0.027*	*0.004, 0.011*	*-0.021, 0.058*	*-0.007, 0.011*	▲ *-0.009, 0.011*
Young	High	*-0.004, 0.065*	*0.012, 0.02*	*-0.027, 0.069*	*0.001, 0.025*	*0.004, 0.011*	*-0.025, 0.045*	▲ *-0.006, 0.01*	▲ *-0.003, 0.008*
Average	Low	▲ *0.039, 0.121*	*0.055, 0.057*	▲ *0.004, 0.083*	▲ *-0.003, 0.025*	▲ *-0.01, 0.013*	*0, 0.055*	*-0.002, 0.01*	▲ *-0.002, 0.008*
Average	Medium	▲ *0.041, 0.12*	*0.056, 0.057*	▲ *-0.006, 0.076*	▲ *-0.006, 0.025*	▲ *-0.011, 0.014*	*-0.018, 0.056*	▲ *-0.005, 0.011*	▲ *-0.007, 0.01*
Average	High	*0.044, 0.071*	*-0.054, 0.054*	*0.001, 0.066*	▲ *-0.012, 0.025*	▲ *-0.015, 0.017*	*-0.018, 0.044*	▲ *-0.002, 0.008*	▲ *0.001, 0.008*
Old	Low	▲ *0.017, 0.02*	▲ *0.009, 0.012*	▲ *-0.001, 0.028*	▲ *-0.009, 0.013*	▲ *-0.019, 0.02*	▲ *0.005, 0.051*	*0.005, 0.011*	*0, 0.008*
Old	Medium	▲ *-0.009, 0.018*	▲ *-0.018, 0.019*	▲ *-0.01, 0.028*	▲ *-0.015, 0.017*	▲ *-0.025, 0.025*	*0.005, 0.051*	*0, 0.01*	*-0.007, 0.01*
Old	High	▲ *-0.033, 0.035*	▲ *-0.04, 0.04*	▲ *-0.028, 0.037*	▲ *-0.031, 0.032*	▲ *0.02, 0.021*	*-0.001, 0.04*	*-0.001, 0.009*	*-0.02, 0.02*

Table 39:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of survival probability at time t per scenario for log-logistic distributed data. Information and heterogeneity as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

		Information: Low		Information: Medium			Information: High		
		Survival (Log-logistic)		*Survival (Log-logistic)*			*Survival (Log-logistic)*		
Heterogeneity	Age	Medium	Low	High	Medium	Low*	High	Medium	Low*
Low	Young	0.005, 0.09	0, 0.057	-0.034, 0.137	-0.001, 0.027	-0.004, 0.01	0.018, 0.028	0.004, 0.011	-0.004, 0.009
Low	Average	▲ 0.004, 0.083	0, 0.055	▲ 0.039, 0.121	▲ -0.003, 0.025	-0.002, 0.01	0.055, 0.057	▲ -0.01, 0.013	▲ -0.002, 0.008
Low	Old	▲ -0.001, 0.028	▲ 0.005, 0.051	▲ 0.017, 0.02	▲ -0.009, 0.013	0.005, 0.011	▲ 0.009, 0.012	▲ -0.019, 0.02	0, 0.008
Medium	Young	0.001, 0.089	-0.021, 0.058	-0.032, 0.132	0.003, 0.027	-0.007, 0.011	0.018, 0.028	0.004, 0.011	▲ -0.009, 0.011
Medium	Average	▲ -0.006, 0.076	-0.018, 0.056	▲ 0.041, 0.12	▲ -0.006, 0.025	▲ -0.005, 0.011	0.056, 0.057	▲ -0.011, 0.014	▲ -0.007, 0.01
Medium	Old	▲ -0.01, 0.028	0.005, 0.051	▲ -0.009, 0.018	▲ -0.015, 0.017	0, 0.01	▲ -0.018, 0.019	▲ -0.025, 0.025	-0.007, 0.01
High	Young	-0.027, 0.069	-0.025, 0.045	-0.004, 0.065	0.001, 0.025	▲ -0.006, 0.01	0.012, 0.02	0.004, 0.011	▲ -0.003, 0.008
High	Average	0.001, 0.066	-0.018, 0.044	0.044, 0.071	▲ -0.012, 0.025	▲ -0.002, 0.008	-0.054, 0.054	▲ -0.015, 0.017	▲ 0.001, 0.008
High	Old	▲ -0.028, 0.037	-0.001, 0.04	▲ -0.033, 0.035	▲ -0.031, 0.032	-0.001, 0.009	▲ -0.04, 0.04	▲ 0.02, 0.021	-0.02, 0.02

Lognormal distributed data

Table 40:

Legend for filtered colour-coded table and times a GPM incorporating method ranked best for absolute mean bias of survival probability at time t for lognormal distributed data. *Amount of times the method ranked best*

	Non-GPM	Internal additive	External additive	Converging
Times method ranked best	*6*	*15*	*37*	*13*

Table 41:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of survival probability at time t per scenario for lognormal distributed data. Survival and ages as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

		Survival (Lognormal)							
		High		Medium			Low		
		Information		Information			Information		
Age	Heterogeneity	Medium	High	Low	Medium	High	Low	Medium*	High*
Young	Low	▲ 0.008, 0.125	-0.027, 0.03	-0.003, 0.078	-0.002, 0.025	0.003, 0.011	-0.002, 0.059	-0.004, 0.009	-0.004, 0.009
Young	Medium	NA	▲ -0.024, 0.027	-0.001, 0.077	-0.002, 0.026	0.004, 0.011	-0.008, 0.059	-0.004, 0.009	-0.004, 0.009
Young	High	▲ -0.003, 0.044	▲ -0.013, 0.02	-0.001, 0.064	▲ 0.003, 0.024	▲ -0.002, 0.01	-0.001, 0.041	-0.001, 0.008	▲ -0.001, 0.009
Average	Low	▲ -0.027, 0.069	▲ -0.06, 0.06	▲ 0.002, 0.073	▲ -0.003, 0.024	▲ -0.011, 0.014	0.001, 0.06	-0.002, 0.009	▲ -0.002, 0.008
Average	Medium	▲ -0.021, 0.065	▲ -0.055, 0.055	▲ 0.002, 0.071	▲ -0.004, 0.024	▲ -0.012, 0.014	-0.001, 0.058	▲ -0.001, 0.009	▲ -0.001, 0.009
Average	High	▲ 0, 0.044	▲ -0.033, 0.037	▲ 0.004, 0.058	▲ 0.004, 0.023	▲ -0.009, 0.014	0, 0.043	-0.002, 0.008	▲ 0.003, 0.01
Old	Low	▲ 0.02, 0.021	▲ 0.008, 0.011	▲ -0.002, 0.027	▲ -0.009, 0.014	▲ -0.021, 0.021	0.017, 0.061	0.007, 0.012	0, 0.009
Old	Medium	▲ -0.007, 0.018	▲ -0.019, 0.021	▲ -0.006, 0.028	▲ -0.014, 0.017	▲ -0.026, 0.026	▲ 0.005, 0.054	0.005, 0.011	-0.004, 0.009
Old	High	▲ -0.032, 0.035	▲ -0.043, 0.043	▲ -0.016, 0.031	▲ -0.025, 0.027	▲ 0.02, 0.021	▲ -0.013, 0.04	0.004, 0.009	-0.021, 0.021

Table 42:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of survival probability at time t per scenario for lognormal distributed data. Information and heterogeneity as main axes.
*(Bias, RMSE), ▲ = Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

		Information: Low		Information: Medium			Information: High		
		Survival (Lognormal)		Survival (Lognormal)			Survival (Lognormal)		
Heterogeneity	Age	Medium	Low	High	Medium	Low*	High	Medium	Low*
Low	Young	-0.003, 0.078	-0.002, 0.059	▲ 0.008, 0.125	-0.002, 0.025	-0.004, 0.009	-0.027, 0.03	0.003, 0.011	-0.004, 0.009
Low	Average	▲ 0.002, 0.073	0.001, 0.06	▲ -0.027, 0.069	▲ -0.003, 0.024	-0.002, 0.009	▲ -0.06, 0.06	▲ -0.011, 0.014	▲ -0.002, 0.008
Low	Old	▲ -0.002, 0.027	0.017, 0.061	▲ 0.02, 0.021	▲ -0.009, 0.014	0.007, 0.012	▲ 0.008, 0.011	▲ -0.021, 0.021	0, 0.009
Medium	Young	-0.001, 0.077	-0.008, 0.059	NA	-0.002, 0.026	-0.004, 0.009	▲ -0.024, 0.027	0.004, 0.011	-0.004, 0.009
Medium	Average	▲ 0.002, 0.071	-0.001, 0.058	▲ -0.021, 0.065	▲ -0.004, 0.024	▲ -0.001, 0.009	▲ -0.055, 0.055	▲ -0.012, 0.014	▲ -0.001, 0.009
Medium	Old	▲ -0.006, 0.028	▲ 0.005, 0.054	▲ -0.007, 0.018	▲ -0.014, 0.017	0.005, 0.011	▲ -0.019, 0.021	▲ -0.026, 0.026	-0.004, 0.009
High	Young	-0.001, 0.064	-0.001, 0.041	▲ -0.003, 0.044	▲ 0.003, 0.024	-0.001, 0.008	▲ -0.013, 0.02	▲ -0.002, 0.01	▲ -0.001, 0.009
High	Average	▲ 0.004, 0.058	0, 0.043	▲ 0, 0.044	▲ 0.004, 0.023	-0.002, 0.008	▲ -0.033, 0.037	▲ -0.009, 0.014	▲ 0.003, 0.01
High	Old	▲ -0.016, 0.031	▲ -0.013, 0.04	▲ -0.032, 0.035	▲ -0.025, 0.027	0.004, 0.009	▲ -0.043, 0.043	▲ 0.02, 0.021	-0.021, 0.021

Gompertz distributed data

Table 43:

Legend for filtered colour-coded table and times a GPM incorporating method ranked best for absolute mean bias of survival probability at time t for Gompertz distributed data. *Amount of times the method ranked best*

	Non-GPM	Internal additive	External additive	Converging
Times method ranked best	*16*	*14*	*30*	*12*

Table 44:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of survival probability at time t per scenario for Gompertz distributed data. Survival and ages as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

Age	Heterogeneity	Survival (Gompertz): High, Information: Medium	High, Information: High	Medium, Information: Low	Medium, Information: Medium	Medium, Information: High	Low, Information: Low	Low, Information: Medium*	Low, Information: High*
Young	Low	▲ -0.198, 0.375	-0.007, 0.054	0.147, 0.314	0.009, 0.117	-0.001, 0.016	0.011, 0.144	▲ -0.005, 0.007	-0.005, 0.009
Young	Medium	▲ -0.167, 0.369	-0.012, 0.059	0.139, 0.317	0.033, 0.117	▲ 0.001, 0.015	-0.004, 0.139	▲ -0.004, 0.007	-0.003, 0.009
Young	High	▲ 0.015, 0.19	▲ 0.011, 0.029	0.288, 0.311	▲ 0.191, 0.191	▲ 0.027, 0.027	0.12, 0.124	▲ -0.002, 0.005	▲ -0.05, 0.05
Average	Low	▲ -0.127, 0.376	-0.014, 0.068	0.143, 0.292	▲ 0.003, 0.101	0.002, 0.015	0.005, 0.14	▲ -0.003, 0.006	▲ -0.003, 0.009
Average	Medium	▲ -0.117, 0.378	-0.014, 0.067	0.118, 0.279	▲ 0.022, 0.104	0.011, 0.018	-0.001, 0.138	▲ -0.002, 0.006	▲ 0, 0.008
Average	High	▲ 0.003, 0.168	▲ -0.033, 0.042	▲ 0.27, 0.292	▲ 0.174, 0.174	▲ 0.011, 0.015	▲ 0.117, 0.121	▲ 0.001, 0.005	▲ -0.046, 0.046
Old	Low	▲ -0.023, 0.042	▲ -0.007, 0.016	▲ 0.04, 0.089	▲ -0.018, 0.032	0.008, 0.012	0.01, 0.136	-0.006, 0.007	-0.007, 0.01
Old	Medium	▲ -0.044, 0.05	-0.003, 0.03	▲ 0.034, 0.088	▲ -0.02, 0.033	▲ 0.01, 0.013	0.001, 0.128	-0.005, 0.007	-0.004, 0.009
Old	High	▲ -0.039, 0.042	▲ -0.035, 0.035	▲ 0.064, 0.082	▲ 0.02, 0.027	▲ 0.004, 0.011	▲ 0.087, 0.096	-0.003, 0.006	▲ 0.004, 0.012

Table 45:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of survival probability at time t per scenario for Gompertz distributed data. Information and heterogeneity as main axes.
*(Bias, RMSE), ▲ = Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

		Information							
		Low		Medium			High		
		Survival (Gompertz)		Survival (Gompertz)			Survival (Gompertz)		
Heterogeneity	Age	Medium	Low	High	Medium	Low*	High	Medium	Low*
Low	Young	0.147, 0.314	0.011, 0.144	-0.198, 0.375 ▲	0.009, 0.117	-0.005, 0.007 ▲	-0.007, 0.054	-0.001, 0.016	-0.005, 0.009
Low	Average	0.143, 0.292	0.005, 0.14	-0.127, 0.376 ▲	0.003, 0.101 ▲	-0.003, 0.006 ▲	-0.014, 0.068	0.002, 0.015	-0.003, 0.009 ▲
Low	Old	0.04, 0.089 ▲	0.01, 0.136	-0.023, 0.042 ▲	-0.018, 0.032 ▲	-0.006, 0.007	-0.007, 0.016 ▲	0.008, 0.012	-0.007, 0.01
Medium	Young	0.139, 0.317	-0.004, 0.139	-0.167, 0.369 ▲	0.033, 0.117	-0.004, 0.007 ▲	-0.012, 0.059	0.001, 0.015 ▲	-0.003, 0.009
Medium	Average	0.118, 0.279	-0.001, 0.138	-0.117, 0.378 ▲	0.022, 0.104 ▲	-0.002, 0.006 ▲	-0.014, 0.067	0.011, 0.018	0, 0.008 ▲
Medium	Old	0.034, 0.088 ▲	0.001, 0.128	-0.044, 0.05 ▲	-0.02, 0.033 ▲	-0.005, 0.007	-0.003, 0.03	0.01, 0.013 ▲	-0.004, 0.009
High	Young	0.288, 0.311	0.12, 0.124	0.015, 0.19 ▲	0.191, 0.191 ▲	-0.002, 0.005 ▲	0.011, 0.029 ▲	0.027, 0.027 ▲	-0.05, 0.05 ▲
High	Average	0.27, 0.292 ▲	0.117, 0.121 ▲	0.003, 0.168 ▲	0.174, 0.174 ▲	0.001, 0.005 ▲	-0.033, 0.042 ▲	0.011, 0.015 ▲	-0.046, 0.046 ▲
High	Old	0.064, 0.082 ▲	0.087, 0.096 ▲	-0.039, 0.042 ▲	0.02, 0.027 ▲	-0.003, 0.006	-0.035, 0.035 ▲	0.004, 0.011 ▲	0.004, 0.012 ▲

A.3: RMST

Weibull distributed data

Table 46:

Legend for filtered colour-coded table and times a GPM incorporating method ranked best for absolute mean bias of RMST for Weibull distributed data. *Amount of times the method ranked best*

	Non-GPM	Internal additive	External additive	Converging
Times method ranked best	*26*	*22*	*12*	*8*

Table 47:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of RMST per scenario for Weibull distributed data. Survival and ages as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

		Survival (Weibull)							
		High		Medium			Low		
		Information		*Information*			*Information*		
Age	*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium	High
Young	Low	*NA*	*-0.007, 0.02*	*-0.003, 0.008*	*0, 0.015* ▲	*-0.015, 0.04*	*-0.015, 0.015* ▲	*-0.037, 0.037* ▲	*-0.046, 0.046* ▲
Young	Medium	*NA*	*-0.008, 0.021*	*-0.002, 0.008*	*0.002, 0.016*	*0.005, 0.037*	*-0.016, 0.016* ▲	*-0.024, 0.025* ▲	*-0.029, 0.029* ▲
Young	High	*0.001, 0.008*	*0.026, 0.036*	*0.006, 0.012*	*0.039, 0.04* ▲	*0.128, 0.128* ▲	*0.001, 0.011*	*0.04, 0.041* ▲	*0.121, 0.121* ▲
Average	Low	*NA*	*-0.018, 0.026*	*-0.002, 0.008* ▲	*0.008, 0.016*	*-0.026, 0.043*	*-0.014, 0.014* ▲	*-0.032, 0.032* ▲	*-0.036, 0.036* ▲
Average	Medium	*NA*	*-0.018, 0.028*	*-0.001, 0.008* ▲	*-0.006, 0.016*	*-0.006, 0.038*	*-0.015, 0.015* ▲	*-0.02, 0.021* ▲	*-0.014, 0.017* ▲
Average	High	*-0.001, 0.008*	*0.004, 0.032*	*0.004, 0.011*	*0.028, 0.031* ▲	*0.017, 0.05* ▲	*-0.001, 0.011*	*0.032, 0.033* ▲	*0.059, 0.063* ▲
Old	Low	*-0.009, 0.011* ▲	*-0.087, 0.087* ▲	*-0.006, 0.009*	*-0.018, 0.022*	*-0.082, 0.082*	*-0.003, 0.007* ▲	*0.023, 0.024* ▲	*-0.053, 0.053*
Old	Medium	*-0.009, 0.012* ▲	*-0.077, 0.078* ▲	*-0.006, 0.009*	*-0.016, 0.021*	*-0.067, 0.068*	*-0.004, 0.008* ▲	*-0.032, 0.033*	*-0.045, 0.045*
Old	High	*-0.009, 0.014*	*-0.051, 0.058*	*0.001, 0.011*	*0.018, 0.024*	*-0.002, 0.036*	*-0.002, 0.011*	*0.024, 0.027*	*0.025, 0.034*

Table 48:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of RMST per scenario for Weibull distributed data. Information and heterogeneity as main axes. *(Bias, RMSE),* ▲ *= Significant result (Wilcoxon)*

		Information							
		Low		Medium			High		
		Survival (Weibull)		*Survival (Weibull)*			*Survival (Weibull)*		
Heterogeneity	Age	Medium	Low	High	Medium	Low	High	Medium	Low
Low	Young	*-0.003, 0.008*	*-0.015, 0.015* ▲	*NA*	*0, 0.015* ▲	*-0.037, 0.037* ▲	*-0.007, 0.02*	*-0.015, 0.04*	*-0.046, 0.046* ▲
Low	Average	*-0.002, 0.008* ▲	*-0.014, 0.014* ▲	*NA*	*0.008, 0.016*	*-0.032, 0.032* ▲	*-0.018, 0.026*	*-0.026, 0.043*	*-0.036, 0.036* ▲
Low	Old	*-0.006, 0.009*	*-0.003, 0.007* ▲	*-0.009, 0.011* ▲	*-0.018, 0.022*	*0.023, 0.024* ▲	*-0.087, 0.087* ▲	*-0.082, 0.082*	*-0.053, 0.053*
Medium	Young	*-0.002, 0.008*	*-0.016, 0.016* ▲	*NA*	*0.002, 0.016*	*-0.024, 0.025* ▲	*-0.008, 0.021*	*0.005, 0.037*	*-0.029, 0.029* ▲
Medium	Average	*-0.001, 0.008* ▲	*-0.015, 0.015* ▲	*NA*	*-0.006, 0.016*	*-0.02, 0.021* ▲	*-0.018, 0.028*	*-0.006, 0.038*	*-0.014, 0.017* ▲
Medium	Old	*-0.006, 0.009*	*-0.004, 0.008* ▲	*-0.009, 0.012* ▲	*-0.016, 0.021*	*-0.032, 0.033*	*-0.077, 0.078* ▲	*-0.067, 0.068*	*-0.045, 0.045*
High	Young	*0.006, 0.012*	*0.001, 0.011*	*0.001, 0.008*	*0.039, 0.04* ▲	*0.04, 0.041* ▲	*0.026, 0.036*	*0.128, 0.128* ▲	*0.121, 0.121* ▲
High	Average	*0.004, 0.011*	*-0.001, 0.011*	*-0.001, 0.008*	*0.028, 0.031* ▲	*0.032, 0.033* ▲	*0.004, 0.032*	*0.017, 0.05* ▲	*0.059, 0.063* ▲
High	Old	*0.001, 0.011*	*-0.002, 0.011*	*-0.009, 0.014*	*0.018, 0.024*	*0.024, 0.027*	*-0.051, 0.058*	*-0.002, 0.036*	*0.025, 0.034*

Log-logistic distributed data

Table 49:

Legend for filtered colour-coded table and times a GPM incorporating method ranked best for absolute mean bias of RMST for log-logistic distributed data. *Amount of times the method ranked best*

	Non-GPM	Internal additive	External additive	Converging
Times method ranked best	27	35	3	7

Table 50:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of RMST per scenario for log-logistic distributed data. Survival and ages as main axes. *(Bias, RMSE),* ▲ *= Significant result (Wilcoxon)*

		Survival (Log-logistic)							
		High		Medium			Low		
		Information		*Information*			*Information*		
Age	*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium	High
Young	Low	0, 0.004	-0.008, 0.019	-0.007, 0.008 ▲	-0.014, 0.017 ▲	0.011, 0.029 ▲	-0.015, 0.015 ▲	-0.038, 0.038 ▲	-0.04, 0.043 ▲
Young	Medium	-0.001, 0.004	-0.008, 0.02	-0.007, 0.009 ▲	-0.012, 0.016 ▲	0.014, 0.03 ▲	-0.016, 0.016 ▲	-0.04, 0.04 ▲	-0.039, 0.043 ▲
Young	High	0, 0.008	0.007, 0.028	0, 0.009	0.007, 0.016 ▲	-0.024, 0.04	-0.01, 0.013	-0.01, 0.016 ▲	-0.018, 0.034 ▲
Average	Low	-0.001, 0.005 ▲	-0.022, 0.028	-0.006, 0.008 ▲	-0.005, 0.013 ▲	-0.045, 0.048	-0.015, 0.015 ▲	-0.034, 0.034 ▲	-0.027, 0.035 ▲
Average	Medium	-0.001, 0.005	-0.021, 0.028	-0.006, 0.008 ▲	-0.003, 0.013 ▲	-0.046, 0.049	-0.016, 0.016 ▲	-0.035, 0.035 ▲	-0.02, 0.031 ▲
Average	High	-0.001, 0.008	-0.019, 0.032	0, 0.01	-0.005, 0.016 ▲	0, 0.033	-0.009, 0.012 ▲	-0.005, 0.015 ▲	0.041, 0.051 ▲
Old	Low	-0.01, 0.012 ▲	-0.101, 0.101 ▲	0.007, 0.009 ▲	-0.03, 0.03	-0.043, 0.046	-0.004, 0.006 ▲	0.016, 0.018 ▲	0, 0.026
Old	Medium	-0.011, 0.013 ▲	-0.093, 0.093 ▲	0.006, 0.008 ▲	-0.029, 0.03	-0.043, 0.046	-0.006, 0.008 ▲	0.015, 0.018 ▲	0.001, 0.027
Old	High	-0.013, 0.016	-0.091, 0.092 ▲	-0.006, 0.011	-0.014, 0.019	-0.041, 0.048 ▲	-0.003, 0.01 ▲	-0.031, 0.031	-0.02, 0.036

Table 51:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of RMST per scenario for log-logistic distributed data. Information and heterogeneity as main axes. *(Bias, RMSE)*, ▲ *= Significant result (Wilcoxon)*

Heterogeneity	Age	Information: Low Survival (Log-logistic) Medium	Information: Low Survival (Log-logistic) Low	Information: Medium Survival (Log-logistic) High	Information: Medium Survival (Log-logistic) Medium	Information: Medium Survival (Log-logistic) Low	Information: High Survival (Log-logistic) High	Information: High Survival (Log-logistic) Medium	Information: High Survival (Log-logistic) Low
Low	Young	*-0.007, 0.008* ▲	*-0.015, 0.015* ▲	*0, 0.004*	*-0.014, 0.017* ▲	*-0.038, 0.038* ▲	*-0.008, 0.019*	*0.011, 0.029* ▲	*-0.04, 0.043* ▲
Low	Average	*-0.006, 0.008* ▲	*-0.015, 0.015* ▲	*-0.001, 0.005* ▲	*-0.005, 0.013* ▲	*-0.034, 0.034* ▲	*-0.022, 0.028*	*-0.045, 0.048*	*-0.027, 0.035* ▲
Low	Old	*0.007, 0.009* ▲	*-0.004, 0.006* ▲	*-0.01, 0.012* ▲	*-0.03, 0.03*	*0.016, 0.018* ▲	*-0.101, 0.101* ▲	*-0.043, 0.046*	*0, 0.026*
Medium	Young	*-0.007, 0.009* ▲	*-0.016, 0.016* ▲	*-0.001, 0.004*	*-0.012, 0.016* ▲	*-0.04, 0.04* ▲	*-0.008, 0.02*	*0.014, 0.03* ▲	*-0.039, 0.043* ▲
Medium	Average	*-0.006, 0.008* ▲	*-0.016, 0.016* ▲	*-0.001, 0.005*	*-0.003, 0.013* ▲	*-0.035, 0.035* ▲	*-0.021, 0.028*	*-0.046, 0.049*	*-0.02, 0.031* ▲
Medium	Old	*0.006, 0.008* ▲	*-0.006, 0.008* ▲	*-0.011, 0.013* ▲	*-0.029, 0.03*	*0.015, 0.018* ▲	*-0.093, 0.093* ▲	*-0.043, 0.046*	*0.001, 0.027*
High	Young	*0, 0.009*	*-0.01, 0.013*	*0, 0.008*	*0.007, 0.016* ▲	*-0.01, 0.016* ▲	*0.007, 0.028*	*-0.024, 0.04*	*-0.018, 0.034* ▲
High	Average	*0, 0.01*	*-0.009, 0.012* ▲	*-0.001, 0.008*	*-0.005, 0.016* ▲	*-0.005, 0.015* ▲	*-0.019, 0.032*	*0, 0.033*	*0.041, 0.051* ▲
High	Old	*-0.006, 0.011*	*-0.003, 0.01* ▲	*-0.013, 0.016*	*-0.014, 0.019*	*-0.031, 0.031*	*-0.091, 0.092* ▲	*-0.041, 0.048* ▲	*-0.02, 0.036*

Lognormal distributed data

Table 52:

Legend for filtered colour-coded table and times a GPM incorporating method ranked best for absolute mean bias of RMST for lognormal distributed data. *Amount of times the method ranked best*

	Non-GPM	Internal additive	External additive	Converging
Times method ranked best	*26*	*42*	*0*	*3*

Table 53:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of RMST per scenario for lognormal distributed data. Survival and ages as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

			Survival (Lognormal)							
			High		Medium			Low		
			Information		*Information*			*Information*		
Age		*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium	High
Young		Low	*-0.001, 0.004* ▲	*-0.016, 0.026*	*-0.011, 0.011* ▲	*-0.021, 0.023* ▲	*-0.006, 0.034* ▲	*-0.016, 0.016* ▲	*-0.037, 0.037* ▲	*-0.041, 0.041* ▲
		Medium	*NA*	*-0.017, 0.027*	*-0.012, 0.012* ▲	*-0.022, 0.024* ▲	*-0.002, 0.034* ▲	*-0.017, 0.017* ▲	*-0.035, 0.035* ▲	*-0.037, 0.038* ▲
		High	*0.004, 0.009*	*-0.035, 0.042*	*-0.008, 0.012*	*-0.005, 0.017* ▲	*-0.02, 0.042*	*-0.02, 0.02*	*-0.028, 0.029* ▲	*-0.008, 0.034* ▲
Average		Low	*-0.003, 0.006* ▲	*-0.041, 0.045* ▲	*-0.01, 0.011* ▲	*-0.012, 0.017* ▲	*-0.064, 0.066*	*-0.015, 0.015* ▲	*-0.032, 0.032* ▲	*-0.028, 0.029* ▲
		Medium	*-0.004, 0.006* ▲	*-0.044, 0.047* ▲	*-0.01, 0.011* ▲	*-0.014, 0.019* ▲	*-0.068, 0.07*	*-0.016, 0.016* ▲	*-0.029, 0.03* ▲	*-0.019, 0.023* ▲
		High	*-0.008, 0.011*	*-0.073, 0.075*	*-0.008, 0.012* ▲	*0.001, 0.017* ▲	*-0.05, 0.057*	*-0.019, 0.02* ▲	*-0.023, 0.025* ▲	*0.049, 0.055* ▲
Old		Low	*-0.019, 0.02* ▲	*-0.211, 0.211* ▲	*0.002, 0.007* ▲	*-0.04, 0.04*	*-0.11, 0.11*	*-0.003, 0.007* ▲	*0.019, 0.021* ▲	*-0.016, 0.026* ▲
		Medium	*-0.021, 0.022* ▲	*-0.197, 0.197* ▲	*0.001, 0.007* ▲	*-0.043, 0.043*	*-0.106, 0.106*	*-0.005, 0.008* ▲	*0.022, 0.023* ▲	*-0.019, 0.028*
		High	*-0.029, 0.029*	*-0.187, 0.187* ▲	*0, 0.011* ▲	*-0.041, 0.042*	*-0.12, 0.12* ▲	*-0.012, 0.015* ▲	*0.03, 0.033* ▲	*-0.065, 0.066*

Table 54:

Filtered colour-coded table for best ranking GPM incorporating method for absolute mean bias of RMST per scenario for lognormal distributed data. Information and heterogeneity as main axes. *(Bias, RMSE),* ▲ *= Significant result (Wilcoxon)*

			Information							
			Low		Medium			High		
			Survival (Lognormal)		*Survival (Lognormal)*			*Survival (Lognormal)*		
Heterogeneity			Medium	Low	High	Medium	Low	High	Medium	Low
Low	Age	Young	▲ *-0.011, 0.011*	▲ *-0.016, 0.016*	▲ *-0.001, 0.004*	▲ *-0.021, 0.023*	▲ *-0.037, 0.037*	*-0.016, 0.026*	▲ *-0.006, 0.034*	▲ *-0.041, 0.041*
		Average	▲ *-0.01, 0.011*	▲ *-0.015, 0.015*	▲ *-0.003, 0.006*	▲ *-0.012, 0.017*	▲ *-0.032, 0.032*	▲ *-0.041, 0.045*	*-0.064, 0.066*	▲ *-0.028, 0.029*
		Old	▲ *0.002, 0.007*	▲ *-0.003, 0.007*	▲ *-0.019, 0.02*	*-0.04, 0.04*	▲ *0.019, 0.021*	▲ *-0.211, 0.211*	*-0.11, 0.11*	▲ *-0.016, 0.026*
Medium	Age	Young	▲ *-0.012, 0.012*	▲ *-0.017, 0.017*	*NA*	▲ *-0.022, 0.024*	▲ *-0.035, 0.035*	*-0.017, 0.027*	▲ *-0.002, 0.034*	▲ *-0.037, 0.038*
		Average	▲ *-0.01, 0.011*	▲ *-0.016, 0.016*	▲ *-0.004, 0.006*	▲ *-0.014, 0.019*	▲ *-0.029, 0.03*	▲ *-0.044, 0.047*	*-0.068, 0.07*	▲ *-0.019, 0.023*
		Old	▲ *0.001, 0.007*	▲ *-0.005, 0.008*	▲ *-0.021, 0.022*	*-0.043, 0.043*	▲ *0.022, 0.023*	▲ *-0.197, 0.197*	*-0.106, 0.106*	*-0.019, 0.028*
High	Age	Young	*-0.008, 0.012*	*-0.02, 0.02*	*0.004, 0.009*	▲ *-0.005, 0.017*	▲ *-0.028, 0.029*	*-0.035, 0.042*	*-0.02, 0.042*	▲ *-0.008, 0.034*
		Average	▲ *-0.008, 0.012*	▲ *-0.019, 0.02*	*-0.008, 0.011*	▲ *0.001, 0.017*	▲ *-0.023, 0.025*	*-0.073, 0.075*	*-0.05, 0.057*	▲ *0.049, 0.055*
		Old	▲ *0, 0.011*	▲ *-0.012, 0.015*	*-0.029, 0.029*	*-0.041, 0.042*	▲ *0.03, 0.033*	▲ *-0.187, 0.187*	▲ *-0.12, 0.12*	*-0.065, 0.066*

Gompertz distributed data

Table 55:

Legend for colour-coded table and times a GPM incorporating method ranked best for absolute mean bias of RMST for Gompertz distributed data. *Amount of times the method ranked best*

	Non-GPM	Internal additive	External additive	Converging
Times method ranked best	*28*	*36*	*0*	*8*

Table 56:

Colour-coded table for best ranking GPM incorporating method for absolute mean bias of RMST per scenario for Gompertz distributed data. Survival and ages as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

		Survival (Gompertz)							
		High		Medium			Low		
		Information		*Information*			*Information*		
Age	*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium	High
Young	Low	-0.001, 0.003	-0.006, 0.012	-0.007, 0.007	-0.01, 0.012 ▲	0.013, 0.025 ▲	-0.017, 0.017 ▲	-0.035, 0.035 ▲	-0.042, 0.042 ▲
Young	Medium	-0.002, 0.003	-0.007, 0.013	-0.008, 0.008	-0.012, 0.013 ▲	0.012, 0.026 ▲	-0.018, 0.018 ▲	-0.033, 0.033 ▲	-0.038, 0.038 ▲
Young	High	-0.007, 0.008	-0.019, 0.024	-0.012, 0.014 ▲	-0.026, 0.026 ▲	-0.076, 0.076 ▲	-0.024, 0.024 ▲	-0.058, 0.058 ▲	-0.214, 0.214 ▲
Average	Low	-0.002, 0.004 ▲	-0.007, 0.014	-0.007, 0.008	-0.002, 0.008 ▲	-0.039, 0.041	-0.016, 0.016 ▲	-0.03, 0.03 ▲	-0.032, 0.032 ▲
Average	Medium	-0.002, 0.004	-0.009, 0.015	-0.008, 0.008	-0.002, 0.009 ▲	-0.042, 0.043	-0.017, 0.017 ▲	-0.028, 0.028 ▲	-0.02, 0.022 ▲
Average	High	-0.007, 0.009	-0.022, 0.027	-0.009, 0.015 ▲	-0.018, 0.019 ▲	0.033, 0.05 ▲	-0.023, 0.023 ▲	-0.052, 0.052 ▲	-0.141, 0.141 ▲
Old	Low	-0.007, 0.008 ▲	-0.027, 0.03	-0.009, 0.009	-0.021, 0.022	-0.046, 0.047	-0.002, 0.011 ▲	0.024, 0.024 ▲	-0.049, 0.049
Old	Medium	-0.008, 0.01	-0.027, 0.031	-0.01, 0.01	-0.023, 0.023	-0.051, 0.051	-0.006, 0.009 ▲	0.024, 0.025 ▲	-0.05, 0.05
Old	High	-0.014, 0.015	-0.04, 0.042	-0.016, 0.016	-0.038, 0.038	-0.139, 0.139	-0.016, 0.016 ▲	0.003, 0.016 ▲	-0.213, 0.213 ▲

Table 57:

Colour-coded table for best ranking GPM incorporating method for absolute mean bias of RMST per scenario for Gompertz distributed data. Information and heterogeneity as main axes. *(Bias, RMSE),* ▲ = *Significant result (Wilcoxon)*

			Information: Low		Information: Medium			Information: High		
			Survival (Gompertz)		*Survival (Gompertz)*			*Survival (Gompertz)*		
Heterogeneity	**Age**		Medium	Low	High	Medium	Low	High	Medium	Low
Low	Age	Young	*-0.007, 0.007*	▲ *-0.017, 0.017*	*-0.001, 0.003*	▲ *-0.01, 0.012*	▲ *-0.035, 0.035*	*-0.006, 0.012*	▲ *0.013, 0.025*	▲ *-0.042, 0.042*
Low	Age	Average	*-0.007, 0.008*	▲ *-0.016, 0.016*	▲ *-0.002, 0.004*	▲ *-0.002, 0.008*	▲ *-0.03, 0.03*	*-0.007, 0.014*	*-0.039, 0.041*	▲ *-0.032, 0.032*
Low	Age	Old	*-0.009, 0.009*	▲ *-0.002, 0.011*	▲ *-0.007, 0.008*	*-0.021, 0.022*	▲ *0.024, 0.024*	*-0.027, 0.03*	*-0.046, 0.047*	*-0.049, 0.049*
Medium	Age	Young	*-0.008, 0.008*	▲ *-0.018, 0.018*	*-0.002, 0.003*	▲ *-0.012, 0.013*	▲ *-0.033, 0.033*	*-0.007, 0.013*	▲ *0.012, 0.026*	▲ *-0.038, 0.038*
Medium	Age	Average	*-0.008, 0.008*	▲ *-0.017, 0.017*	*-0.002, 0.004*	▲ *-0.002, 0.009*	▲ *-0.028, 0.028*	*-0.009, 0.015*	*-0.042, 0.043*	▲ *-0.02, 0.022*
Medium	Age	Old	*-0.01, 0.01*	▲ *-0.006, 0.009*	*-0.008, 0.01*	*-0.023, 0.023*	▲ *0.024, 0.025*	*-0.027, 0.031*	*-0.051, 0.051*	*-0.05, 0.05*
High	Age	Young	▲ *-0.012, 0.014*	▲ *-0.024, 0.024*	*-0.007, 0.008*	▲ *-0.026, 0.026*	▲ *-0.058, 0.058*	*-0.019, 0.024*	▲ *-0.076, 0.076*	▲ *-0.214, 0.214*
High	Age	Average	▲ *-0.009, 0.015*	▲ *-0.023, 0.023*	*-0.007, 0.009*	▲ *-0.018, 0.019*	▲ *-0.052, 0.052*	*-0.022, 0.027*	▲ *0.033, 0.05*	▲ *-0.141, 0.141*
High	Age	Old	*-0.016, 0.016*	▲ *-0.016, 0.016*	*-0.014, 0.015*	*-0.038, 0.038*	▲ *0.003, 0.016*	*-0.04, 0.042*	*-0.139, 0.139*	▲ *-0.213, 0.213*

Appendix B: Colour-coded tables for best ranking methods per scenario

This Appendix shows the results of the simulation by colour-coding the tested methods and showing the best performing method for each scenario based on the absolute mean bias for the estimand. Separate tables are shown for each estimand and underlying survival distribution used to generate data separately. The DGM dimensions (survival, age, heterogeneity, level of information) are used as rows and columns for the tables, tables are shown in pairs with flipped "minor" "and major" dimension of rows and columns. Cells show the mean bias and mean RMSE for the method over 2,500 replications. A triangle denotes whether the result was significant based on a Wilcoxon signed-rank test ($\alpha = 0.05$). Each pair of tables is accompanied by another that mostly serves as a legend, but also shows the overall number of times a method ranked best and the percentage of that result which was significant for each survival distribution used to generate the data.

B.1: Overall mean survival

Weibull distributed data

Table 58:

Legend for colour-coded table and times a model ranked best for absolute mean bias of mean survival for Weibull distributed data. *Amount of times the model ranked best*

	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F	Total
Non-GPM	*0*	*4*	*0*	*0*	*5*	*2*	*1*	*12*
Internal additive	*4*	*2*	*0*	*0*	*1*	*2*	*1*	*10*
External additive	*5*	*9*	*7*	*7*	*3*	*4*	*1*	*36*
Converging	*2*	*6*	*1*	*1*	*1*	*3*	*0*	*14*
Total	*11*	*21*	*8*	*8*	*10*	*11*	*3*	

Table 59:

Colour-coded table for best ranking model for absolute mean bias of mean survival per scenario for Weibull distributed data. Survival and ages as main axes. *(Bias, RMSE),* ▲ *= Significant result (Wilcoxon)*

Age	*Heterogeneity*	Survival (Weibull): High, *Information* Medium	High, *Information* High	Medium, *Information* Low	Medium, *Information* Medium	Medium, *Information* High	Low, *Information* Low	Low, *Information* Medium	Low, *Information* High
Young	Low	*0.567, 22.586*	*-0.12, 0.765* ▲	*0.556, 3.7* ▲	*-0.011, 1.081* ▲	*-0.153, 0.339* ▲	*0.035, 0.376*	*-0.031, 0.073*	*-0.033, 0.033* ▲
Young	Medium	*-0.195, 22.2* ▲	*-0.388, 0.941* ▲	*-0.325, 3.627*	*0.05, 1.3*	*0.117, 0.407* ▲	*0.217, 0.331*	*-0.057, 0.083*	*-0.04, 0.048*
Young	High	*0.091, 5.364*	*0.028, 1.12*	*0.145, 2.593* ▲	*-0.127, 1.728* ▲	*-0.019, 0.512* ▲	*0.005, 1.368*	*0.035, 0.908* ▲	*0.064, 0.347*
Average	Low	*-0.498, 1.207* ▲	*0.719, 4.128* ▲	*0.054, 2.756* ▲	*-0.188, 0.817* ▲	*0.339, 0.514*	*0.034, 0.358*	*-0.023, 0.071*	*-0.02, 0.02* ▲
Average	Medium	*0.646, 1.953* ▲	*0.564, 3.953* ▲	*-0.447, 2.72* ▲	*0.309, 1*	*0.329, 0.395* ▲	*0.068, 1.358*	*-0.035, 0.073* ▲	*-0.007, 0.03* ▲
Average	High	*-0.056, 1.447* ▲	*-0.324, 0.597* ▲	*0.042, 1.783* ▲	*-0.076, 1.093*	*0.059, 0.432*	*-0.097, 1.027* ▲	*-0.092, 0.614* ▲	*0.007, 0.225*
Old	Low	*0.17, 0.927*	*-0.01, 0.323*	*-0.083, 1.593* ▲	*0.066, 0.482*	*0.017, 0.067* ▲	*-0.044, 0.691*	*-0.013, 0.043* ▲	*-0.01, 0.026* ▲
Old	Medium	*0.027, 0.554*	*-0.161, 0.537* ▲	*0.1, 0.778* ▲	*-0.002, 0.462*	*0.013, 0.089*	*0.021, 0.203* ▲	*-0.012, 0.065*	*0.002, 0.034* ▲
Old	High	*-0.055, 0.449*	*0.025, 0.363*	*-0.071, 0.692* ▲	*0.012, 0.315* ▲	*0.02, 0.149*	*-0.107, 0.479*	*0.193, 0.316*	*0.016, 0.076*

Table 60:

Colour-coded table for best ranking model for absolute mean bias of mean survival per scenario for Weibull distributed data. Information and heterogeneity as main axes. *(Bias, RMSE),* ▲ *= Significant result (Wilcoxon)*

			Information							
			Low		Medium			High		
			Survival (Weibull)		*Survival (Weibull)*			*Survival (Weibull)*		
Heterogeneity			Medium	Low	High	Medium	Low	High	Medium	Low
Low	**Age**	Young	▲ 0.556, 3.7	0.035, 0.376	0.567, 22.586	▲ -0.011, 1.081	-0.031, 0.073	▲ -0.12, 0.765	▲ -0.153, 0.339	▲ -0.033, 0.033
		Average	▲ 0.054, 2.756	0.034, 0.358	▲ -0.498, 1.207	▲ -0.188, 0.817	-0.023, 0.071	▲ 0.719, 4.128	0.339, 0.514	▲ -0.02, 0.02
		Old	▲ -0.083, 1.593	-0.044, 0.691	0.17, 0.927	0.066, 0.482	▲ -0.013, 0.043	-0.01, 0.323	▲ 0.017, 0.067	▲ -0.01, 0.026
Medium	**Age**	Young	-0.325, 3.627	0.217, 0.331	▲ -0.195, 22.2	0.05, 1.3	-0.057, 0.083	▲ -0.388, 0.941	▲ 0.117, 0.407	-0.04, 0.048
		Average	▲ -0.447, 2.72	0.068, 1.358	▲ 0.646, 1.953	0.309, 1	▲ -0.035, 0.073	▲ 0.564, 3.953	▲ 0.329, 0.395	▲ -0.007, 0.03
		Old	▲ 0.1, 0.778	▲ 0.021, 0.203	0.027, 0.554	-0.002, 0.462	-0.012, 0.065	▲ -0.161, 0.537	0.013, 0.089	▲ 0.002, 0.034
High	**Age**	Young	▲ 0.145, 2.593	0.005, 1.368	0.091, 5.364	▲ -0.127, 1.728	▲ 0.035, 0.908	0.028, 1.12	▲ -0.019, 0.512	0.064, 0.347
		Average	▲ 0.042, 1.783	▲ -0.097, 1.027	▲ -0.056, 1.447	-0.076, 1.093	▲ -0.092, 0.614	▲ -0.324, 0.597	0.059, 0.432	0.007, 0.225
		Old	▲ -0.071, 0.692	-0.107, 0.479	-0.055, 0.449	▲ 0.012, 0.315	0.193, 0.316	0.025, 0.363	0.02, 0.149	0.016, 0.076

Log-logistic distributed data

Table 61:

Legend for colour-coded table and times a model ranked best for absolute mean bias of mean survival for log-logistic distributed data. *Amount of times the model ranked best*

	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F	Total
Non-GPM	*2*	*1*	*1*	*0*	*5*	*4*	*2*	*15*
Internal additive	*2*	*4*	*2*	*0*	*0*	*0*	*1*	*9*
External additive	*6*	*0*	*14*	*4*	*3*	*3*	*1*	*31*
Converging	*1*	*3*	*5*	*2*	*1*	*5*	*0*	*17*
Total	*11*	*8*	*22*	*6*	*9*	*12*	*4*	

Table 62:

Colour-coded table for best ranking model for absolute mean bias of mean survival per scenario for log-logistic distributed data. Survival and ages as main axes. *(Bias, RMSE)*, ▲ *= Significant result (Wilcoxon)*

Survival (Log-logistic)

		Survival: High		Survival: Medium			Survival: Low		
		Information		*Information*			*Information*		
Age	*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium	High
Young	Low	*1.042, 6.347*	*-0.48, 0.844* ▲	*0.381, 3.233* ▲	*0.052, 0.903* ▲	*-0.042, 0.621*	*0.059, 0.435*	*-0.01, 0.113*	*-0.016, 0.063*
Young	Medium	*0.454, 2.1* ▲	*0.525, 7.918* ▲	*-0.084, 3.122*	*-0.07, 0.888* ▲	*-0.061, 0.551* ▲	*0.002, 0.336*	*0.036, 0.144*	*0.001, 0.082*
Young	High	*0.618, 2.848* ▲	*0.08, 1.066* ▲	*-0.015, 2.67*	*-0.024, 1.74* ▲	*-0.104, 0.522*	*0.208, 1.515*	*0.078, 0.541*	*0.024, 0.345*
Average	Low	*-0.456, 1.228* ▲	*0.06, 3.948* ▲	*0.121, 2.388* ▲	*-0.067, 0.625* ▲	*0.093, 0.354*	*0.024, 0.396*	*-0.002, 0.101*	*-0.006, 0.056*
Average	Medium	*-0.184, 14.968*	*0.269, 3.915* ▲	*-0.148, 2.209* ▲	*0.04, 2.064*	*0.039, 0.339*	*0.031, 0.303*	*0.013, 0.127*	*-0.012, 0.065*
Average	High	*-0.029, 4.214*	*-0.241, 0.58* ▲	*0.235, 3.827*	*-0.258, 1.075* ▲	*0.203, 0.379* ▲	*0.164, 1.14*	*0.013, 0.452*	*0.008, 0.233*
Old	Low	*0.165, 0.929*	*-0.028, 0.315*	*-0.123, 0.889*	*0.1, 0.616*	*0.028, 0.075*	*0.021, 0.289* ▲	*0.028, 0.082*	*0.004, 0.034*
Old	Medium	*0.011, 0.589*	*-0.141, 0.534*	*-0.082, 2.266*	*0.035, 0.577* ▲	*0.013, 0.08*	*-0.016, 0.206* ▲	*-0.015, 0.069* ▲	*-0.027, 0.037* ▲
Old	High	*-0.096, 0.547* ▲	*0.011, 0.384*	*0.223, 1.445*	*-0.036, 0.377*	*-0.015, 0.137*	*-0.238, 0.497* ▲	*0.081, 0.263* ▲	*-0.014, 0.076*

Table 63:

Colour-coded table for best ranking model for absolute mean bias of mean survival per scenario for log-logistic distributed data. Information and heterogeneity as main axes. *(Bias, RMSE),* ▲ *= Significant result (Wilcoxon)*

Heterogeneity	Age	Information: Low — Survival (Log-logistic): Medium	Information: Low — Survival (Log-logistic): Low	Information: Medium — Survival (Log-logistic): High	Information: Medium — Survival (Log-logistic): Medium	Information: Medium — Survival (Log-logistic): Low	Information: High — Survival (Log-logistic): High	Information: High — Survival (Log-logistic): Medium	Information: High — Survival (Log-logistic): Low
Low	Young	0.381, 3.233 ▲	0.059, 0.435	1.042, 6.347	0.052, 0.903 ▲	-0.01, 0.113	-0.48, 0.844 ▲	-0.042, 0.621	-0.016, 0.063
Low	Average	0.121, 2.388 ▲	0.024, 0.396	-0.456, 1.228 ▲	-0.067, 0.625 ▲	-0.002, 0.101	0.06, 3.948 ▲	0.093, 0.354	-0.006, 0.056
Low	Old	-0.123, 0.889	0.021, 0.289 ▲	0.165, 0.929	0.1, 0.616	0.028, 0.082	-0.028, 0.315	0.028, 0.075	0.004, 0.034
Medium	Young	-0.084, 3.122	0.002, 0.336	0.454, 2.1 ▲	-0.07, 0.888 ▲	0.036, 0.144	0.525, 7.918 ▲	-0.061, 0.551 ▲	0.001, 0.082
Medium	Average	-0.148, 2.209 ▲	0.031, 0.303	-0.184, 14.968	0.04, 2.064	0.013, 0.127	0.269, 3.915 ▲	0.039, 0.339	-0.012, 0.065
Medium	Old	-0.082, 2.266	-0.016, 0.206 ▲	0.011, 0.589	0.035, 0.577 ▲	-0.015, 0.069 ▲	-0.141, 0.534	0.013, 0.08	-0.027, 0.037 ▲
High	Young	-0.015, 2.67	0.208, 1.515	0.618, 2.848 ▲	-0.024, 1.74 ▲	0.078, 0.541	0.08, 1.066 ▲	-0.104, 0.522	0.024, 0.345
High	Average	0.235, 3.827	0.164, 1.14	-0.029, 4.214	-0.258, 1.075 ▲	0.013, 0.452	-0.241, 0.58 ▲	0.203, 0.379 ▲	0.008, 0.233
High	Old	0.223, 1.445	-0.238, 0.497 ▲	-0.096, 0.547 ▲	-0.036, 0.377	0.081, 0.263 ▲	0.011, 0.384	-0.015, 0.137	-0.014, 0.076

Lognormal distributed data

Table 64:

Legend for colour-coded table and times a model ranked best for absolute mean bias of mean survival for lognormal distributed data. *Amount of times the model ranked best*

	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F	Total
Non-GPM	*1*	*1*	*1*	*1*	*3*	*3*	*0*	*10*
Internal additive	*3*	*5*	*2*	*1*	*0*	*3*	*1*	*15*
External additive	*2*	*0*	*2*	*20*	*1*	*4*	*3*	*32*
Converging	*1*	*3*	*4*	*3*	*2*	*2*	*0*	*15*
Total	*7*	*9*	*9*	*25*	*6*	*12*	*4*	

Table 65:

Colour-coded table for best ranking model for absolute mean bias of mean survival per scenario for lognormal distributed data. Survival and ages as main axes. *(Bias, RMSE),* ▲ *= Significant result (Wilcoxon)*

		Survival (Lognormal)							
		High		Medium			Low		
		Information		*Information*			*Information*		
Age	*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium	High
Young	Low	*-0.562, 2.06* ▲	*-0.462, 0.967* ▲	*0.107, 2.786*	*0.022, 0.86*	*0.025, 0.482*	*0.037, 0.372*	*-0.001, 0.123*	*-0.035, 0.038*
Young	Medium	*-8.809, 21.976*	*-0.546, 1.098* ▲	*0.158, 2.766*	*0.043, 0.879*	*0.036, 0.487*	*0.045, 0.496*	*-0.002, 0.193*	*-0.003, 0.054*
Young	High	*0.075, 2.113* ▲	*-0.413, 1.221*	*0.195, 2.584*	*0.298, 1.021*	*0.051, 0.503*	*0.188, 1.527*	*0.09, 0.601*	*0.023, 0.331*
Average	Low	*0.298, 3.329*	*0.571, 3.947* ▲	*0.082, 2.1*	*-0.063, 1.156*	*-0.119, 0.314* ▲	*-0.052, 0.337*	*0.014, 0.116*	*0.013, 0.037* ▲
Average	Medium	*0.149, 3.036*	*-0.808, 2.123*	*-0.011, 2.451*	*-0.114, 1.123* ▲	*0.015, 0.316* ▲	*-0.016, 0.423*	*0, 0.108*	*0.007, 0.04*
Average	High	*-0.37, 2.158*	*-0.214, 0.57* ▲	*0.072, 1.749*	*-0.015, 1.007* ▲	*0.137, 0.35* ▲	*0.051, 1.124*	*-0.016, 0.429*	*-0.018, 0.236*
Old	Low	*0.092, 0.957*	*0.027, 0.545* ▲	*0.162, 0.92*	*-0.062, 0.261* ▲	*0.023, 0.071* ▲	*-0.094, 0.267*	*-0.029, 0.05*	*0.001, 0.028*
Old	Medium	*0.076, 3.395*	*-0.055, 0.505*	*0.069, 2.525*	*-0.282, 0.357* ▲	*0.009, 0.078*	*0.002, 0.199* ▲	*0.028, 0.068* ▲	*0.022, 0.038*
Old	High	*0.081, 0.443*	*0.002, 0.409*	*0.01, 1.378*	*0, 0.354*	*-0.012, 0.135*	*-0.226, 0.508*	*0.082, 0.277* ▲	*-0.014, 0.075*

Table 66:

Colour-coded table for best ranking model for absolute mean bias of mean survival per scenario for lognormal distributed data. Information and heterogeneity as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

		Information: Low		Information: Medium			Information: High		
		Survival (Lognormal)		*Survival (Lognormal)*			*Survival (Lognormal)*		
Heterogeneity	Age	Medium	Low	High	Medium	Low	High	Medium	Low
Low	Young	*0.107, 2.786*	*0.037, 0.372*	▲ *-0.562, 2.06*	*0.022, 0.86*	*-0.001, 0.123*	▲ *-0.462, 0.967*	*0.025, 0.482*	*-0.035, 0.038*
Low	Average	*0.082, 2.1*	*-0.052, 0.337*	*0.298, 3.329*	*-0.063, 1.156*	*0.014, 0.116*	▲ *0.571, 3.947*	▲ *-0.119, 0.314*	▲ *0.013, 0.037*
Low	Old	*0.162, 0.92*	*-0.094, 0.267*	*0.092, 0.957*	▲ *-0.062, 0.261*	*-0.029, 0.05*	▲ *0.027, 0.545*	▲ *0.023, 0.071*	*0.001, 0.028*
Medium	Young	*0.158, 2.766*	*0.045, 0.496*	*-8.809, 21.976*	*0.043, 0.879*	*-0.002, 0.193*	▲ *-0.546, 1.098*	*0.036, 0.487*	*-0.003, 0.054*
Medium	Average	*-0.011, 2.451*	*-0.016, 0.423*	*0.149, 3.036*	▲ *-0.114, 1.123*	*0, 0.108*	*-0.808, 2.123*	▲ *0.015, 0.316*	*0.007, 0.04*
Medium	Old	*0.069, 2.525*	▲ *0.002, 0.199*	*0.076, 3.395*	▲ *-0.282, 0.357*	▲ *0.028, 0.068*	*-0.055, 0.505*	*0.009, 0.078*	*0.022, 0.038*
High	Young	*0.195, 2.584*	*0.188, 1.527*	▲ *0.075, 2.113*	*0.298, 1.021*	*0.09, 0.601*	*-0.413, 1.221*	*0.051, 0.503*	*0.023, 0.331*
High	Average	*0.072, 1.749*	*0.051, 1.124*	*-0.37, 2.158*	▲ *-0.015, 1.007*	*-0.016, 0.429*	▲ *-0.214, 0.57*	▲ *0.137, 0.35*	*-0.018, 0.236*
High	Old	*0.01, 1.378*	*-0.226, 0.508*	*0.081, 0.443*	*0, 0.354*	▲ *0.082, 0.277*	*0.002, 0.409*	*-0.012, 0.135*	*-0.014, 0.075*

Gompertz distributed data

Table 67:

Legend for colour-coded table and times a model ranked best for absolute mean bias of mean survival for Gompertz distributed data. *Amount of times the model ranked best*

	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F	Total
Non-GPM	*0*	*7*	*0*	*0*	*2*	*4*	*0*	*13*
Internal additive	*2*	*7*	*0*	*0*	*3*	*5*	*0*	*17*
External additive	*4*	*2*	*3*	*8*	*1*	*6*	*0*	*24*
Converging	*1*	*3*	*3*	*0*	*5*	*6*	*0*	*18*
Total	*7*	*19*	*6*	*8*	*11*	*21*	*0*	

Table 68:

Colour-coded table for best ranking model for absolute mean bias of mean survival per scenario for Gompertz distributed data. Survival and ages as main axes. *(Bias, RMSE),* ▲ *= Significant result (Wilcoxon)*

Survival (Gompertz)

Age		Survival:	High		Medium			Low		
		Information:	Medium	High	Low	Medium	High	Low	Medium	High
Young	*Heterogeneity*	Low	*0.318, 5.312* ▲	*0.319, 1.456*	*-0.525, 4.772*	*-0.088, 3.057*	*-0.184, 0.724*	*0.244, 0.296*	*-0.004, 0.088*	*-0.036, 0.036* ▲
		Medium	*-3.34, 14.042*	*0.016, 2.406*	*-1.481, 4.723*	*0.019, 3.086*	*0.021, 0.625*	*-0.06, 0.473*	*-0.056, 0.157*	*-0.017, 0.046* ▲
		High	*-0.415, 5.307*	*0.095, 1.043*	*-0.136, 3.006*	*-0.242, 1.903* ▲	*0.051, 0.629*	*0.463, 1.427* ▲	*0.042, 0.89*	*-0.052, 0.321* ▲
Average	*Heterogeneity*	Low	*-0.246, 1.139* ▲	*-0.382, 2.601*	*-0.087, 3.94*	*-0.127, 2.153* ▲	*-0.028, 0.264* ▲	*0.233, 0.282*	*0.001, 0.096*	*-0.023, 0.023* ▲
		Medium	*-0.093, 1.246* ▲	*-0.37, 0.738* ▲	*-0.73, 4.001*	*0.062, 2.451*	*-0.015, 0.364*	*-0.016, 0.488*	*-0.04, 0.153*	*-0.001, 0.034*
		High	*0.107, 1.426* ▲	*-0.44, 0.622* ▲	*0.296, 4.17*	*0.286, 0.795*	*0.097, 0.357* ▲	*0.272, 1.041* ▲	*-0.107, 0.603*	*-0.089, 0.229*
Old	*Heterogeneity*	Low	*0.167, 0.888*	*0.006, 0.325*	*0.034, 1.68* ▲	*-0.01, 0.236*	*0.026, 0.089*	*0.029, 0.306* ▲	*-0.012, 0.073*	*0.041, 0.046* ▲
		Medium	*0.092, 0.543*	*-0.156, 0.373*	*0.031, 1.836*	*-0.02, 0.761*	*0.001, 0.089*	*-0.018, 0.378*	*0.005, 0.163*	*-0.028, 0.044*
		High	*0.018, 0.412* ▲	*0.05, 0.338*	*0.19, 1.755*	*0.094, 0.436*	*-0.012, 0.137*	*0.022, 0.47*	*-0.146, 0.259*	*-0.006, 0.07*

Table 69:

Colour-coded table for best ranking model for absolute mean bias of mean survival per scenario for Gompertz distributed data. Information and heterogeneity as main axes. *(Bias, RMSE),* ▲ *= Significant result (Wilcoxon)*

		Information							
		Low		Medium			High		
		Survival (Gompertz)		*Survival (Gompertz)*			*Survival (Gompertz)*		
Heterogeneity	**Age**	Medium	Low	High	Medium	Low	High	Medium	Low
Low	Young	*-0.525, 4.772*	*0.244, 0.296*	*0.318, 5.312* ▲	*-0.088, 3.057*	*-0.004, 0.088*	*0.319, 1.456*	*-0.184, 0.724*	*-0.036, 0.036* ▲
Low	Average	*-0.087, 3.94*	*0.233, 0.282*	*-0.246, 1.139* ▲	*-0.127, 2.153* ▲	*0.001, 0.096*	*-0.382, 2.601*	*-0.028, 0.264* ▲	*-0.023, 0.023* ▲
Low	Old	*0.034, 1.68* ▲	*0.029, 0.306* ▲	*0.167, 0.888*	*-0.01, 0.236*	*-0.012, 0.073*	*0.006, 0.325*	*0.026, 0.089*	*0.041, 0.046* ▲
Medium	Young	*-1.481, 4.723*	*-0.06, 0.473*	*-3.34, 14.042*	*0.019, 3.086*	*-0.056, 0.157*	*0.016, 2.406*	*0.021, 0.625*	*-0.017, 0.046* ▲
Medium	Average	*-0.73, 4.001*	*-0.016, 0.488*	*-0.093, 1.246* ▲	*0.062, 2.451*	*-0.04, 0.153*	*-0.37, 0.738* ▲	*-0.015, 0.364*	*-0.001, 0.034*
Medium	Old	*0.031, 1.836*	*-0.018, 0.378*	*0.092, 0.543*	*-0.02, 0.761*	*0.005, 0.163*	*-0.156, 0.373*	*0.001, 0.089*	*-0.028, 0.044*
High	Young	*-0.136, 3.006*	*0.463, 1.427* ▲	*-0.415, 5.307*	*-0.242, 1.903* ▲	*0.042, 0.89*	*0.095, 1.043*	*0.051, 0.629*	*-0.052, 0.321* ▲
High	Average	*0.296, 4.17*	*0.272, 1.041* ▲	*0.107, 1.426* ▲	*0.286, 0.795*	*-0.107, 0.603*	*-0.44, 0.622* ▲	*0.097, 0.357* ▲	*-0.089, 0.229*
High	Old	*0.19, 1.755*	*0.022, 0.47*	*0.018, 0.412* ▲	*0.094, 0.436*	*-0.146, 0.259*	*0.05, 0.338*	*-0.012, 0.137*	*-0.006, 0.07*

B.2: Survival probability at time *t*

Weibull distributed data

Table 70:

Legend for colour-coded table and times a model ranked best for absolute mean bias of survival probability at time t for Weibull distributed data. *Amount of times the model ranked best*

	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F	Total
Non-GPM	2	1	1	0	2	10	0	16
Internal additive	2	4	3	1	0	9	0	19
External additive	4	2	4	7	1	1	1	20
Converging	1	6	1	3	1	5	0	17
Total	9	13	9	11	4	25	1	

Table 71:

Colour-coded table for best ranking model for absolute mean bias of survival probability at time t per scenario for Weibull distributed data. Survival and ages as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

		Survival (Weibull)							
		High		Medium			Low		
		Information		Information			Information		
Age	Heterogeneity	Medium	High	Low	Medium	High	Low	Medium*	High*
Young	Low	0.006, 0.042 ▲	-0.004, 0.018 ▲	0.001, 0.115	-0.003, 0.092	0, 0.015	-0.006, 0.083	-0.006, 0.008	0, 0.009
Young	Medium	0.008, 0.044 ▲	0.001, 0.02 ▲	-0.01, 0.116	-0.006, 0.037	0, 0.014	-0.003, 0.136	-0.006, 0.008	0, 0.009
Young	High	-0.005, 0.065	-0.005, 0.018 ▲	-0.006, 0.066	0.006, 0.043	-0.001, 0.011	-0.001, 0.064	0, 0.008	0, 0.009
Average	Low	0.01, 0.066 ▲	-0.025, 0.068	-0.002, 0.103 ▲	-0.007, 0.092	0.004, 0.015	-0.001, 0.08	-0.004, 0.007	0.001, 0.009
Average	Medium	-0.017, 0.045 ▲	-0.022, 0.066	0.002, 0.111	0.001, 0.036	0.002, 0.012	-0.005, 0.134	-0.004, 0.007	0.002, 0.009
Average	High	0.016, 0.16	-0.019, 0.025	0.001, 0.059 ▲	0.007, 0.024 ▲	-0.002, 0.011	-0.002, 0.067	0, 0.008	0.002, 0.009
Old	Low	0.002, 0.027	-0.001, 0.049	-0.008, 0.032	-0.012, 0.016	0.001, 0.008	0.004, 0.075	-0.006, 0.009	-0.007, 0.01
Old	Medium	-0.004, 0.017	-0.004, 0.049 ▲	0.006, 0.056	-0.002, 0.016 ▲	0.008, 0.012	-0.001, 0.126	-0.005, 0.011	0.001, 0.009
Old	High	-0.023, 0.028	0.025, 0.04	0, 0.03	0.011, 0.02	0.004, 0.011	-0.005, 0.039	0, 0.006	0.006, 0.012

Table 72:

Colour-coded table for best ranking model for absolute mean bias of survival probability at time t per scenario for Weibull distributed data. Information and heterogeneity as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

		Information: Low		Information: Medium			Information: High		
		Survival (Weibull)		*Survival (Weibull)*			*Survival (Weibull)*		
Heterogeneity	**Age**	Medium	Low	High	Medium	Low*	High	Medium	Low*
Low	Young	*0.001, 0.115*	*-0.006, 0.083*	*0.006, 0.042* ▲	*-0.003, 0.092*	*-0.006, 0.008*	*-0.004, 0.018* ▲	*0, 0.015*	*0, 0.009*
Low	Average	*-0.002, 0.103* ▲	*-0.001, 0.08*	*0.01, 0.066* ▲	*-0.007, 0.092*	*-0.004, 0.007*	*-0.025, 0.068*	*0.004, 0.015*	*0.001, 0.009*
Low	Old	*-0.008, 0.032*	*0.004, 0.075*	*0.002, 0.027*	*-0.012, 0.016*	*-0.006, 0.009*	*-0.001, 0.049*	*0.001, 0.008*	*-0.007, 0.01*
Medium	Young	*-0.01, 0.116*	*-0.003, 0.136*	*0.008, 0.044* ▲	*-0.006, 0.037*	*-0.006, 0.008*	*0.001, 0.02* ▲	*0, 0.014*	*0, 0.009*
Medium	Average	*0.002, 0.111*	*-0.005, 0.134*	*-0.017, 0.045* ▲	*0.001, 0.036*	*-0.004, 0.007*	*-0.022, 0.066*	*0.002, 0.012*	*0.002, 0.009*
Medium	Old	*0.006, 0.056*	*-0.001, 0.126*	*-0.004, 0.017*	*-0.002, 0.016* ▲	*-0.005, 0.011*	*-0.004, 0.049* ▲	*0.008, 0.012*	*0.001, 0.009*
High	Young	*-0.006, 0.066*	*-0.001, 0.064*	*-0.005, 0.065*	*0.006, 0.043*	*0, 0.008*	*-0.005, 0.018* ▲	*-0.001, 0.011*	*0, 0.009*
High	Average	*0.001, 0.059* ▲	*-0.002, 0.067*	*0.016, 0.16*	*0.007, 0.024* ▲	*0, 0.008*	*-0.019, 0.025*	*-0.002, 0.011*	*0.002, 0.009*
High	Old	*0, 0.03*	*-0.005, 0.039*	*-0.023, 0.028*	*0.011, 0.02*	*0, 0.006*	*0.025, 0.04*	*0.004, 0.011*	*0.006, 0.012*

Log-logistic distributed data

Table 73:

Legend for colour-coded table and times a model ranked best for absolute mean bias of survival probability at time t for log-logistic distributed data. *Amount of model the method ranked best*

	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F	Total
Non-GPM	*1*	*1*	*4*	*0*	*1*	*4*	*0*	*11*
Internal additive	*1*	*0*	*10*	*2*	*0*	*2*	*0*	*15*
External additive	*2*	*0*	*8*	*15*	*1*	*3*	*0*	*29*
Converging	*0*	*6*	*5*	*2*	*1*	*2*	*1*	*17*
Total	*4*	*7*	*27*	*19*	*3*	*11*	*1*	

Table 74:

Colour-coded table for best ranking model for absolute mean bias of survival probability at time t per scenario for log-logistic distributed data. Survival and ages as main axes. *(Bias, RMSE),* ▲ *= Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

		Survival (Log-logistic)							
		High		Medium			Low		
		Information		*Information*			*Information*		
Age	*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium*	High*
Young	Low	*0.008, 0.08*	*0.007, 0.027*	*0.005, 0.09*	*-0.001, 0.027*	*0.004, 0.011*	*0, 0.057*	*-0.004, 0.01*	*-0.004, 0.009*
Young	Medium	▲ *0.005, 0.044*	*0.005, 0.027*	*0.001, 0.089*	*0.003, 0.027*	*-0.001, 0.012*	*-0.014, 0.125*	*0.001, 0.009*	*0.003, 0.009*
Young	High	*-0.004, 0.065*	*-0.005, 0.02*	*0.002, 0.137*	*0.001, 0.025*	*0.001, 0.01*	*-0.001, 0.041*	*0, 0.008*	*0, 0.009*
Average	Low	▲ *0.009, 0.064*	*0.03, 0.048*	*0.004, 0.083*	▲ *-0.003, 0.025*	▲ *0.001, 0.011*	*0, 0.055*	*-0.002, 0.01*	*-0.002, 0.008*
Average	Medium	▲ *0.013, 0.066*	*-0.026, 0.067*	*-0.006, 0.076*	▲ *-0.006, 0.025*	*0, 0.011*	*-0.018, 0.056*	*0, 0.01*	*0, 0.009*
Average	High	▲ *0.008, 0.045*	*-0.016, 0.023*	*0.001, 0.066*	*-0.006, 0.036*	*-0.007, 0.013*	*-0.003, 0.042*	*-0.001, 0.006*	*0.001, 0.008*
Old	Low	*0.002, 0.028*	*-0.001, 0.049*	▲ *-0.001, 0.028*	*0, 0.015*	*0.007, 0.01*	▲ *0.005, 0.051*	*-0.002, 0.011*	*0, 0.008*
Old	Medium	*-0.006, 0.017*	▲ *-0.001, 0.049*	*-0.003, 0.055*	▲ *-0.006, 0.013*	*0.002, 0.008*	*0.005, 0.051*	*0, 0.01*	*0.001, 0.008*
Old	High	*-0.028, 0.031*	*0.025, 0.042*	*-0.007, 0.055*	▲ *0.002, 0.018*	*0.002, 0.01*	*-0.001, 0.04*	*-0.001, 0.009*	*0.007, 0.013*

Table 75:

Colour-coded table for best ranking model for absolute mean bias of survival probability at time t per scenario for log-logistic distributed data. Information and heterogeneity as main axes. *(Bias, RMSE),* ▲ *= Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

		Information: Low		Information: Medium			Information: High		
		Survival (Log-logistic)		*Survival (Log-logistic)*			*Survival (Log-logistic)*		
Heterogeneity	**Age**	Medium	Low	High	Medium	Low*	High	Medium	Low*
Low	Young	*0.005, 0.09*	*0, 0.057*	*0.008, 0.08*	*-0.001, 0.027*	*-0.004, 0.01*	*0.007, 0.027*	*0.004, 0.011*	*-0.004, 0.009*
Low	Average	*0.004, 0.083*	*0, 0.055*	*0.009, 0.064* ▲	*-0.003, 0.025* ▲	*-0.002, 0.01*	*0.03, 0.048*	*0.001, 0.011* ▲	*-0.002, 0.008*
Low	Old	*-0.001, 0.028* ▲	*0.005, 0.051* ▲	*0.002, 0.028*	*0, 0.015*	*-0.002, 0.011*	*-0.001, 0.049*	*0.007, 0.01*	*0, 0.008*
Medium	Young	*0.001, 0.089*	*-0.014, 0.125*	*0.005, 0.044* ▲	*0.003, 0.027*	*0.001, 0.009*	*0.005, 0.027*	*-0.001, 0.012*	*0.003, 0.009*
Medium	Average	*-0.006, 0.076*	*-0.018, 0.056*	*0.013, 0.066* ▲	*-0.006, 0.025* ▲	*0, 0.01*	*-0.026, 0.067*	*0, 0.011*	*0, 0.009*
Medium	Old	*-0.003, 0.055*	*0.005, 0.051*	*-0.006, 0.017*	*-0.006, 0.013* ▲	*0, 0.01*	*-0.001, 0.049* ▲	*0.002, 0.008*	*0.001, 0.008*
High	Young	*0.002, 0.137*	*-0.001, 0.041*	*-0.004, 0.065*	*0.001, 0.025*	*0, 0.008*	*-0.005, 0.02*	*0.001, 0.01*	*0, 0.009*
High	Average	*0.001, 0.066*	*-0.003, 0.042*	*0.008, 0.045* ▲	*-0.006, 0.036*	*-0.001, 0.006*	*-0.016, 0.023*	*-0.007, 0.013*	*0.001, 0.008*
High	Old	*-0.007, 0.055*	*-0.001, 0.04*	*-0.028, 0.031*	*0.002, 0.018* ▲	*-0.001, 0.009*	*0.025, 0.042*	*0.002, 0.01*	*0.007, 0.013*

Lognormal distributed data

Table 76:

Legend for colour-coded table and times a model ranked best for absolute mean bias of survival probability at time t for lognormal distributed data. *Amount of times the model ranked best*

	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F	Total
Non-GPM	*2*	*1*	*5*	*3*	*2*	*4*	*0*	*17*
Internal additive	*0*	*1*	*4*	*7*	*0*	*3*	*4*	*19*
External additive	*2*	*0*	*0*	*12*	*4*	*4*	*0*	*22*
Converging	*0*	*5*	*3*	*2*	*1*	*1*	*2*	*14*
Total	*4*	*7*	*12*	*24*	*7*	*12*	*6*	

Table 77:

Colour-coded table for best ranking model for absolute mean bias of survival probability at time t per scenario for lognormal distributed data. Survival and ages as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

		Survival (Lognormal)							
		High		Medium			Low		
		Information		*Information*			*Information*		
Age	*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium*	High*
Young	Low	*0.008, 0.125*	*-0.002, 0.023*	*-0.003, 0.078*	*0.002, 0.046*	*0.001, 0.011*	*-0.002, 0.059*	*-0.002, 0.006*	*-0.004, 0.009*
Young	Medium	*0.272, 0.272*	*0.001, 0.022*	*-0.001, 0.077*	*0.001, 0.046*	*0.002, 0.01*	*-0.008, 0.059*	*-0.002, 0.006*	*-0.004, 0.009*
Young	High	▲ *-0.003, 0.044*	*0.001, 0.019*	*-0.001, 0.064*	*0.003, 0.024*	*-0.002, 0.01*	*-0.001, 0.061*	*-0.001, 0.006*	*-0.001, 0.009*
Average	Low	*-0.004, 0.132*	*0.016, 0.032*	*0.002, 0.073*	*0.003, 0.024*	*0.004, 0.012*	*0.001, 0.06*	*0, 0.006*	*-0.002, 0.008*
Average	Medium	*0, 0.12*	*0.018, 0.032*	*0.002, 0.071*	*-0.003, 0.024*	*0.006, 0.013*	*-0.001, 0.058*	*0.001, 0.006*	*-0.001, 0.009*
Average	High	▲ *0, 0.044*	*-0.012, 0.021*	*0.004, 0.058*	*0.001, 0.034*	*-0.009, 0.014*	*0, 0.043*	*-0.001, 0.006*	*0.001, 0.008*
Old	Low	*-0.002, 0.028*	▲ *0.005, 0.027*	*-0.002, 0.027*	▲ *-0.002, 0.026*	*0.005, 0.009*	*-0.004, 0.117*	*-0.003, 0.01*	*0, 0.009*
Old	Medium	*-0.007, 0.018*	*-0.004, 0.028*	*0.005, 0.05*	*0.001, 0.025*	*0.001, 0.008*	*-0.002, 0.109*	*-0.005, 0.011*	*-0.002, 0.009*
Old	High	*0.029, 0.035*	*0.025, 0.045*	*0.01, 0.06*	▲ *-0.001, 0.018*	*0.002, 0.01*	*0, 0.04*	*0, 0.006*	*0.007, 0.013*

Table 78:

Colour-coded table for best ranking model for absolute mean bias of survival probability at time t per scenario for lognormal distributed data. Information and heterogeneity as main axes. *(Bias, RMSE),* ▲ *= Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

		Information: Low		Information: Medium			Information: High		
		Survival (Lognormal)		*Survival (Lognormal)*			*Survival (Lognormal)*		
Heterogeneity	Age	Medium	Low	High	Medium	Low*	High	Medium	Low*
Low	Young	*-0.003, 0.078*	*-0.002, 0.059*	*0.008, 0.125*	*0.002, 0.046*	*-0.002, 0.006*	*-0.002, 0.023*	*0.001, 0.011*	*-0.004, 0.009*
Low	Average	*0.002, 0.073*	*0.001, 0.06*	*-0.004, 0.132*	*0.003, 0.024*	*0, 0.006*	*0.016, 0.032*	*0.004, 0.012*	*-0.002, 0.008*
Low	Old	*-0.002, 0.027*	*-0.004, 0.117*	*-0.002, 0.028*	▲ *-0.002, 0.026*	*-0.003, 0.01*	▲ *0.005, 0.027*	*0.005, 0.009*	*0, 0.009*
Medium	Young	*-0.001, 0.077*	*-0.008, 0.059*	*0.272, 0.272*	*0.001, 0.046*	*-0.002, 0.006*	*0.001, 0.022*	*0.002, 0.01*	*-0.004, 0.009*
Medium	Average	*0.002, 0.071*	*-0.001, 0.058*	*0, 0.12*	*-0.003, 0.024*	*0.001, 0.006*	*0.018, 0.032*	*0.006, 0.013*	*-0.001, 0.009*
Medium	Old	*0.005, 0.05*	*-0.002, 0.109*	*-0.007, 0.018*	*0.001, 0.025*	*-0.005, 0.011*	*-0.004, 0.028*	*0.001, 0.008*	*-0.002, 0.009*
High	Young	*-0.001, 0.064*	*-0.001, 0.061*	▲ *-0.003, 0.044*	*0.003, 0.024*	*-0.001, 0.006*	*0.001, 0.019*	*-0.002, 0.01*	*-0.001, 0.009*
High	Average	*0.004, 0.058*	*0, 0.043*	▲ *0, 0.044*	*0.001, 0.034*	*-0.001, 0.006*	*-0.012, 0.021*	*-0.009, 0.014*	*0.001, 0.008*
High	Old	*0.01, 0.06*	*0, 0.04*	*0.029, 0.035*	▲ *-0.001, 0.018*	*0, 0.006*	*0.025, 0.045*	*0.002, 0.01*	*0.007, 0.013*

Gompertz distributed data

Table 79:

Legend for colour-coded table and times a model ranked best for absolute mean bias of survival probability at time t for Gompertz distributed data. *Amount of times the model ranked best*

	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F	Total
Non-GPM	*0*	*3*	*1*	*0*	*7*	*3*	*0*	*14*
Internal additive	*0*	*7*	*4*	*0*	*5*	*4*	*0*	*20*
External additive	*2*	*1*	*5*	*6*	*6*	*2*	*0*	*22*
Converging	*0*	*5*	*2*	*3*	*1*	*4*	*1*	*16*
Total	*2*	*16*	*12*	*9*	*19*	*13*	*1*	

Table 80:

Colour-coded table for best ranking model for absolute mean bias of survival probability at time t per scenario for Gompertz distributed data. Survival and ages as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

Survival (Gompertz)

Age	Heterogeneity	High: *Information* Medium	High: *Information* High	Medium: *Information* Low	Medium: *Information* Medium	Medium: *Information* High	Low: *Information* Low	Low: *Information* Medium*	Low: *Information* High*
Young	Low	*-0.001, 0.124*	*0.005, 0.018*	*-0.031, 0.135*	*0.009, 0.117*	*-0.001, 0.016*	*0.011, 0.144*	*-0.001, 0.008*	*-0.001, 0.009*
Young	Medium	*-0.167, 0.369* ▲	*0, 0.018* ▲	*0.054, 0.114*	*-0.025, 0.084*	*0.001, 0.015*	*-0.004, 0.139*	*-0.004, 0.007*	*0, 0.009*
Young	High	*0.004, 0.137*	*-0.003, 0.018*	*0.018, 0.165*	*-0.002, 0.025*	*0, 0.011*	*0, 0.062*	*0, 0.008*	*0, 0.008*
Average	Low	*0.016, 0.063* ▲	*-0.014, 0.068*	*-0.012, 0.13*	*0.003, 0.101*	*0.002, 0.015*	*0.005, 0.14*	*0.001, 0.008*	*0.002, 0.009*
Average	Medium	*-0.005, 0.043* ▲	*-0.014, 0.067*	*-0.044, 0.13*	*-0.008, 0.083*	*-0.003, 0.013*	*-0.001, 0.138*	*-0.002, 0.006*	*0, 0.008*
Average	High	*0.003, 0.168*	*0.019, 0.027*	*0, 0.068*	*-0.012, 0.041*	*-0.007, 0.013* ▲	*-0.002, 0.04*	*0, 0.007*	*0.001, 0.009*
Old	Low	*0.002, 0.028*	*-0.006, 0.049*	*0.017, 0.035*	*0.005, 0.013* ▲	*0.008, 0.012*	*-0.002, 0.11*	*-0.001, 0.008*	*0.005, 0.011*
Old	Medium	*-0.004, 0.017*	*-0.003, 0.03*	*0.006, 0.031*	*0, 0.022*	*0.004, 0.009*	*0.001, 0.128*	*0.001, 0.01* ▲	*-0.001, 0.009*
Old	High	*-0.018, 0.024* ▲	*0.026, 0.038*	*-0.011, 0.06*	*-0.001, 0.018* ▲	*0.003, 0.009* ▲	*0.005, 0.039*	*0.003, 0.007*	*0.004, 0.012*

Table 81:

Colour-coded table for best ranking model for absolute mean bias of survival probability at time t per scenario for Gompertz distributed data. Information and heterogeneity as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon), * denotes scenarios that are not extrapolations*

		Information							
		Low		Medium			High		
		Survival (Gompertz)		*Survival (Gompertz)*			*Survival (Gompertz)*		
Heterogeneity	**Age**	Medium	Low	High	Medium	Low*	High	Medium	Low*
Low	Young	*-0.031, 0.135*	*0.011, 0.144*	*-0.001, 0.124*	*0.009, 0.117*	*-0.001, 0.008*	*0.005, 0.018*	*-0.001, 0.016*	*-0.001, 0.009*
Low	Average	*-0.012, 0.13*	*0.005, 0.14*	*0.016, 0.063* ▲	*0.003, 0.101*	*0.001, 0.008*	*-0.014, 0.068*	*0.002, 0.015*	*0.002, 0.009*
Low	Old	*0.017, 0.035*	*-0.002, 0.11*	*0.002, 0.028*	*0.005, 0.013* ▲	*-0.001, 0.008*	*-0.006, 0.049*	*0.008, 0.012*	*0.005, 0.011*
Medium	Young	*0.054, 0.114*	*-0.004, 0.139*	*-0.167, 0.369* ▲	*-0.025, 0.084*	*-0.004, 0.007*	*0, 0.018* ▲	*0.001, 0.015*	*0, 0.009*
Medium	Average	*-0.044, 0.13*	*-0.001, 0.138*	*-0.005, 0.043* ▲	*-0.008, 0.083*	*-0.002, 0.006*	*-0.014, 0.067*	*-0.003, 0.013*	*0, 0.008*
Medium	Old	*0.006, 0.031*	*0.001, 0.128*	*-0.004, 0.017*	*0, 0.022*	*0.001, 0.01* ▲	*-0.003, 0.03*	*0.004, 0.009*	*-0.001, 0.009*
High	Young	*0.018, 0.165*	*0, 0.062*	*0.004, 0.137*	*-0.002, 0.025*	*0, 0.008*	*-0.003, 0.018*	*0, 0.011*	*0, 0.008*
High	Average	*0, 0.068*	*-0.002, 0.04*	*0.003, 0.168*	*-0.012, 0.041*	*0, 0.007*	*0.019, 0.027*	*-0.007, 0.013* ▲	*0.001, 0.009*
High	Old	*-0.011, 0.06*	*0.005, 0.039*	*-0.018, 0.024* ▲	*-0.001, 0.018* ▲	*0.003, 0.007*	*0.026, 0.038*	*0.003, 0.009* ▲	*0.004, 0.012*

B.3: RMST

Weibull distributed data

Table 82:

Legend for colour-coded table and times a model ranked best for absolute mean bias of RMST for Weibull distributed data. *Amount of times the model ranked best*

	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F	Total
Non-GPM	*2*	*3*	*6*	*0*	*6*	*1*	*0*	***18***
Internal additive	*3*	*9*	*7*	*1*	*6*	*1*	*0*	***27***
External additive	*5*	*2*	*4*	*0*	*0*	*1*	*1*	***13***
Converging	*3*	*2*	*1*	*0*	*2*	*5*	*1*	***14***
Total	***13***	***16***	***18***	***1***	***14***	***8***	***2***	

Table 83:

Colour-coded table for best ranking model for absolute mean bias of RMST per scenario for Weibull distributed data. Survival and ages as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

		Survival (Weibull)							
		High		Medium			Low		
		Information		*Information*			*Information*		
Age	*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium	High
Young	Low	*0, 0.004*	*-0.006, 0.012*	*0.001, 0.009*	*0, 0.015*	*-0.007, 0.029*	*-0.015, 0.015*	*-0.032, 0.032* ▲	*-0.045, 0.045*
Young	Medium	*0, 0.004*	*-0.006, 0.013*	*0.001, 0.009*	*0.002, 0.016*	*0.004, 0.029*	*-0.016, 0.016*	*-0.024, 0.025* ▲	*-0.029, 0.029* ▲
Young	High	*-0.001, 0.008*	*0.004, 0.027*	*0.001, 0.01*	*-0.002, 0.017*	*-0.004, 0.044*	*0.001, 0.011*	*-0.005, 0.015* ▲	*0.001, 0.032*
Average	Low	*-0.001, 0.005*	*-0.008, 0.015*	*-0.001, 0.009*	*0.003, 0.013*	*-0.013, 0.03*	*-0.013, 0.014*	*-0.027, 0.027* ▲	*-0.032, 0.032* ▲
Average	Medium	*-0.002, 0.005*	*-0.01, 0.016*	*0, 0.01*	*-0.003, 0.014*	*-0.006, 0.038*	*-0.015, 0.015*	*-0.02, 0.021* ▲	*-0.014, 0.017*
Average	High	*-0.001, 0.008*	*0.004, 0.032*	*0, 0.01*	*0, 0.016*	*0.002, 0.044*	*-0.001, 0.011*	*-0.001, 0.014* ▲	*-0.03, 0.042*
Old	Low	*-0.008, 0.009* ▲	*-0.026, 0.031*	*0, 0.01*	*-0.016, 0.018*	*0.015, 0.04*	*-0.003, 0.007*	*0.014, 0.018* ▲	*0.033, 0.035* ▲
Old	Medium	*-0.008, 0.009*	*-0.027, 0.031*	*0, 0.01*	*-0.016, 0.018*	*-0.024, 0.031*	*0.002, 0.013*	*0.005, 0.019* ▲	*-0.045, 0.045*
Old	High	*-0.002, 0.019*	*-0.025, 0.057*	*0.001, 0.011*	*0, 0.018*	*0, 0.033*	*-0.001, 0.01*	*0.01, 0.021* ▲	*0, 0.033*

Table 84:

Colour-coded table for best ranking model for absolute mean bias of RMST per scenario for Weibull distributed data. Information and heterogeneity as main axes. *(Bias, RMSE),* ▲ *= Significant result (Wilcoxon)*

Heterogeneity	Age	Information: Low Survival (Weibull) Medium	Low Survival (Weibull) Low	Medium Survival (Weibull) High	Medium Survival (Weibull) Medium	Medium Survival (Weibull) Low	High Survival (Weibull) High	High Survival (Weibull) Medium	High Survival (Weibull) Low
Low	Young	*0.001, 0.009*	*-0.015, 0.015*	*0, 0.004*	*0, 0.015*	▲ *-0.032, 0.032*	*-0.006, 0.012*	*-0.007, 0.029*	*-0.045, 0.045*
Low	Average	*-0.001, 0.009*	*-0.013, 0.014*	*-0.001, 0.005*	*0.003, 0.013*	▲ *-0.027, 0.027*	*-0.008, 0.015*	*-0.013, 0.03*	▲ *-0.032, 0.032*
Low	Old	*0, 0.01*	*-0.003, 0.007*	▲ *-0.008, 0.009*	*-0.016, 0.018*	▲ *0.014, 0.018*	*-0.026, 0.031*	*0.015, 0.04*	▲ *0.033, 0.035*
Medium	Young	*0.001, 0.009*	*-0.016, 0.016*	*0, 0.004*	*0.002, 0.016*	▲ *-0.024, 0.025*	*-0.006, 0.013*	*0.004, 0.029*	▲ *-0.029, 0.029*
Medium	Average	*0, 0.01*	*-0.015, 0.015*	*-0.002, 0.005*	*-0.003, 0.014*	▲ *-0.02, 0.021*	*-0.01, 0.016*	*-0.006, 0.038*	*-0.014, 0.017*
Medium	Old	*0, 0.01*	*0.002, 0.013*	*-0.008, 0.009*	*-0.016, 0.018*	▲ *0.005, 0.019*	*-0.027, 0.031*	*-0.024, 0.031*	*-0.045, 0.045*
High	Young	*0.001, 0.01*	*0.001, 0.011*	*-0.001, 0.008*	*-0.002, 0.017*	▲ *-0.005, 0.015*	*0.004, 0.027*	*-0.004, 0.044*	*0.001, 0.032*
High	Average	*0, 0.01*	*-0.001, 0.011*	*-0.001, 0.008*	*0, 0.016*	▲ *-0.001, 0.014*	*0.004, 0.032*	*0.002, 0.044*	*-0.03, 0.042*
High	Old	*0.001, 0.011*	*-0.001, 0.01*	*-0.002, 0.019*	*0, 0.018*	▲ *0.01, 0.021*	*-0.025, 0.057*	*0, 0.033*	*0, 0.033*

Log-logistic distributed data

Table 85:

Legend for colour-coded table and times a model ranked best for absolute mean bias of RMST for log-logistic distributed data. *Amount of times the model ranked best*

	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F	Total
Non-GPM	*1*	*5*	*8*	*0*	*5*	*1*	*1*	***21***
Internal additive	*3*	*15*	*6*	*6*	*2*	*1*	*0*	***33***
External additive	*3*	*0*	*0*	*0*	*0*	*0*	*0*	***3***
Converging	*3*	*5*	*4*	*0*	*2*	*1*	*0*	***15***
Total	***10***	***25***	***18***	***6***	***9***	***3***	***1***	

Table 86:

Colour-coded table for best ranking model for absolute mean bias of RMST per scenario for log-logistic distributed data. Survival and ages as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

Survival (Log-logistic)

Age	Heterogeneity	High: Information Medium	High: Information High	Medium: Information Low	Medium: Information Medium	Medium: Information High	Low: Information Low	Low: Information Medium	Low: Information High
Young	Low	*0, 0.004*	*-0.005, 0.012*	*-0.006, 0.008*	*0, 0.014*	*-0.001, 0.035*	▲ *-0.014, 0.014*	▲ *-0.014, 0.016*	▲ *-0.015, 0.018*
Young	Medium	*-0.001, 0.004*	*-0.005, 0.02*	*-0.005, 0.011*	*-0.002, 0.014*	*0.001, 0.034*	*-0.014, 0.014*	▲ *-0.007, 0.013*	*-0.004, 0.016*
Young	High	*0, 0.008*	*0.007, 0.028*	*0, 0.009*	*0, 0.018*	*-0.004, 0.037*	*0, 0.01*	▲ *-0.01, 0.016*	*-0.006, 0.032*
Average	Low	*-0.001, 0.005*	*-0.008, 0.015*	*-0.005, 0.007*	▲ *-0.003, 0.02*	*-0.025, 0.037*	▲ *-0.014, 0.014*	▲ *-0.009, 0.013*	*-0.004, 0.014*
Average	Medium	*-0.001, 0.005*	*-0.009, 0.016*	*-0.004, 0.01*	*-0.003, 0.013*	*-0.028, 0.039*	▲ *-0.013, 0.014*	▲ *-0.003, 0.012*	▲ *0.002, 0.016*
Average	High	*0, 0.009*	*0.003, 0.031*	*0, 0.01*	*0.002, 0.018*	*0, 0.033*	*0, 0.01*	▲ *-0.005, 0.015*	▲ *0.041, 0.051*
Old	Low	▲ *-0.008, 0.009*	*-0.027, 0.031*	▲ *0.002, 0.007*	*0.002, 0.02*	*-0.024, 0.035*	*-0.003, 0.006*	▲ *0.016, 0.018*	*0, 0.026*
Old	Medium	*-0.008, 0.01*	*-0.026, 0.03*	*0.001, 0.008*	*0.01, 0.022*	*-0.018, 0.033*	*-0.001, 0.011*	*-0.002, 0.018*	*0.001, 0.027*
Old	High	*-0.002, 0.019*	*-0.027, 0.058*	*0, 0.01*	*-0.006, 0.023*	*0.002, 0.036*	*-0.003, 0.01*	*0.007, 0.019*	*-0.02, 0.036*

Table 87:

Colour-coded table for best ranking model for absolute mean bias of RMST per scenario for log-logistic distributed data. Information and heterogeneity as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

Heterogeneity		Information	Low		Medium			High		
			Survival (Log-logistic)		*Survival (Log-logistic)*			*Survival (Log-logistic)*		
			Medium	Low	High	Medium	Low	High	Medium	Low
Low	Age	Young	*-0.006, 0.008*	▲ *-0.014, 0.014*	*0, 0.004*	*0, 0.014*	▲ *-0.014, 0.016*	*-0.005, 0.012*	*-0.001, 0.035*	▲ *-0.015, 0.018*
		Average	*-0.005, 0.007*	▲ *-0.014, 0.014*	*-0.001, 0.005*	▲ *-0.003, 0.02*	▲ *-0.009, 0.013*	*-0.008, 0.015*	*-0.025, 0.037*	*-0.004, 0.014*
		Old	▲ *0.002, 0.007*	*-0.003, 0.006*	▲ *-0.008, 0.009*	*0.002, 0.02*	▲ *0.016, 0.018*	*-0.027, 0.031*	*-0.024, 0.035*	*0, 0.026*
Medium	Age	Young	*-0.005, 0.011*	*-0.014, 0.014*	*-0.001, 0.004*	*-0.002, 0.014*	▲ *-0.007, 0.013*	*-0.005, 0.02*	*0.001, 0.034*	*-0.004, 0.016*
		Average	*-0.004, 0.01*	▲ *-0.013, 0.014*	*-0.001, 0.005*	*-0.003, 0.013*	▲ *-0.003, 0.012*	*-0.009, 0.016*	*-0.028, 0.039*	▲ *0.002, 0.016*
		Old	*0.001, 0.008*	*-0.001, 0.011*	*-0.008, 0.01*	*0.01, 0.022*	*-0.002, 0.018*	*-0.026, 0.03*	*-0.018, 0.033*	*0.001, 0.027*
High	Age	Young	*0, 0.009*	*0, 0.01*	*0, 0.008*	*0, 0.018*	▲ *-0.01, 0.016*	*0.007, 0.028*	*-0.004, 0.037*	*-0.006, 0.032*
		Average	*0, 0.01*	*0, 0.01*	*0, 0.009*	*0.002, 0.018*	▲ *-0.005, 0.015*	*0.003, 0.031*	*0, 0.033*	▲ *0.041, 0.051*
		Old	*0, 0.01*	*-0.003, 0.01*	*-0.002, 0.019*	*-0.006, 0.023*	*0.007, 0.019*	*-0.027, 0.058*	*0.002, 0.036*	*-0.02, 0.036*

Lognormal distributed data

Table 88:

Legend for colour-coded table and times a model ranked best for absolute mean bias of RMST for lognormal distributed data. *Amount of times the model ranked best*

	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F	Total
Non-GPM	*2*	*6*	*5*	*0*	*5*	*2*	*0*	*20*
Internal additive	*2*	*15*	*9*	*4*	*4*	*3*	*0*	*37*
External additive	*2*	*4*	*0*	*0*	*0*	*0*	*0*	*6*
Converging	*0*	*4*	*5*	*0*	*0*	*0*	*0*	*9*
Total	*6*	*29*	*19*	*4*	*9*	*5*	*0*	

Table 89:

Colour-coded table for best ranking model for absolute mean bias of RMST per scenario for lognormal distributed data. Survival and ages as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

		Survival (Lognormal)							
		High		Medium			Low		
		Information		*Information*			*Information*		
Age	*Heterogeneity*	Medium	High	Low	Medium	High	Low	Medium	High
Young	Low	*0, 0.004*	*0.003, 0.018*	*-0.002, 0.016*	*0.002, 0.013*	*-0.005, 0.032*	*-0.012, 0.013*	*-0.009, 0.013* ▲	*-0.024, 0.024* ▲
Young	Medium	*-0.002, 0.004*	*0.004, 0.019*	*-0.004, 0.016*	*0.003, 0.014*	*-0.002, 0.032*	*-0.013, 0.013*	*-0.005, 0.012*	*-0.011, 0.017* ▲
Young	High	*0.001, 0.008*	*-0.005, 0.03*	*0, 0.009*	*-0.005, 0.017*	*-0.014, 0.04*	*0, 0.01*	*-0.011, 0.016* ▲	*-0.008, 0.034*
Average	Low	*-0.001, 0.005*	*-0.003, 0.023*	*-0.004, 0.007* ▲	*0.001, 0.012* ▲	*-0.025, 0.039* ▲	*-0.012, 0.012* ▲	*-0.005, 0.012* ▲	*-0.014, 0.016* ▲
Average	Medium	*-0.001, 0.005*	*-0.003, 0.023*	*-0.004, 0.007* ▲	*0.001, 0.012*	*-0.023, 0.041* ▲	*-0.012, 0.012* ▲	*0, 0.012* ▲	*-0.001, 0.015*
Average	High	*0, 0.008*	*0.002, 0.031*	*0.001, 0.009*	*-0.001, 0.017*	*-0.003, 0.043*	*0.001, 0.01*	*-0.007, 0.014* ▲	*0.036, 0.046* ▲
Old	Low	*-0.007, 0.009*	*-0.026, 0.03*	*0.002, 0.007*	*-0.001, 0.014*	*-0.009, 0.03*	*-0.002, 0.006* ▲	*0.009, 0.014* ▲	*0.002, 0.024*
Old	Medium	*-0.008, 0.01*	*-0.027, 0.031*	*0.001, 0.007*	*-0.001, 0.014*	*-0.004, 0.03*	*-0.002, 0.011*	*0, 0.017*	*-0.006, 0.027*
Old	High	*0, 0.018*	*-0.022, 0.055*	*0, 0.01*	*-0.001, 0.022* ▲	*0.009, 0.035*	*-0.003, 0.01*	*0.006, 0.019*	*-0.022, 0.038*

Table 90:

Colour-coded table for best ranking model for absolute mean bias of RMST per scenario for lognormal distributed data. Information and heterogeneity as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

		Information							
		Low		Medium			High		
		Survival (Lognormal)		*Survival (Lognormal)*			*Survival (Lognormal)*		
Heterogeneity	**Age**	Medium	Low	High	Medium	Low	High	Medium	Low
Low	Young	*-0.002, 0.016*	*-0.012, 0.013*	*0, 0.004*	*0.002, 0.013*	▲ *-0.009, 0.013*	*0.003, 0.018*	*-0.005, 0.032*	▲ *-0.024, 0.024*
Low	Average	▲ *-0.004, 0.007*	▲ *-0.012, 0.012*	*-0.001, 0.005*	▲ *0.001, 0.012*	▲ *-0.005, 0.012*	*-0.003, 0.023*	▲ *-0.025, 0.039*	▲ *-0.014, 0.016*
Low	Old	*0.002, 0.007*	▲ *-0.002, 0.006*	*-0.007, 0.009*	*-0.001, 0.014*	▲ *0.009, 0.014*	*-0.026, 0.03*	*-0.009, 0.03*	*0.002, 0.024*
Medium	Young	*-0.004, 0.016*	*-0.013, 0.013*	*-0.002, 0.004*	*0.003, 0.014*	*-0.005, 0.012*	*0.004, 0.019*	*-0.002, 0.032*	▲ *-0.011, 0.017*
Medium	Average	▲ *-0.004, 0.007*	▲ *-0.012, 0.012*	*-0.001, 0.005*	*0.001, 0.012*	▲ *0, 0.012*	*-0.003, 0.023*	▲ *-0.023, 0.041*	*-0.001, 0.015*
Medium	Old	*0.001, 0.007*	*-0.002, 0.011*	*-0.008, 0.01*	*-0.001, 0.014*	*0, 0.017*	*-0.027, 0.031*	*-0.004, 0.03*	*-0.006, 0.027*
High	Young	*0, 0.009*	*0, 0.01*	*0.001, 0.008*	*-0.005, 0.017*	▲ *-0.011, 0.016*	*-0.005, 0.03*	*-0.014, 0.04*	*-0.008, 0.034*
High	Average	*0.001, 0.009*	*0.001, 0.01*	*0, 0.008*	*-0.001, 0.017*	▲ *-0.007, 0.014*	*0.002, 0.031*	*-0.003, 0.043*	▲ *0.036, 0.046*
High	Old	*0, 0.01*	*-0.003, 0.01*	*0, 0.018*	▲ *-0.001, 0.022*	*0.006, 0.019*	*-0.022, 0.055*	*0.009, 0.035*	*-0.022, 0.038*

Gompertz distributed data

Table 91:

Legend for colour-coded table and times a model ranked best for absolute mean bias of RMST for Gompertz distributed data. *Amount of times the model ranked best*

	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F	Total
Non-GPM	*2*	*2*	*6*	*1*	*8*	*0*	*0*	*19*
Internal additive	*5*	*9*	*5*	*5*	*3*	*4*	*2*	*33*
External additive	*2*	*2*	*2*	*1*	*0*	*0*	*0*	*7*
Converging	*5*	*4*	*1*	*1*	*1*	*1*	*0*	*13*
Total	*14*	*17*	*14*	*8*	*12*	*5*	*2*	

Table 92:

Colour-coded table for best ranking model for absolute mean bias of RMST per scenario for Gompertz distributed data. Survival and ages as main axes. *(Bias, RMSE), ▲ = Significant result (Wilcoxon)*

Age	*Heterogeneity*	Survival (Gompertz): High, *Information*: Medium	High, High	Medium, Low	Medium, Medium	Medium, High	Low, Low	Low, Medium	Low, High
Young	Low	*0, 0.004*	*-0.006, 0.012*	*-0.004, 0.006*	*-0.002, 0.02*	*-0.002, 0.026*	*-0.016, 0.017*	*-0.034, 0.034*	*-0.04, 0.04*
Young	Medium	*-0.002, 0.003*	*-0.007, 0.013*	*-0.005, 0.007*	*0, 0.02*	*0.005, 0.028*	*-0.017, 0.017*	*-0.027, 0.029*	*-0.028, 0.029*
Young	High	*0, 0.009*	*0.002, 0.029*	*0.001, 0.009*	*0.001, 0.015*	*-0.008, 0.04*	*-0.005, 0.011*	*-0.001, 0.015*	*-0.002, 0.033*
Average	Low	*-0.001, 0.005*	*-0.007, 0.014*	*-0.003, 0.006*	*-0.002, 0.008* ▲	*-0.001, 0.034*	*-0.016, 0.016*	*-0.029, 0.029*	*-0.03, 0.03*
Average	Medium	*-0.002, 0.005*	*-0.009, 0.015*	*-0.003, 0.006*	*-0.002, 0.02*	*-0.012, 0.046*	*-0.015, 0.016* ▲	*-0.022, 0.026*	*-0.013, 0.018*
Average	High	*-0.001, 0.009*	*0.007, 0.034*	*-0.001, 0.009*	*-0.001, 0.016*	*0.002, 0.047*	*-0.004, 0.011*	*-0.002, 0.017*	*-0.001, 0.038*
Old	Low	*-0.007, 0.008* ▲	*-0.027, 0.03*	*0.007, 0.008*	*-0.019, 0.026*	*-0.031, 0.037*	*-0.002, 0.011* ▲	*0.001, 0.015* ▲	*-0.026, 0.033* ▲
Old	Medium	*-0.008, 0.01*	*-0.027, 0.031*	*0.005, 0.008* ▲	*-0.012, 0.023*	*-0.006, 0.037*	*-0.006, 0.01*	*0.006, 0.016* ▲	*-0.031, 0.04* ▲
Old	High	*-0.004, 0.019*	*-0.023, 0.054*	*0, 0.01*	*0.01, 0.019*	*-0.004, 0.034*	*0, 0.011*	*0.003, 0.016* ▲	*-0.008, 0.039*

Table 93:

Colour-coded table for best ranking model for absolute mean bias of RMST per scenario for Gompertz distributed data. Information and heterogeneity as main axes. *(Bias, RMSE),* ▲ *= Significant result (Wilcoxon)*

		Information: Low		Information: Medium			Information: High		
		Survival (Gompertz)		*Survival (Gompertz)*			*Survival (Gompertz)*		
Heterogeneity	**Age**	Medium	Low	High	Medium	Low	High	Medium	Low
Low	Young	-0.004, 0.006	-0.016, 0.017	0, 0.004	-0.002, 0.02	-0.034, 0.034	-0.006, 0.012	-0.002, 0.026	-0.04, 0.04
Low	Average	-0.003, 0.006	-0.016, 0.016	-0.001, 0.005	-0.002, 0.008 ▲	-0.029, 0.029	-0.007, 0.014	-0.001, 0.034	-0.03, 0.03
Low	Old	0.007, 0.008	-0.002, 0.011 ▲	-0.007, 0.008 ▲	-0.019, 0.026	0.001, 0.015 ▲	-0.027, 0.03	-0.031, 0.037	-0.026, 0.033 ▲
Medium	Young	-0.005, 0.007	-0.017, 0.017	-0.002, 0.003	0, 0.02	-0.027, 0.029	-0.007, 0.013	0.005, 0.028	-0.028, 0.029
Medium	Average	-0.003, 0.006	-0.015, 0.016 ▲	-0.002, 0.005	-0.002, 0.02	-0.022, 0.026	-0.009, 0.015	-0.012, 0.046	-0.013, 0.018
Medium	Old	0.005, 0.008 ▲	-0.006, 0.01	-0.008, 0.01	-0.012, 0.023	0.006, 0.016 ▲	-0.027, 0.031	-0.006, 0.037	-0.031, 0.04 ▲
High	Young	0.001, 0.009	-0.005, 0.011	0, 0.009	0.001, 0.015	-0.001, 0.015	0.002, 0.029	-0.008, 0.04	-0.002, 0.033
High	Average	-0.001, 0.009	-0.004, 0.011	-0.001, 0.009	-0.001, 0.016	-0.002, 0.017	0.007, 0.034	0.002, 0.047	-0.001, 0.038
High	Old	0, 0.01	0, 0.011	-0.004, 0.019	0.01, 0.019	0.003, 0.016 ▲	-0.023, 0.054	-0.004, 0.034	-0.008, 0.039

Appendix C: Raw results of bias of overall extrapolated mean survival per scenario per method and distribution

Table 94:

Mean bias of overall mean survival – Non-GPM models

Scenario characteristics						*Distribution for extrapolation*						
Age	**Survival**	**Heterogeneity**	**Information**	**Distribution**	**Mean of method**	**Exponential**	**Weibull**	**Log-logistic**	**Lognormal**	**Gompertz**	**Gen. Gamma**	**Gen. F**
Average	High	High	High	Gompertz	4.76	1.01	3.17	5.58	7.22	7.41	4.15	NA
Average	High	High	High	Log-logistic	5.84	1.91	4.51	6.53	8.20	8.81	5.11	NA
Average	High	High	High	Lognormal	6.34	2.29	5.18	7.00	8.62	9.48	5.47	NA
Average	High	High	High	Weibull	5.37	1.50	3.94	6.14	7.81	8.22	4.63	NA
Average	High	High	Medium	Gompertz	4.91	-1.76	0.89	5.29	9.80	10.31	NA	NA
Average	High	High	Medium	Log-logistic	6.00	-1.34	2.46	6.22	10.54	12.09	NA	NA
Average	High	High	Medium	Lognormal	6.64	-1.32	3.31	6.70	10.76	13.77	NA	NA
Average	High	High	Medium	Weibull	6.00	-1.50	2.01	6.00	10.39	11.75	7.36	NA
Average	High	Low	High	Gompertz	7.47	9.46	7.16	9.12	11.81	1.38	5.91	NA
Average	High	Low	High	Log-logistic	6.29	8.48	5.76	7.95	10.71	0.06	4.80	NA
Average	High	Low	High	Lognormal	5.84	7.48	4.77	7.07	9.65	0.57	5.52	NA
Average	High	Low	High	Weibull	6.65	8.64	6.16	8.29	11.01	0.72	5.08	NA
Average	High	Low	Medium	Gompertz	10.21	11.84	NA	10.36	15.31	3.33	NA	NA
Average	High	Low	Medium	Log-logistic	8.78	11.41	NA	8.82	14.25	0.64	NA	NA
Average	High	Low	Medium	Lognormal	5.98	10.26	NA	5.68	11.93	-3.95	NA	NA
Average	High	Low	Medium	Weibull	9.04	11.23	NA	9.07	14.37	1.50	NA	NA
Average	High	Medium	High	Gompertz	6.79	8.74	6.45	8.63	11.40	0.80	4.71	NA
Average	High	Medium	High	Log-logistic	5.73	7.49	5.05	7.36	10.05	0.27	4.17	NA
Average	High	Medium	High	Lognormal	5.60	6.64	4.44	6.75	9.22	1.21	5.27	5.68
Average	High	Medium	High	Weibull	5.97	7.69	5.36	7.64	10.32	0.56	4.26	NA
Average	High	Medium	Medium	Gompertz	10.06	11.25	NA	10.39	15.46	3.14	NA	NA
Average	High	Medium	Medium	Log-logistic	7.92	10.22	NA	8.02	13.63	-0.18	NA	NA
Average	High	Medium	Medium	Lognormal	5.43	8.75	NA	5.30	11.47	-3.79	NA	NA
Average	High	Medium	Medium	Weibull	8.69	10.32	NA	8.73	14.15	1.56	NA	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Average	Low	High	High	Gompertz	1.03	-1.53	-0.09	1.68	1.57	3.58	0.92	1.07
Average	Low	High	High	Log-logistic	0.90	-1.55	-0.30	1.33	1.14	3.70	0.94	1.04
Average	Low	High	High	Lognormal	0.89	-1.55	-0.32	1.30	1.11	3.69	0.94	1.03
Average	Low	High	High	Weibull	0.97	-1.54	-0.21	1.50	1.34	3.72	0.92	1.05
Average	Low	High	Low	Gompertz	3.42	-3.71	-2.04	1.92	2.70	18.24	NA	NA
Average	Low	High	Low	Log-logistic	2.75	-3.77	-2.59	1.05	1.68	17.38	NA	NA
Average	Low	High	Low	Lognormal	2.60	-3.74	-2.65	0.91	1.50	16.96	NA	NA
Average	Low	High	Low	Weibull	2.92	-3.79	-2.42	1.39	2.14	17.70	2.52	NA
Average	Low	High	Medium	Gompertz	1.97	-2.94	-1.15	1.95	2.08	10.74	1.15	NA
Average	Low	High	Medium	Log-logistic	1.75	-2.95	-1.52	1.44	1.41	10.77	1.35	NA
Average	Low	High	Medium	Lognormal	1.74	-2.96	-1.55	1.40	1.37	10.80	1.39	NA
Average	Low	High	Medium	Weibull	1.86	-2.96	-1.39	1.66	1.71	10.88	1.24	NA
Average	Low	Low	High	Gompertz	0.14	-0.04	-0.05	0.74	0.46	-0.05	-0.05	-0.05
Average	Low	Low	High	Log-logistic	-0.08	-0.12	-0.12	-0.01	-0.08	-0.11	-0.11	-0.02
Average	Low	Low	High	Lognormal	-0.04	-0.06	-0.06	0.06	-0.04	-0.07	-0.05	NA
Average	Low	Low	High	Weibull	0.06	-0.04	-0.05	0.44	0.22	-0.05	-0.05	-0.05
Average	Low	Low	Low	Gompertz	2.42	0.26	0.28	3.00	4.29	4.86	1.85	NA
Average	Low	Low	Low	Log-logistic	0.05	0.59	-0.66	0.10	0.45	-0.91	0.75	NA
Average	Low	Low	Low	Lognormal	-0.16	0.65	-0.65	-0.05	0.12	-0.85	NA	NA
Average	Low	Low	Low	Weibull	1.46	0.78	0.06	2.03	3.17	0.57	2.16	NA
Average	Low	Low	Medium	Gompertz	0.44	0.08	0.03	1.32	1.26	-0.02	0.00	NA
Average	Low	Low	Medium	Log-logistic	-0.22	-0.18	-0.37	0.01	-0.11	-0.41	-0.26	NA
Average	Low	Low	Medium	Lognormal	-0.09	-0.05	-0.25	0.11	-0.02	-0.29	-0.02	NA
Average	Low	Low	Medium	Weibull	0.22	0.13	-0.05	0.77	0.65	-0.13	-0.04	NA
Average	Low	Medium	High	Gompertz	0.15	-0.07	-0.06	0.80	0.53	-0.04	-0.06	-0.04
Average	Low	Medium	High	Log-logistic	-0.07	-0.16	-0.17	0.08	-0.05	-0.09	-0.10	-0.01
Average	Low	Medium	High	Lognormal	-0.03	-0.10	-0.11	0.13	-0.02	-0.07	-0.04	NA
Average	Low	Medium	High	Weibull	0.07	-0.06	-0.07	0.50	0.27	-0.06	-0.06	-0.04
Average	Low	Medium	Low	Gompertz	2.19	-0.14	-0.04	2.62	3.73	5.34	1.62	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Average	Low	Medium	Low	Log-logistic	-0.28	0.06	-0.86	0.08	0.35	-1.07	NA	NA
Average	Low	Medium	Low	Lognormal	-0.32	0.15	-0.84	-0.02	0.13	-1.02	NA	NA
Average	Low	Medium	Low	Weibull	1.05	0.24	-0.30	1.60	2.50	0.58	1.67	NA
Average	Low	Medium	Medium	Gompertz	0.41	-0.14	-0.12	1.32	1.26	0.19	-0.07	NA
Average	Low	Medium	Medium	Log-logistic	-0.23	-0.30	-0.47	0.13	-0.05	-0.49	-0.21	NA
Average	Low	Medium	Medium	Lognormal	-0.13	-0.21	-0.39	0.20	0.00	-0.40	0.01	NA
Average	Low	Medium	Medium	Weibull	0.16	-0.06	-0.19	0.84	0.70	-0.23	-0.09	NA
Average	Medium	High	High	Gompertz	2.36	-2.76	0.85	3.03	3.37	6.88	2.45	2.70
Average	Medium	High	High	Log-logistic	2.36	-2.86	1.03	3.12	3.42	6.83	2.63	NA
Average	Medium	High	High	Lognormal	2.31	-2.83	0.89	2.92	3.12	6.64	2.66	2.78
Average	Medium	High	High	Weibull	2.52	-2.70	1.28	3.49	3.92	6.76	2.39	NA
Average	Medium	High	Low	Gompertz	3.76	-7.27	-3.12	2.30	5.15	20.51	4.98	NA
Average	Medium	High	Low	Log-logistic	3.76	-7.37	-3.00	2.34	4.89	20.95	4.75	NA
Average	Medium	High	Low	Lognormal	3.43	-7.24	-3.49	1.74	3.93	20.55	5.11	NA
Average	Medium	High	Low	Weibull	4.72	-7.10	-1.61	3.82	6.64	21.98	4.57	NA
Average	Medium	High	Medium	Gompertz	3.19	-5.72	-1.15	3.15	4.46	15.31	3.09	NA
Average	Medium	High	Medium	Log-logistic	3.27	-5.81	-0.90	3.26	4.43	15.19	3.42	NA
Average	Medium	High	Medium	Lognormal	3.08	-5.69	-1.16	2.91	3.85	14.87	3.72	NA
Average	Medium	High	Medium	Weibull	3.71	-5.56	-0.08	4.08	5.47	15.40	2.97	NA
Average	Medium	Low	High	Gompertz	0.61	-0.74	-0.46	2.00	2.29	0.67	-0.09	NA
Average	Medium	Low	High	Log-logistic	0.87	-1.50	-0.62	1.69	1.94	3.29	0.42	NA
Average	Medium	Low	High	Lognormal	1.04	-1.54	-0.58	1.49	1.50	4.11	1.27	NA
Average	Medium	Low	High	Weibull	1.89	-0.69	0.60	3.42	3.93	3.60	0.47	NA
Average	Medium	Low	Low	Gompertz	4.91	-1.07	-0.09	4.94	10.04	10.72	NA	NA
Average	Medium	Low	Low	Log-logistic	2.98	-3.43	-2.32	2.51	6.55	9.46	5.11	NA
Average	Medium	Low	Low	Lognormal	-0.03	-3.81	-4.10	-0.23	2.29	5.71	NA	NA
Average	Medium	Low	Low	Weibull	7.63	-3.17	1.69	7.24	11.76	20.63	NA	NA
Average	Medium	Low	Medium	Gompertz	1.71	-1.31	-1.00	3.55	5.80	3.03	0.18	NA
Average	Medium	Low	Medium	Log-logistic	1.06	-2.98	-2.17	2.04	3.59	6.14	-0.27	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Average	Medium	Low	Medium	Lognormal	1.14	-3.21	-2.62	1.13	1.88	7.63	2.01	NA
Average	Medium	Low	Medium	Weibull	4.64	-2.10	0.93	5.74	8.07	14.24	0.99	NA
Average	Medium	Medium	High	Gompertz	0.78	-0.99	-0.47	2.01	2.30	1.68	0.14	NA
Average	Medium	Medium	High	Log-logistic	1.03	-1.63	-0.53	1.79	2.04	3.91	0.62	NA
Average	Medium	Medium	High	Lognormal	1.14	-1.67	-0.51	1.59	1.61	4.50	1.36	NA
Average	Medium	Medium	High	Weibull	1.86	-0.97	0.50	3.26	3.77	4.09	0.55	NA
Average	Medium	Medium	Low	Gompertz	3.92	-2.04	-0.99	4.00	8.76	9.86	NA	NA
Average	Medium	Medium	Low	Log-logistic	2.45	-3.94	-2.67	2.22	6.08	10.55	NA	NA
Average	Medium	Medium	Low	Lognormal	1.21	-4.22	-4.18	-0.14	2.40	7.46	5.93	NA
Average	Medium	Medium	Low	Weibull	7.13	-3.66	1.07	6.63	10.95	20.65	NA	NA
Average	Medium	Medium	Medium	Gompertz	1.61	-1.95	-1.35	3.16	5.23	4.51	0.06	NA
Average	Medium	Medium	Medium	Log-logistic	1.39	-3.29	-2.19	2.06	3.55	8.18	0.04	NA
Average	Medium	Medium	Medium	Lognormal	1.40	-3.48	-2.63	1.22	1.99	9.25	2.08	NA
Average	Medium	Medium	Medium	Weibull	4.31	-2.52	0.51	5.26	7.48	14.29	0.87	NA
Old	High	High	High	Gompertz	0.46	0.07	0.15	1.05	1.28	0.18	0.05	NA
Old	High	High	High	Log-logistic	0.48	0.10	0.19	1.07	1.32	0.19	0.01	NA
Old	High	High	High	Lognormal	0.51	0.12	0.22	1.10	1.36	0.23	0.00	NA
Old	High	High	High	Weibull	0.48	0.09	0.18	1.07	1.31	0.19	0.03	NA
Old	High	High	Medium	Gompertz	1.47	-0.13	0.23	2.24	3.66	1.90	0.90	NA
Old	High	High	Medium	Log-logistic	1.71	-0.10	0.36	2.30	3.73	2.16	1.13	2.38
Old	High	High	Medium	Lognormal	1.77	-0.11	0.50	2.39	3.81	2.63	1.39	NA
Old	High	High	Medium	Weibull	1.62	-0.10	0.40	2.36	3.78	2.24	1.05	NA
Old	High	Low	High	Gompertz	1.18	1.63	1.00	2.03	2.49	0.01	-0.10	NA
Old	High	Low	High	Log-logistic	1.13	1.57	0.93	1.97	2.41	-0.03	-0.06	NA
Old	High	Low	High	Lognormal	1.10	1.51	0.87	1.90	2.32	-0.03	0.03	NA
Old	High	Low	High	Weibull	1.14	1.58	0.95	1.98	2.42	-0.01	-0.05	NA
Old	High	Low	Medium	Gompertz	3.65	3.74	2.85	4.47	6.37	1.31	3.16	NA
Old	High	Low	Medium	Log-logistic	3.48	3.72	2.64	4.32	6.25	0.94	3.00	NA
Old	High	Low	Medium	Lognormal	3.16	3.59	2.25	4.03	6.00	0.39	2.72	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Old	High	Low	Medium	Weibull	3.56	3.72	2.74	4.39	6.31	1.17	3.03	NA
Old	High	Medium	High	Gompertz	0.85	1.11	0.63	1.63	2.05	-0.16	-0.18	NA
Old	High	Medium	High	Log-logistic	0.80	1.05	0.56	1.55	1.96	-0.19	-0.14	NA
Old	High	Medium	High	Lognormal	0.77	0.98	0.50	1.49	1.87	-0.18	-0.06	NA
Old	High	Medium	High	Weibull	0.81	1.07	0.58	1.58	1.98	-0.18	-0.16	NA
Old	High	Medium	Medium	Gompertz	2.96	2.80	2.10	3.83	5.69	1.01	2.32	NA
Old	High	Medium	Medium	Log-logistic	2.75	2.75	1.84	3.62	5.50	0.61	2.17	NA
Old	High	Medium	Medium	Lognormal	2.41	2.55	1.43	3.31	5.22	0.08	1.88	NA
Old	High	Medium	Medium	Weibull	2.75	2.69	1.84	3.63	5.50	0.68	2.16	NA
Old	Low	High	High	Gompertz	0.30	-0.36	0.06	0.77	0.66	0.45	0.19	NA
Old	Low	High	High	Log-logistic	0.22	-0.36	-0.01	0.61	0.48	0.46	0.17	NA
Old	Low	High	High	Lognormal	0.22	-0.35	-0.01	0.61	0.47	0.46	0.17	NA
Old	Low	High	High	Weibull	0.24	-0.36	0.02	0.69	0.56	0.46	0.17	0.17
Old	Low	High	Low	Gompertz	1.64	-1.77	-0.59	1.46	1.89	7.60	1.24	NA
Old	Low	High	Low	Log-logistic	1.32	-1.76	-0.95	1.01	1.35	6.94	1.31	NA
Old	Low	High	Low	Lognormal	1.37	-1.76	-0.93	1.03	1.38	7.03	1.43	NA
Old	Low	High	Low	Weibull	1.53	-1.78	-0.78	1.25	1.67	7.41	1.38	NA
Old	Low	High	Medium	Gompertz	0.92	-1.15	-0.15	1.23	1.31	3.70	0.59	NA
Old	Low	High	Medium	Log-logistic	0.79	-1.14	-0.34	0.98	0.98	3.61	0.63	NA
Old	Low	High	Medium	Lognormal	0.79	-1.13	-0.34	0.98	0.98	3.62	0.65	NA
Old	Low	High	Medium	Weibull	0.86	-1.14	-0.25	1.11	1.15	3.71	0.62	NA
Old	Low	Low	High	Gompertz	0.10	-0.04	-0.05	0.55	0.37	-0.05	-0.05	-0.05
Old	Low	Low	High	Log-logistic	-0.03	-0.06	-0.06	0.06	0.01	-0.06	-0.07	NA
Old	Low	Low	High	Lognormal	-0.02	-0.05	-0.05	0.08	0.01	-0.06	-0.05	NA
Old	Low	Low	High	Weibull	0.05	-0.04	-0.05	0.37	0.21	-0.05	-0.05	-0.05
Old	Low	Low	Low	Gompertz	1.37	0.22	0.22	1.98	2.77	1.99	1.05	NA
Old	Low	Low	Low	Log-logistic	0.09	0.58	-0.45	0.31	0.70	-0.69	NA	NA
Old	Low	Low	Low	Lognormal	-0.02	0.57	-0.50	0.12	0.40	-0.71	NA	NA
Old	Low	Low	Low	Weibull	1.00	0.67	0.08	1.57	2.38	0.17	1.11	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Old	Low	Low	Medium	Gompertz	0.31	0.06	0.02	0.92	0.92	-0.03	-0.01	NA
Old	Low	Low	Medium	Log-logistic	-0.11	-0.09	-0.24	0.10	0.03	-0.27	-0.17	NA
Old	Low	Low	Medium	Lognormal	-0.05	-0.03	-0.19	0.14	0.06	-0.22	-0.07	NA
Old	Low	Low	Medium	Weibull	0.18	0.10	-0.04	0.61	0.56	-0.11	-0.04	NA
Old	Low	Medium	High	Gompertz	0.10	-0.05	-0.05	0.53	0.38	-0.04	-0.05	-0.04
Old	Low	Medium	High	Log-logistic	-0.02	-0.07	-0.08	0.12	0.04	-0.06	-0.06	-0.03
Old	Low	Medium	High	Lognormal	-0.02	-0.06	-0.06	0.12	0.03	-0.06	-0.05	-0.03
Old	Low	Medium	High	Weibull	0.05	-0.05	-0.06	0.37	0.22	-0.05	-0.05	-0.05
Old	Low	Medium	Low	Gompertz	1.12	-0.08	-0.02	1.70	2.38	2.00	0.73	NA
Old	Low	Medium	Low	Log-logistic	-0.06	0.17	-0.58	0.28	0.60	-0.77	NA	NA
Old	Low	Medium	Low	Lognormal	0.01	0.19	-0.61	0.17	0.41	-0.79	0.72	NA
Old	Low	Medium	Low	Weibull	0.66	0.21	-0.21	1.20	1.82	0.11	0.82	NA
Old	Low	Medium	Medium	Gompertz	0.26	-0.09	-0.08	0.87	0.87	0.03	-0.05	NA
Old	Low	Medium	Medium	Log-logistic	-0.10	-0.15	-0.29	0.19	0.10	-0.30	-0.13	NA
Old	Low	Medium	Medium	Lognormal	-0.06	-0.12	-0.26	0.20	0.11	-0.27	-0.04	NA
Old	Low	Medium	Medium	Weibull	0.13	-0.04	-0.14	0.60	0.56	-0.17	-0.07	NA
Old	Medium	High	High	Gompertz	0.43	-0.44	0.18	0.99	0.97	0.76	0.27	0.28
Old	Medium	High	High	Log-logistic	0.45	-0.48	0.22	1.02	0.99	0.79	0.30	0.31
Old	Medium	High	High	Lognormal	0.44	-0.47	0.21	0.98	0.92	0.79	0.34	0.34
Old	Medium	High	High	Weibull	0.51	-0.45	0.29	1.16	1.17	0.80	0.28	0.30
Old	Medium	High	Low	Gompertz	2.04	-2.69	-0.55	1.98	3.23	8.24	2.05	NA
Old	Medium	High	Low	Log-logistic	2.15	-2.80	-0.47	2.01	3.10	8.78	2.28	NA
Old	Medium	High	Low	Lognormal	1.99	-2.78	-0.63	1.83	2.81	8.70	NA	NA
Old	Medium	High	Low	Weibull	2.49	-2.74	0.15	2.61	3.85	9.13	1.94	NA
Old	Medium	High	Medium	Gompertz	1.44	-1.68	0.09	1.83	2.30	5.04	1.05	NA
Old	Medium	High	Medium	Log-logistic	1.49	-1.76	0.18	1.87	2.28	5.18	1.18	NA
Old	Medium	High	Medium	Lognormal	1.43	-1.76	0.10	1.74	2.07	5.10	1.32	NA
Old	Medium	High	Medium	Weibull	1.64	-1.71	0.47	2.17	2.69	5.26	0.93	NA
Old	Medium	Low	High	Gompertz	0.38	0.06	0.07	0.99	1.00	0.07	0.07	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Old	Medium	Low	High	Log-logistic	0.39	-0.04	0.10	0.93	0.94	0.34	0.19	0.24
Old	Medium	Low	High	Lognormal	0.38	-0.07	0.09	0.81	0.76	0.46	0.32	0.32
Old	Medium	Low	High	Weibull	0.59	0.05	0.28	1.32	1.36	0.38	0.13	NA
Old	Medium	Low	Low	Gompertz	2.79	0.25	0.43	3.00	5.26	3.93	3.89	NA
Old	Medium	Low	Low	Log-logistic	1.95	-0.80	-0.32	2.23	4.11	3.94	2.52	NA
Old	Medium	Low	Low	Lognormal	0.86	-1.14	-1.47	0.80	2.12	1.96	2.91	NA
Old	Medium	Low	Low	Weibull	3.59	-1.08	1.36	3.91	5.84	8.27	3.26	NA
Old	Medium	Low	Medium	Gompertz	1.10	0.07	0.13	2.20	3.08	0.81	0.33	NA
Old	Medium	Low	Medium	Log-logistic	0.90	-0.57	-0.21	1.71	2.36	1.72	0.39	NA
Old	Medium	Low	Medium	Lognormal	0.82	-0.75	-0.48	1.29	1.65	2.11	1.12	NA
Old	Medium	Low	Medium	Weibull	1.94	-0.42	0.85	2.84	3.71	4.13	0.52	NA
Old	Medium	Medium	High	Gompertz	0.33	-0.05	0.02	0.91	0.92	0.12	0.05	NA
Old	Medium	Medium	High	Log-logistic	0.35	-0.15	0.03	0.85	0.85	0.35	0.14	NA
Old	Medium	Medium	High	Lognormal	0.33	-0.18	0.03	0.74	0.69	0.45	0.25	NA
Old	Medium	Medium	High	Weibull	0.50	-0.07	0.19	1.21	1.24	0.37	0.09	NA
Old	Medium	Medium	Low	Gompertz	2.20	-0.30	0.03	2.62	4.74	3.92	NA	NA
Old	Medium	Medium	Low	Log-logistic	1.72	-1.16	-0.59	1.96	3.70	4.14	2.26	NA
Old	Medium	Medium	Low	Lognormal	0.98	-1.40	-1.43	0.89	2.17	2.89	2.76	NA
Old	Medium	Medium	Low	Weibull	3.23	-1.30	0.99	3.58	5.44	7.96	2.71	NA
Old	Medium	Medium	Medium	Gompertz	0.92	-0.28	-0.08	1.93	2.71	1.02	0.23	NA
Old	Medium	Medium	Medium	Log-logistic	0.85	-0.77	-0.30	1.59	2.18	2.09	0.35	NA
Old	Medium	Medium	Medium	Lognormal	0.83	-0.91	-0.51	1.25	1.59	2.55	1.03	NA
Old	Medium	Medium	Medium	Weibull	1.73	-0.65	0.61	2.57	3.38	4.06	0.43	NA
Young	High	High	High	Gompertz	6.42	-1.13	3.48	6.89	9.39	13.29	6.61	NA
Young	High	High	High	Log-logistic	7.92	0.08	5.61	8.29	10.75	15.06	7.71	NA
Young	High	High	High	Lognormal	9.17	0.69	6.91	9.16	11.46	16.07	NA	10.73
Young	High	High	High	Weibull	7.63	-0.42	4.83	7.81	10.31	14.41	7.26	9.24
Young	High	High	Medium	Gompertz	5.01	-5.87	-1.45	4.79	11.30	14.01	7.26	NA
Young	High	High	Medium	Log-logistic	6.10	-5.67	0.99	6.18	12.34	16.64	NA	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Young	High	High	Medium	Lognormal	6.54	-5.95	1.84	6.52	12.31	17.99	NA	NA
Young	High	High	Medium	Weibull	5.64	-5.75	0.09	5.72	12.03	16.10	NA	NA
Young	High	Low	High	Gompertz	12.11	13.71	11.36	13.62	17.44	4.44	NA	NA
Young	High	Low	High	Log-logistic	8.84	11.08	7.59	10.40	14.51	0.60	NA	NA
Young	High	Low	High	Lognormal	7.69	8.90	5.43	8.48	12.36	1.35	9.58	NA
Young	High	Low	High	Weibull	9.69	11.49	8.64	11.27	15.26	2.23	9.28	NA
Young	High	Low	Medium	Gompertz	13.09	15.60	NA	12.59	19.46	4.71	NA	NA
Young	High	Low	Medium	Log-logistic	9.22	13.94	NA	8.30	16.35	-1.70	NA	NA
Young	High	Low	Medium	Lognormal	3.86	11.32	NA	2.21	11.72	-9.80	NA	NA
Young	High	Low	Medium	Weibull	10.38	13.92	NA	9.71	17.32	0.57	NA	NA
Young	High	Medium	High	Gompertz	10.66	12.89	9.99	12.75	16.91	2.16	9.25	NA
Young	High	Medium	High	Log-logistic	8.07	9.82	6.62	9.63	13.75	0.53	NA	NA
Young	High	Medium	High	Lognormal	7.63	7.92	5.27	8.31	12.05	2.83	9.37	NA
Young	High	Medium	High	Weibull	8.63	10.43	7.44	10.40	14.51	1.20	7.79	NA
Young	High	Medium	Medium	Gompertz	5.49	NA	NA	NA	NA	5.49	NA	NA
Young	High	Medium	Medium	Log-logistic	8.28	12.40	NA	7.58	15.93	-2.78	NA	NA
Young	High	Medium	Medium	Lognormal	-8.81	NA	NA	NA	NA	-8.81	NA	NA
Young	High	Medium	Medium	Weibull	9.80	12.92	NA	9.26	17.19	-0.20	NA	NA
Young	Low	High	High	Gompertz	1.19	-2.36	-0.59	1.86	1.68	5.50	0.96	1.28
Young	Low	High	High	Log-logistic	0.99	-2.40	-0.90	1.36	1.08	5.63	0.97	1.19
Young	Low	High	High	Lognormal	1.01	-2.38	-0.89	1.38	1.09	5.68	1.00	1.20
Young	Low	High	High	Weibull	1.11	-2.38	-0.76	1.62	1.39	5.69	0.97	1.23
Young	Low	High	Low	Gompertz	4.03	-4.68	-2.95	1.91	2.82	24.40	2.66	NA
Young	Low	High	Low	Log-logistic	3.42	-4.72	-3.53	0.80	1.46	23.26	3.23	NA
Young	Low	High	Low	Lognormal	3.44	-4.71	-3.53	0.78	1.43	23.28	3.40	NA
Young	Low	High	Low	Weibull	3.77	-4.78	-3.33	1.32	2.18	24.13	3.11	NA
Young	Low	High	Medium	Gompertz	2.42	-3.86	-1.91	2.08	2.18	14.90	1.15	NA
Young	Low	High	Medium	Log-logistic	2.10	-3.91	-2.38	1.36	1.26	14.90	1.34	NA
Young	Low	High	Medium	Lognormal	2.12	-3.88	-2.37	1.37	1.25	14.94	1.43	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Young	Low	High	Medium	Weibull	2.27	-3.90	-2.19	1.72	1.74	15.07	1.17	NA
Young	Low	Low	High	Gompertz	0.15	-0.04	-0.05	0.79	0.48	-0.05	-0.05	-0.05
Young	Low	Low	High	Log-logistic	-0.11	-0.14	-0.15	-0.02	-0.10	-0.13	-0.13	NA
Young	Low	Low	High	Lognormal	-0.04	-0.06	-0.06	0.07	-0.04	-0.07	-0.05	NA
Young	Low	Low	High	Weibull	0.08	-0.04	-0.05	0.46	0.22	-0.05	-0.05	NA
Young	Low	Low	Low	Gompertz	3.03	0.26	0.31	3.40	4.86	6.93	2.44	NA
Young	Low	Low	Low	Log-logistic	0.12	0.60	-0.68	0.11	0.43	-0.92	1.17	NA
Young	Low	Low	Low	Lognormal	0.25	0.65	-0.68	-0.09	0.05	-0.88	2.41	NA
Young	Low	Low	Low	Weibull	1.62	0.79	0.04	2.17	3.40	0.80	2.51	NA
Young	Low	Low	Medium	Gompertz	0.48	0.08	0.03	1.44	1.33	-0.01	0.00	NA
Young	Low	Low	Medium	Log-logistic	-0.25	-0.21	-0.40	-0.01	-0.14	-0.44	-0.28	NA
Young	Low	Low	Medium	Lognormal	-0.09	-0.06	-0.26	0.11	-0.04	-0.30	-0.02	NA
Young	Low	Low	Medium	Weibull	0.24	0.13	-0.05	0.82	0.68	-0.13	-0.04	NA
Young	Low	Medium	High	Gompertz	0.17	-0.07	-0.06	0.89	0.57	-0.03	-0.06	-0.03
Young	Low	Medium	High	Log-logistic	-0.09	-0.20	-0.20	0.07	-0.09	-0.10	-0.13	-0.02
Young	Low	Medium	High	Lognormal	-0.04	-0.12	-0.12	0.14	-0.04	-0.08	-0.04	NA
Young	Low	Medium	High	Weibull	0.07	-0.06	-0.07	0.54	0.27	-0.06	-0.06	-0.04
Young	Low	Medium	Low	Gompertz	2.61	-0.16	-0.07	2.90	4.14	7.05	1.80	NA
Young	Low	Medium	Low	Log-logistic	-0.04	0.02	-0.92	0.04	0.31	-1.10	1.43	NA
Young	Low	Medium	Low	Lognormal	-0.36	0.12	-0.88	-0.06	0.08	-1.06	NA	NA
Young	Low	Medium	Low	Weibull	1.27	0.23	-0.31	1.78	2.76	1.10	2.08	NA
Young	Low	Medium	Medium	Gompertz	0.46	-0.15	-0.12	1.48	1.35	0.28	-0.07	NA
Young	Low	Medium	Medium	Log-logistic	-0.28	-0.36	-0.53	0.10	-0.10	-0.55	-0.25	NA
Young	Low	Medium	Medium	Lognormal	-0.14	-0.23	-0.41	0.20	-0.03	-0.43	0.03	NA
Young	Low	Medium	Medium	Weibull	0.17	-0.07	-0.20	0.90	0.72	-0.24	-0.09	NA
Young	Medium	High	High	Gompertz	2.56	-5.07	0.05	3.29	3.80	10.38	2.88	NA
Young	Medium	High	High	Log-logistic	2.65	-5.18	0.33	3.43	3.89	10.24	3.16	NA
Young	Medium	High	High	Lognormal	2.52	-5.06	0.18	3.20	3.52	9.99	3.31	NA
Young	Medium	High	High	Weibull	2.94	-4.82	0.76	4.03	4.66	10.19	2.82	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Young	Medium	High	Low	Gompertz	3.73	-10.07	-5.41	1.59	5.35	27.21	NA	NA
Young	Medium	High	Low	Log-logistic	4.11	-10.08	-5.25	1.73	5.06	27.90	5.33	NA
Young	Medium	High	Low	Lognormal	3.78	-9.84	-5.70	1.02	3.80	27.40	6.02	NA
Young	Medium	High	Low	Weibull	5.36	-9.65	-3.47	3.68	7.36	28.87	NA	NA
Young	Medium	High	Medium	Gompertz	3.32	-8.37	-3.05	2.83	4.56	20.78	3.16	NA
Young	Medium	High	Medium	Log-logistic	3.42	-8.47	-2.69	3.05	4.60	20.52	3.51	NA
Young	Medium	High	Medium	Lognormal	3.25	-8.28	-2.97	2.65	3.89	20.18	4.06	NA
Young	Medium	High	Medium	Weibull	4.15	-7.99	-1.50	4.28	6.15	20.98	2.98	NA
Young	Medium	Low	High	Gompertz	0.40	-1.48	-1.11	2.15	2.50	0.91	-0.57	NA
Young	Medium	Low	High	Log-logistic	0.72	-2.71	-1.59	1.59	1.88	5.21	-0.04	NA
Young	Medium	Low	High	Lognormal	1.17	-2.68	-1.46	1.45	1.39	6.49	1.33	1.68
Young	Medium	Low	High	Weibull	2.51	-1.22	0.44	4.36	5.09	5.96	0.41	NA
Young	Medium	Low	Low	Gompertz	5.87	-1.92	-0.67	5.59	12.14	14.21	NA	NA
Young	Medium	Low	Low	Log-logistic	3.05	-4.78	-3.51	2.43	7.49	13.61	NA	NA
Young	Medium	Low	Low	Lognormal	0.99	-5.21	-5.55	-1.09	1.89	7.56	8.35	NA
Young	Medium	Low	Low	Weibull	10.09	-3.88	1.82	9.07	15.06	28.43	10.04	NA
Young	Medium	Low	Medium	Gompertz	1.79	-2.17	-1.78	3.91	6.80	4.37	-0.39	NA
Young	Medium	Low	Medium	Log-logistic	0.90	-4.34	-3.45	1.81	3.74	8.72	-1.08	NA
Young	Medium	Low	Medium	Lognormal	1.21	-4.52	-3.86	0.76	1.61	11.15	2.15	NA
Young	Medium	Low	Medium	Weibull	6.14	-2.78	0.66	7.10	10.20	20.48	1.17	NA
Young	Medium	Medium	High	Gompertz	0.69	-1.89	-1.19	2.16	2.54	2.79	-0.28	NA
Young	Medium	Medium	High	Log-logistic	1.03	-2.91	-1.47	1.79	2.09	6.32	0.34	NA
Young	Medium	Medium	High	Lognormal	1.33	-2.89	-1.41	1.57	1.56	7.12	1.49	1.85
Young	Medium	Medium	High	Weibull	2.44	-1.68	0.22	4.10	4.81	6.74	0.46	NA
Young	Medium	Medium	Low	Gompertz	5.05	-3.11	-1.61	4.67	10.88	14.43	NA	NA
Young	Medium	Medium	Low	Log-logistic	2.71	-5.44	-4.05	1.93	6.79	14.32	NA	NA
Young	Medium	Medium	Low	Lognormal	0.07	-5.69	-5.60	-0.87	2.10	10.40	NA	NA
Young	Medium	Medium	Low	Weibull	9.33	-4.66	0.82	8.07	13.76	28.66	NA	NA
Young	Medium	Medium	Medium	Gompertz	1.71	-3.00	-2.32	3.35	5.98	6.62	-0.38	NA
Young	Medium	Medium	Medium	Log-logistic	1.40	-4.77	-3.57	1.80	3.64	11.87	-0.58	NA
Young	Medium	Medium	Medium	Lognormal	1.56	-4.86	-3.91	0.89	1.79	13.25	2.21	NA
Young	Medium	Medium	Medium	Weibull	5.63	-3.45	0.05	6.41	9.31	20.49	0.95	NA

Table 95:

Mean bias of overall mean survival – internal additive hazards models

		Scenario characteristics					*Distribution used for extrapolation*					
Age	**Survival**	**Heterogeneity**	**Information**	**Distribution**	**Mean of method**	**Exponential**	**Weibull**	**Log-logistic**	**Lognormal**	**Gompertz**	**Gen. Gamma**	**Gen. F**
Average	High	High	High	Gompertz	11.72	4.23	7.39	9.23	10.90	19.63	18.97	NA
Average	High	High	High	Log-logistic	14.13	5.84	9.51	10.85	12.42	18.55	27.60	NA
Average	High	High	High	Lognormal	14.41	6.55	10.53	11.65	13.07	16.99	27.69	NA
Average	High	High	High	Weibull	13.53	5.14	8.65	10.19	11.82	20.40	24.98	NA
Average	High	High	Medium	Gompertz	13.56	-0.28	2.72	6.85	11.21	30.41	30.41	NA
Average	High	High	Medium	Log-logistic	13.90	0.48	4.69	8.76	12.18	28.65	28.65	NA
Average	High	High	Medium	Lognormal	14.04	0.60	5.67	9.77	12.48	27.85	27.85	NA
Average	High	High	Medium	Weibull	13.84	0.16	4.08	7.97	11.95	29.44	29.44	NA
Average	High	Low	High	Gompertz	19.48	15.24	13.29	25.58	16.28	20.77	25.75	NA
Average	High	Low	High	Log-logistic	18.61	14.03	11.55	25.04	14.99	20.30	25.73	NA
Average	High	Low	High	Lognormal	17.94	12.75	10.39	22.98	13.98	21.77	25.73	NA
Average	High	Low	High	Weibull	18.82	14.20	12.02	25.24	15.31	20.35	25.80	NA
Average	High	Low	Medium	Gompertz	23.38	16.30	NA	25.74	NA	25.75	25.75	NA
Average	High	Low	Medium	Log-logistic	23.24	15.81	NA	25.71	NA	25.73	25.73	NA
Average	High	Low	Medium	Lognormal	22.91	14.49	NA	25.69	NA	25.74	25.74	NA
Average	High	Low	Medium	Weibull	23.25	15.59	NA	25.78	NA	25.81	25.81	NA
Average	High	Medium	High	Gompertz	19.09	14.65	12.63	25.28	15.91	19.52	26.56	NA
Average	High	Medium	High	Log-logistic	18.00	13.23	11.06	23.26	14.58	19.81	26.06	NA
Average	High	Medium	High	Lognormal	17.38	12.12	10.34	20.69	13.83	21.41	25.91	NA
Average	High	Medium	High	Weibull	18.13	13.41	11.40	23.44	14.89	19.42	26.22	NA
Average	High	Medium	Medium	Gompertz	23.77	15.48	NA	26.51	NA	26.54	26.54	NA
Average	High	Medium	Medium	Log-logistic	23.07	14.38	NA	25.91	NA	26.01	26.01	NA
Average	High	Medium	Medium	Lognormal	20.50	12.68	NA	25.68	12.32	25.90	25.90	NA
Average	High	Medium	Medium	Weibull	21.56	14.47	NA	26.13	14.79	26.21	26.21	NA
Average	Low	High	High	Gompertz	1.54	-1.44	0.23	2.05	2.01	4.70	1.48	1.73
Average	Low	High	High	Log-logistic	1.41	-1.45	0.01	1.69	1.56	4.79	1.52	1.71

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Average	Low	High	High	Lognormal	1.38	-1.44	-0.02	1.66	1.52	4.77	1.51	1.68
Average	Low	High	High	Weibull	1.47	-1.44	0.10	1.87	1.78	4.80	1.49	1.71
Average	Low	High	Low	Gompertz	3.36	-3.70	-1.99	2.00	2.81	18.44	2.62	NA
Average	Low	High	Low	Log-logistic	2.87	-3.76	-2.55	1.13	1.79	17.59	3.02	NA
Average	Low	High	Low	Lognormal	2.69	-3.73	-2.61	0.99	1.60	17.18	NA	NA
Average	Low	High	Low	Weibull	3.03	-3.78	-2.37	1.47	2.25	17.90	2.70	NA
Average	Low	High	Medium	Gompertz	2.15	-2.91	-1.05	2.10	2.26	11.12	1.39	NA
Average	Low	High	Medium	Log-logistic	1.93	-2.92	-1.43	1.58	1.58	11.16	1.61	NA
Average	Low	High	Medium	Lognormal	1.93	-2.93	-1.45	1.55	1.54	11.20	1.66	NA
Average	Low	High	Medium	Weibull	2.04	-2.93	-1.29	1.81	1.89	11.28	1.49	NA
Average	Low	Low	High	Gompertz	0.18	-0.02	-0.03	0.77	0.41	-0.03	-0.03	NA
Average	Low	Low	High	Log-logistic	-0.07	-0.10	-0.10	0.01	-0.07	-0.09	-0.08	NA
Average	Low	Low	High	Lognormal	-0.02	-0.04	-0.05	0.08	-0.02	-0.05	-0.02	NA
Average	Low	Low	High	Weibull	0.07	-0.02	-0.04	0.45	0.19	-0.04	-0.04	-0.03
Average	Low	Low	Low	Gompertz	2.59	0.29	0.31	3.05	4.33	4.97	NA	NA
Average	Low	Low	Low	Log-logistic	-0.08	0.62	-0.66	0.09	0.36	-0.81	NA	NA
Average	Low	Low	Low	Lognormal	-0.15	0.69	-0.65	-0.05	0.08	-0.81	NA	NA
Average	Low	Low	Low	Weibull	1.37	0.82	0.07	2.06	3.13	0.76	NA	NA
Average	Low	Low	Medium	Gompertz	0.45	0.10	0.05	1.36	1.19	0.00	0.02	NA
Average	Low	Low	Medium	Log-logistic	-0.21	-0.17	-0.36	0.02	-0.12	-0.40	-0.23	NA
Average	Low	Low	Medium	Lognormal	-0.07	-0.03	-0.24	0.13	-0.02	-0.28	0.01	NA
Average	Low	Low	Medium	Weibull	0.23	0.15	-0.03	0.79	0.59	-0.11	-0.02	NA
Average	Low	Medium	High	Gompertz	0.18	-0.04	-0.03	0.86	0.52	0.00	-0.02	0.00
Average	Low	Medium	High	Log-logistic	-0.04	-0.13	-0.14	0.11	-0.03	-0.05	-0.05	0.04
Average	Low	Medium	High	Lognormal	0.00	-0.07	-0.08	0.17	0.01	-0.03	0.01	NA
Average	Low	Medium	High	Weibull	0.10	-0.03	-0.04	0.54	0.26	-0.03	-0.03	-0.01
Average	Low	Medium	Low	Gompertz	2.35	-0.12	-0.02	2.67	3.77	5.46	NA	NA
Average	Low	Medium	Low	Log-logistic	-0.28	0.09	-0.85	0.09	0.32	-1.06	NA	NA
Average	Low	Medium	Low	Lognormal	-0.31	0.18	-0.83	-0.02	0.11	-1.01	NA	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Average	Low	Medium	Low	Weibull	0.95	0.27	-0.29	1.63	2.48	0.64	NA	NA
Average	Low	Medium	Medium	Gompertz	0.43	-0.12	-0.10	1.37	1.22	0.24	-0.04	NA
Average	Low	Medium	Medium	Log-logistic	-0.21	-0.28	-0.46	0.15	-0.04	-0.47	-0.17	NA
Average	Low	Medium	Medium	Lognormal	-0.11	-0.19	-0.37	0.22	0.01	-0.39	0.07	NA
Average	Low	Medium	Medium	Weibull	0.18	-0.04	-0.17	0.87	0.67	-0.21	-0.06	NA
Average	Medium	High	High	Gompertz	3.72	-2.32	2.00	4.13	4.62	8.98	4.06	4.55
Average	Medium	High	High	Log-logistic	3.76	-2.45	2.16	4.19	4.64	8.94	4.23	4.64
Average	Medium	High	High	Lognormal	3.57	-2.46	1.94	3.92	4.26	8.68	4.20	4.45
Average	Medium	High	High	Weibull	3.78	-2.30	2.42	4.59	5.19	8.83	3.95	NA
Average	Medium	High	Low	Gompertz	3.80	-7.22	-2.91	2.53	5.43	21.16	NA	NA
Average	Medium	High	Low	Log-logistic	4.02	-7.33	-2.81	2.55	5.15	21.31	5.25	NA
Average	Medium	High	Low	Lognormal	3.28	-7.20	-3.33	1.93	4.16	20.86	NA	NA
Average	Medium	High	Low	Weibull	4.97	-7.06	-1.39	4.05	6.92	22.36	NA	NA
Average	Medium	High	Medium	Gompertz	3.65	-5.60	-0.74	3.56	4.95	15.98	3.77	NA
Average	Medium	High	Medium	Log-logistic	3.71	-5.72	-0.50	3.66	4.90	15.82	4.10	NA
Average	Medium	High	Medium	Lognormal	3.77	-5.60	-0.81	3.27	4.27	15.48	4.36	5.43
Average	Medium	High	Medium	Weibull	4.17	-5.46	0.35	4.49	5.96	16.03	3.65	NA
Average	Medium	Low	High	Gompertz	1.16	-0.40	-0.03	2.57	2.82	1.51	0.47	NA
Average	Medium	Low	High	Log-logistic	1.52	-1.21	-0.13	2.27	2.55	4.52	1.15	NA
Average	Medium	Low	High	Lognormal	1.80	-1.28	-0.12	2.02	2.11	5.36	2.10	2.38
Average	Medium	Low	High	Weibull	2.66	-0.36	1.24	4.17	4.74	4.94	1.21	NA
Average	Medium	Low	Low	Gompertz	10.64	-0.83	0.20	5.34	10.42	38.08	NA	NA
Average	Medium	Low	Low	Log-logistic	5.65	-3.31	-2.14	2.72	6.84	24.15	NA	NA
Average	Medium	Low	Low	Lognormal	1.70	-3.73	-4.02	-0.11	2.48	13.90	NA	NA
Average	Medium	Low	Low	Weibull	9.63	-3.06	2.05	7.58	12.18	29.38	NA	NA
Average	Medium	Low	Medium	Gompertz	2.10	-1.09	-0.73	3.91	6.10	3.67	0.71	NA
Average	Medium	Low	Medium	Log-logistic	1.43	-2.83	-1.95	2.35	3.90	6.88	0.20	NA
Average	Medium	Low	Medium	Lognormal	1.27	-3.09	-2.46	1.38	2.19	8.35	NA	NA
Average	Medium	Low	Medium	Weibull	5.18	-1.93	1.37	6.22	8.60	15.15	1.68	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Average	Medium	Medium	High	Gompertz	1.46	-0.62	0.05	2.65	2.96	2.84	0.85	NA
Average	Medium	Medium	High	Log-logistic	1.79	-1.31	0.04	2.45	2.76	5.30	1.48	NA
Average	Medium	Medium	High	Lognormal	1.98	-1.38	0.01	2.18	2.30	5.88	2.30	2.59
Average	Medium	Medium	High	Weibull	2.71	-0.61	1.20	4.07	4.66	5.57	1.40	NA
Average	Medium	Medium	Low	Gompertz	9.06	-1.84	-0.73	4.28	9.12	34.48	NA	NA
Average	Medium	Medium	Low	Log-logistic	4.65	-3.83	-2.50	2.43	6.37	20.80	NA	NA
Average	Medium	Medium	Low	Lognormal	1.37	-4.14	-4.10	-0.01	2.59	12.49	NA	NA
Average	Medium	Medium	Low	Weibull	8.67	-3.56	1.39	6.96	11.35	27.24	NA	NA
Average	Medium	Medium	Medium	Gompertz	2.02	-1.74	-1.08	3.52	5.55	5.23	0.65	NA
Average	Medium	Medium	Medium	Log-logistic	1.78	-3.15	-1.96	2.37	3.87	8.97	0.59	NA
Average	Medium	Medium	Medium	Lognormal	1.76	-3.37	-2.45	1.48	2.31	9.97	2.62	NA
Average	Medium	Medium	Medium	Weibull	5.48	-2.37	0.93	5.73	7.97	15.16	NA	NA
Old	High	High	High	Gompertz	12.06	10.18	9.13	11.97	8.98	16.14	15.96	NA
Old	High	High	High	Log-logistic	12.95	11.43	10.94	13.83	9.90	15.80	15.80	NA
Old	High	High	High	Lognormal	13.85	11.98	11.58	14.45	NA	15.63	15.63	NA
Old	High	High	High	Weibull	12.66	11.07	10.22	13.18	9.56	15.97	15.95	NA
Old	High	High	Medium	Gompertz	9.09	2.97	3.73	8.59	6.69	16.27	16.27	NA
Old	High	High	Medium	Log-logistic	9.34	3.34	4.20	9.90	6.99	15.81	15.81	NA
Old	High	High	Medium	Lognormal	9.52	3.49	4.54	10.60	7.20	15.64	15.64	NA
Old	High	High	Medium	Weibull	9.31	3.20	4.16	9.48	6.99	16.01	16.01	NA
Old	High	Low	High	Gompertz	14.96	14.96	14.96	14.96	NA	14.96	14.96	NA
Old	High	Low	High	Log-logistic	15.04	15.04	15.04	15.04	NA	15.04	15.04	NA
Old	High	Low	High	Lognormal	15.12	15.12	15.12	15.12	NA	15.12	15.12	NA
Old	High	Low	High	Weibull	15.04	15.04	15.04	15.04	NA	15.04	15.04	NA
Old	High	Low	Medium	Gompertz	13.33	14.12	7.58	14.98	NA	14.98	14.98	NA
Old	High	Low	Medium	Log-logistic	13.27	14.04	7.17	15.04	NA	15.04	15.04	NA
Old	High	Low	Medium	Lognormal	13.19	13.84	6.68	15.13	NA	15.14	15.14	NA
Old	High	Low	Medium	Weibull	13.33	14.01	7.50	15.04	NA	15.05	15.05	NA
Old	High	Medium	High	Gompertz	14.98	14.98	14.98	14.98	NA	14.98	14.98	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Old	High	Medium	High	Log-logistic	15.00	14.99	14.99	15.00	NA	15.00	15.00	NA
Old	High	Medium	High	Lognormal	15.05	15.04	15.03	15.06	NA	15.06	15.06	NA
Old	High	Medium	High	Weibull	15.02	15.01	15.01	15.02	NA	15.02	15.02	NA
Old	High	Medium	Medium	Gompertz	12.76	11.88	6.98	14.96	NA	14.97	14.97	NA
Old	High	Medium	Medium	Log-logistic	12.67	11.62	6.73	14.99	NA	15.00	15.00	NA
Old	High	Medium	Medium	Lognormal	12.06	10.96	6.04	15.06	10.13	15.07	15.07	NA
Old	High	Medium	Medium	Weibull	12.63	11.43	6.68	15.00	NA	15.02	15.02	NA
Old	Low	High	High	Gompertz	1.73	0.16	1.21	2.00	2.11	3.10	1.78	NA
Old	Low	High	High	Log-logistic	1.66	0.21	1.11	1.80	1.85	2.98	1.79	1.87
Old	Low	High	High	Lognormal	1.66	0.22	1.11	1.80	1.85	2.97	1.80	1.86
Old	Low	High	High	Weibull	1.69	0.21	1.18	1.92	2.00	3.02	1.79	NA
Old	Low	High	Low	Gompertz	2.06	-1.70	-0.34	1.79	2.30	8.26	NA	NA
Old	Low	High	Low	Log-logistic	1.69	-1.68	-0.73	1.31	1.74	7.64	1.87	NA
Old	Low	High	Low	Lognormal	1.69	-1.68	-0.71	1.34	1.77	7.74	NA	NA
Old	Low	High	Low	Weibull	1.92	-1.70	-0.54	1.58	2.08	8.11	1.98	NA
Old	Low	High	Medium	Gompertz	1.55	-1.00	0.31	1.78	1.95	4.90	1.37	NA
Old	Low	High	Medium	Log-logistic	1.41	-0.97	0.08	1.50	1.59	4.82	1.44	NA
Old	Low	High	Medium	Lognormal	1.42	-0.96	0.08	1.50	1.58	4.85	1.47	NA
Old	Low	High	Medium	Weibull	1.50	-0.98	0.19	1.65	1.78	4.93	1.43	NA
Old	Low	Low	High	Gompertz	0.24	0.10	0.08	0.76	0.51	0.07	0.08	0.08
Old	Low	Low	High	Log-logistic	0.10	0.09	0.06	0.17	0.11	0.08	0.10	NA
Old	Low	Low	High	Lognormal	0.11	0.10	0.07	0.20	0.11	0.08	0.11	NA
Old	Low	Low	High	Weibull	0.16	0.11	0.06	0.50	0.29	0.06	0.06	0.06
Old	Low	Low	Low	Gompertz	1.78	0.38	0.38	2.24	3.02	2.85	NA	NA
Old	Low	Low	Low	Log-logistic	0.59	0.81	-0.43	0.27	0.47	1.81	NA	NA
Old	Low	Low	Low	Lognormal	0.39	0.80	-0.48	0.10	0.22	1.30	NA	NA
Old	Low	Low	Low	Weibull	1.81	0.91	0.19	1.71	2.41	3.85	NA	NA
Old	Low	Low	Medium	Gompertz	0.46	0.20	0.15	1.15	1.05	0.10	0.12	NA
Old	Low	Low	Medium	Log-logistic	-0.02	0.04	-0.16	0.18	0.07	-0.20	NA	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Old	Low	Low	Medium	Lognormal	0.06	0.11	-0.11	0.23	0.11	-0.14	0.13	NA
Old	Low	Low	Medium	Weibull	0.29	0.26	0.06	0.75	0.60	-0.01	0.08	NA
Old	Low	Medium	High	Gompertz	0.34	0.14	0.16	0.86	0.67	0.19	0.17	0.19
Old	Low	Medium	High	Log-logistic	0.19	0.13	0.10	0.32	0.23	0.18	0.20	NA
Old	Low	Medium	High	Lognormal	0.20	0.14	0.11	0.33	0.22	0.17	0.22	NA
Old	Low	Medium	High	Weibull	0.26	0.15	0.13	0.62	0.44	0.15	0.15	0.17
Old	Low	Medium	Low	Gompertz	1.45	0.05	0.14	1.95	2.63	2.46	NA	NA
Old	Low	Medium	Low	Log-logistic	0.13	0.36	-0.53	0.32	0.52	0.02	NA	NA
Old	Low	Medium	Low	Lognormal	0.06	0.38	-0.57	0.20	0.34	-0.03	NA	NA
Old	Low	Medium	Low	Weibull	0.98	0.40	-0.11	1.35	1.91	1.32	NA	NA
Old	Low	Medium	Medium	Gompertz	0.45	0.05	0.07	1.12	1.06	0.25	0.12	NA
Old	Low	Medium	Medium	Log-logistic	0.02	-0.02	-0.19	0.31	0.19	-0.21	NA	NA
Old	Low	Medium	Medium	Lognormal	0.07	0.03	-0.16	0.34	0.20	-0.17	0.21	NA
Old	Low	Medium	Medium	Weibull	0.27	0.12	-0.01	0.79	0.67	-0.05	0.09	NA
Old	Medium	High	High	Gompertz	3.70	1.40	3.19	3.87	4.22	5.32	3.88	4.04
Old	Medium	High	High	Log-logistic	3.66	1.22	3.16	3.82	4.16	5.41	3.87	4.01
Old	Medium	High	High	Lognormal	3.48	1.12	2.99	3.63	3.91	5.31	NA	3.90
Old	Medium	High	High	Weibull	3.74	1.22	3.29	4.03	4.46	5.38	3.81	3.98
Old	Medium	High	Low	Gompertz	3.27	-2.42	0.20	2.72	4.12	11.73	NA	NA
Old	Medium	High	Low	Log-logistic	3.02	-2.58	0.22	2.70	3.92	10.83	NA	NA
Old	Medium	High	Low	Lognormal	3.08	-2.58	0.01	2.48	3.58	10.45	4.53	NA
Old	Medium	High	Low	Weibull	3.52	-2.53	0.94	3.37	4.73	11.09	NA	NA
Old	Medium	High	Medium	Gompertz	2.85	-1.14	1.37	3.07	3.72	7.15	2.90	NA
Old	Medium	High	Medium	Log-logistic	2.82	-1.28	1.40	3.05	3.64	7.16	2.97	NA
Old	Medium	High	Medium	Lognormal	2.69	-1.32	1.22	2.84	3.33	7.01	3.05	NA
Old	Medium	High	Medium	Weibull	3.00	-1.24	1.76	3.40	4.12	7.24	2.74	NA
Old	Medium	Low	High	Gompertz	2.33	1.67	1.86	2.94	3.17	2.27	2.07	NA
Old	Medium	Low	High	Log-logistic	2.42	1.33	1.89	2.81	3.07	3.05	2.37	NA
Old	Medium	Low	High	Lognormal	2.35	1.18	1.77	2.54	2.73	3.25	2.62	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Old	Medium	Low	High	Weibull	2.92	1.58	2.49	3.65	4.06	3.33	2.42	NA
Old	Medium	Low	Low	Gompertz	9.06	1.41	1.50	5.57	6.23	19.88	19.77	NA
Old	Medium	Low	Low	Log-logistic	5.47	-0.12	0.48	3.17	4.98	18.85	NA	NA
Old	Medium	Low	Low	Lognormal	3.78	-0.62	-1.06	1.34	2.77	16.49	NA	NA
Old	Medium	Low	Low	Weibull	6.63	-0.45	2.61	5.12	7.04	18.85	NA	NA
Old	Medium	Low	Medium	Gompertz	3.13	1.23	1.42	3.57	4.56	4.69	3.31	NA
Old	Medium	Low	Medium	Log-logistic	2.22	0.23	0.85	2.88	3.65	3.81	1.91	NA
Old	Medium	Low	Medium	Lognormal	2.05	-0.06	0.38	2.28	2.81	4.13	2.79	NA
Old	Medium	Low	Medium	Weibull	3.96	0.49	2.49	4.42	5.48	6.91	NA	NA
Old	Medium	Medium	High	Gompertz	2.54	1.67	2.05	3.06	3.33	2.75	2.36	NA
Old	Medium	Medium	High	Log-logistic	2.58	1.35	2.04	2.93	3.21	3.38	2.57	NA
Old	Medium	Medium	High	Lognormal	2.48	1.19	1.90	2.66	2.86	3.52	2.78	NA
Old	Medium	Medium	High	Weibull	3.00	1.56	2.56	3.67	4.08	3.55	2.58	NA
Old	Medium	Medium	Low	Gompertz	6.25	0.62	0.97	4.31	5.66	19.68	NA	NA
Old	Medium	Medium	Low	Log-logistic	4.91	-0.60	0.13	2.75	4.55	17.72	NA	NA
Old	Medium	Medium	Low	Lognormal	3.45	-0.94	-0.99	1.42	2.85	14.93	NA	NA
Old	Medium	Medium	Low	Weibull	6.14	-0.75	2.14	4.65	6.60	18.08	NA	NA
Old	Medium	Medium	Medium	Gompertz	2.53	0.73	1.11	3.22	4.12	3.50	NA	NA
Old	Medium	Medium	Medium	Log-logistic	2.21	-0.04	0.73	2.72	3.45	4.21	NA	NA
Old	Medium	Medium	Medium	Lognormal	1.93	-0.28	0.35	2.23	2.74	4.61	NA	NA
Old	Medium	Medium	Medium	Weibull	3.36	0.16	2.11	4.05	5.04	6.64	2.17	NA
Young	High	High	High	Gompertz	9.07	0.47	5.72	8.79	11.31	16.21	11.90	NA
Young	High	High	High	Log-logistic	13.38	2.13	8.36	10.61	13.00	18.18	28.01	NA
Young	High	High	High	Lognormal	14.95	2.97	9.87	11.67	13.85	19.12	32.21	NA
Young	High	High	High	Weibull	11.74	1.42	7.36	9.94	12.41	17.43	21.18	12.45
Young	High	High	Medium	Gompertz	14.93	-5.07	-0.42	5.66	12.15	38.68	38.55	NA
Young	High	High	Medium	Log-logistic	14.97	-4.63	2.32	7.29	13.31	35.75	35.75	NA
Young	High	High	Medium	Lognormal	14.70	-4.83	3.26	7.79	13.35	34.32	34.32	NA
Young	High	High	Medium	Weibull	15.04	-4.81	1.31	6.72	12.96	37.05	37.03	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Young	High	Low	High	Gompertz	22.46	17.18	14.86	26.50	19.80	23.83	32.58	NA
Young	High	Low	High	Log-logistic	19.15	14.38	10.77	18.77	16.67	22.60	31.70	NA
Young	High	Low	High	Lognormal	17.38	12.00	8.54	12.62	14.67	25.16	31.31	NA
Young	High	Low	High	Weibull	19.71	14.80	11.93	19.69	17.49	22.34	32.00	NA
Young	High	Low	Medium	Gompertz	29.04	18.68	NA	32.31	NA	32.59	32.59	NA
Young	High	Low	Medium	Log-logistic	25.60	16.96	NA	31.30	16.32	31.72	31.72	NA
Young	High	Low	Medium	Lognormal	23.73	14.15	NA	30.20	11.67	31.31	31.31	NA
Young	High	Low	Medium	Weibull	26.06	16.91	NA	31.69	17.71	32.01	32.01	NA
Young	High	Medium	High	Gompertz	21.13	16.42	13.54	21.21	19.27	22.30	34.03	NA
Young	High	Medium	High	Log-logistic	18.05	13.18	10.04	15.36	16.12	21.37	32.22	NA
Young	High	Medium	High	Lognormal	17.05	11.11	8.64	12.18	14.57	24.14	31.64	NA
Young	High	Medium	High	Weibull	18.68	13.81	10.89	16.36	16.88	21.41	32.75	NA
Young	High	Medium	Medium	Gompertz	34.02	NA	NA	NA	NA	NA	34.02	NA
Young	High	Medium	Medium	Log-logistic	25.37	15.24	NA	31.25	15.94	32.22	32.22	NA
Young	High	Medium	Medium	Lognormal	31.64	NA	NA	NA	NA	NA	31.64	NA
Young	High	Medium	Medium	Weibull	28.28	15.77	NA	31.88	NA	32.74	32.74	NA
Young	Low	High	High	Gompertz	1.43	-2.32	-0.46	2.03	1.88	6.05	1.22	1.59
Young	Low	High	High	Log-logistic	1.22	-2.36	-0.78	1.52	1.26	6.15	1.23	1.50
Young	Low	High	High	Lognormal	1.24	-2.34	-0.77	1.54	1.27	6.21	1.27	1.51
Young	Low	High	High	Weibull	1.35	-2.34	-0.63	1.79	1.58	6.24	1.24	1.55
Young	Low	High	Low	Gompertz	4.08	-4.68	-2.93	1.96	2.88	24.51	2.74	NA
Young	Low	High	Low	Log-logistic	3.47	-4.72	-3.51	0.84	1.51	23.38	3.32	NA
Young	Low	High	Low	Lognormal	3.50	-4.70	-3.52	0.82	1.48	23.40	3.49	NA
Young	Low	High	Low	Weibull	3.83	-4.78	-3.31	1.36	2.23	24.24	3.22	NA
Young	Low	High	Medium	Gompertz	2.52	-3.85	-1.87	2.16	2.27	15.11	1.27	NA
Young	Low	High	Medium	Log-logistic	2.19	-3.90	-2.33	1.43	1.34	15.11	1.47	NA
Young	Low	High	Medium	Lognormal	2.22	-3.87	-2.33	1.44	1.34	15.15	1.56	NA
Young	Low	High	Medium	Weibull	2.36	-3.88	-2.14	1.80	1.83	15.28	1.30	NA
Young	Low	Low	High	Gompertz	0.18	-0.04	-0.04	0.80	0.41	-0.04	-0.04	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Young	Low	Low	High	Log-logistic	-0.11	-0.13	-0.14	-0.02	-0.10	-0.12	-0.12	NA
Young	Low	Low	High	Lognormal	-0.03	-0.05	-0.06	0.08	-0.04	-0.06	-0.04	NA
Young	Low	Low	High	Weibull	0.06	-0.03	-0.05	0.46	0.19	-0.05	-0.05	-0.04
Young	Low	Low	Low	Gompertz	3.17	0.27	0.32	3.42	4.88	6.99	NA	NA
Young	Low	Low	Low	Log-logistic	-0.10	0.62	-0.68	0.10	0.39	-0.92	NA	NA
Young	Low	Low	Low	Lognormal	-0.19	0.67	-0.68	-0.09	0.04	-0.88	NA	NA
Young	Low	Low	Low	Weibull	1.46	0.81	0.05	2.19	3.39	0.89	NA	NA
Young	Low	Low	Medium	Gompertz	0.47	0.09	0.04	1.45	1.25	-0.01	0.00	NA
Young	Low	Low	Medium	Log-logistic	-0.24	-0.20	-0.39	-0.01	-0.15	-0.43	-0.27	NA
Young	Low	Low	Medium	Lognormal	-0.09	-0.05	-0.26	0.12	-0.03	-0.30	0.00	NA
Young	Low	Low	Medium	Weibull	0.23	0.14	-0.04	0.82	0.61	-0.12	-0.03	NA
Young	Low	Medium	High	Gompertz	0.18	-0.06	-0.05	0.91	0.53	-0.02	-0.05	-0.02
Young	Low	Medium	High	Log-logistic	-0.08	-0.18	-0.19	0.08	-0.08	-0.09	-0.11	0.00
Young	Low	Medium	High	Lognormal	-0.03	-0.10	-0.11	0.15	-0.03	-0.06	-0.02	0.00
Young	Low	Medium	High	Weibull	0.10	-0.05	-0.06	0.55	0.25	-0.05	-0.05	NA
Young	Low	Medium	Low	Gompertz	2.80	-0.15	-0.06	2.93	4.16	7.11	NA	NA
Young	Low	Medium	Low	Log-logistic	-0.33	0.03	-0.92	0.05	0.28	-1.09	NA	NA
Young	Low	Medium	Low	Lognormal	-0.36	0.14	-0.88	-0.05	0.07	-1.06	NA	NA
Young	Low	Medium	Low	Weibull	1.13	0.24	-0.30	1.79	2.75	1.15	NA	NA
Young	Low	Medium	Medium	Gompertz	0.47	-0.14	-0.12	1.50	1.30	0.30	-0.06	NA
Young	Low	Medium	Medium	Log-logistic	-0.27	-0.35	-0.52	0.11	-0.10	-0.54	-0.23	NA
Young	Low	Medium	Medium	Lognormal	-0.13	-0.22	-0.41	0.21	-0.02	-0.42	0.05	NA
Young	Low	Medium	Medium	Weibull	0.17	-0.06	-0.19	0.91	0.68	-0.23	-0.08	NA
Young	Medium	High	High	Gompertz	3.35	-4.89	0.59	3.82	4.41	11.42	3.68	4.43
Young	Medium	High	High	Log-logistic	3.24	-5.02	0.85	3.94	4.47	11.27	3.93	NA
Young	Medium	High	High	Lognormal	3.09	-4.91	0.66	3.68	4.06	11.00	4.07	NA
Young	Medium	High	High	Weibull	3.53	-4.66	1.29	4.54	5.26	11.20	3.58	NA
Young	Medium	High	Low	Gompertz	3.87	-10.05	-5.30	1.71	5.51	27.48	NA	NA
Young	Medium	High	Low	Log-logistic	3.98	-10.07	-5.16	1.84	5.20	28.07	NA	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Young	Medium	High	Low	Lognormal	3.90	-9.82	-5.62	1.12	3.93	27.57	6.21	NA
Young	Medium	High	Low	Weibull	5.47	-9.64	-3.35	3.81	7.51	29.03	NA	NA
Young	Medium	High	Medium	Gompertz	3.55	-8.32	-2.85	3.05	4.82	21.14	3.49	NA
Young	Medium	High	Medium	Log-logistic	3.65	-8.43	-2.50	3.25	4.84	20.86	3.89	NA
Young	Medium	High	Medium	Lognormal	3.47	-8.24	-2.79	2.84	4.11	20.50	4.40	NA
Young	Medium	High	Medium	Weibull	4.39	-7.95	-1.29	4.50	6.41	21.32	3.37	NA
Young	Medium	Low	High	Gompertz	0.62	-1.35	-0.94	2.39	2.64	1.29	-0.35	NA
Young	Medium	Low	High	Log-logistic	1.01	-2.60	-1.40	1.85	2.12	5.80	0.28	NA
Young	Medium	Low	High	Lognormal	1.49	-2.57	-1.27	1.68	1.66	7.09	1.71	2.10
Young	Medium	Low	High	Weibull	2.84	-1.09	0.70	4.70	5.40	6.61	0.72	NA
Young	Medium	Low	Low	Gompertz	11.83	-1.82	-0.52	5.74	12.36	43.41	NA	NA
Young	Medium	Low	Low	Log-logistic	4.90	-4.72	-3.42	2.54	7.66	22.44	NA	NA
Young	Medium	Low	Low	Lognormal	2.19	-5.18	-5.52	-1.03	1.99	11.35	11.51	NA
Young	Medium	Low	Low	Weibull	11.36	-3.83	2.00	9.26	15.30	34.10	NA	NA
Young	Medium	Low	Medium	Gompertz	1.97	-2.07	-1.66	4.09	6.90	4.68	-0.09	NA
Young	Medium	Low	Medium	Log-logistic	1.09	-4.28	-3.36	1.96	3.86	9.11	-0.75	NA
Young	Medium	Low	Medium	Lognormal	1.39	-4.47	-3.79	0.88	1.76	11.53	2.45	NA
Young	Medium	Low	Medium	Weibull	6.42	-2.71	0.87	7.34	10.46	20.97	1.57	NA
Young	Medium	Medium	High	Gompertz	0.98	-1.75	-0.99	2.45	2.77	3.36	0.02	NA
Young	Medium	Medium	High	Log-logistic	1.37	-2.78	-1.23	2.08	2.40	7.01	0.72	NA
Young	Medium	Medium	High	Lognormal	1.69	-2.78	-1.19	1.84	1.87	7.79	1.93	2.34
Young	Medium	Medium	High	Weibull	2.82	-1.55	0.52	4.47	5.16	7.47	0.83	NA
Young	Medium	Medium	Low	Gompertz	9.96	-3.02	-1.48	4.81	11.08	38.39	NA	NA
Young	Medium	Medium	Low	Log-logistic	4.14	-5.39	-3.97	2.04	6.95	21.09	NA	NA
Young	Medium	Medium	Low	Lognormal	2.18	-5.66	-5.57	-0.81	2.20	13.03	9.86	NA
Young	Medium	Medium	Low	Weibull	10.25	-4.61	0.99	8.24	13.98	32.63	NA	NA
Young	Medium	Medium	Medium	Gompertz	1.92	-2.91	-2.20	3.52	6.09	7.00	0.02	NA
Young	Medium	Medium	Medium	Log-logistic	1.60	-4.72	-3.47	1.95	3.79	12.28	-0.22	NA
Young	Medium	Medium	Medium	Lognormal	1.75	-4.81	-3.83	1.02	1.95	13.64	2.50	NA
Young	Medium	Medium	Medium	Weibull	5.88	-3.39	0.25	6.64	9.56	20.96	1.28	NA

Table 96:

Mean bias of overall mean survival – external additive hazards models

Scenario characteristics						*Distribution used for extrapolation*						
Age	**Survival**	**Heterogeneity**	**Information**	**Distribution**	**Mean of method**	**Exponential**	**Weibull**	**Log-logistic**	**Lognormal**	**Gompertz**	**Gen. Gamma**	**Gen. F**
Average	High	High	High	Gompertz	-2.16	-3.55	-2.70	-1.85	-1.21	-1.28	-2.35	NA
Average	High	High	High	Log-logistic	-2.37	-3.80	-2.81	-2.13	-1.50	-1.40	-2.58	NA
Average	High	High	High	Lognormal	-2.49	-3.94	-2.87	-2.26	-1.66	-1.47	-2.74	NA
Average	High	High	High	Weibull	-2.28	-3.70	-2.76	-2.00	-1.36	-1.35	-2.50	NA
Average	High	High	Medium	Gompertz	-2.25	-5.24	-3.94	-2.03	0.11	-0.12	NA	NA
Average	High	High	Medium	Log-logistic	-2.47	-5.70	-3.91	-2.35	-0.35	-0.03	NA	NA
Average	High	High	Medium	Lognormal	-2.51	-6.03	-3.85	-2.47	-0.63	0.44	NA	NA
Average	High	High	Medium	Weibull	-2.11	-5.47	-3.78	-2.10	-0.06	0.17	-1.46	NA
Average	High	Low	High	Gompertz	-2.05	-1.35	-2.14	-1.55	-0.60	-4.20	-2.48	NA
Average	High	Low	High	Log-logistic	-2.63	-1.83	-2.79	-2.11	-1.12	-4.86	-3.04	NA
Average	High	Low	High	Lognormal	-2.93	-2.32	-3.28	-2.56	-1.63	-4.79	-3.00	NA
Average	High	Low	High	Weibull	-2.44	-1.72	-2.59	-1.93	-0.96	-4.56	-2.88	NA
Average	High	Low	Medium	Gompertz	-1.43	-0.25	NA	-1.15	0.97	-5.29	NA	NA
Average	High	Low	Medium	Log-logistic	-2.15	-0.46	NA	-1.88	0.47	-6.72	NA	NA
Average	High	Low	Medium	Lognormal	-3.55	-0.99	NA	-3.40	-0.61	-9.21	NA	NA
Average	High	Low	Medium	Weibull	-1.99	-0.50	NA	-1.75	0.56	-6.28	NA	NA
Average	High	Medium	High	Gompertz	-1.96	-1.25	-2.06	-1.37	-0.37	-4.14	-2.56	NA
Average	High	Medium	High	Log-logistic	-2.80	-2.15	-3.02	-2.29	-1.31	-4.78	-3.26	NA
Average	High	Medium	High	Lognormal	-3.03	-2.64	-3.42	-2.69	-1.79	-4.58	-3.11	-2.98
Average	High	Medium	High	Weibull	-2.59	-1.95	-2.78	-2.06	-1.09	-4.55	-3.10	NA
Average	High	Medium	Medium	Gompertz	-1.10	-0.09	NA	-0.74	1.44	-5.02	NA	NA
Average	High	Medium	Medium	Log-logistic	-2.44	-0.86	NA	-2.15	0.31	-7.05	NA	NA
Average	High	Medium	Medium	Lognormal	-3.74	-1.63	NA	-3.55	-0.78	-9.02	NA	NA
Average	High	Medium	Medium	Weibull	-1.97	-0.72	NA	-1.72	0.65	-6.07	NA	NA
Average	Low	High	High	Gompertz	-0.25	-1.62	-0.67	0.13	0.10	0.68	-0.22	-0.15
Average	Low	High	High	Log-logistic	-0.36	-1.65	-0.82	-0.11	-0.17	0.72	-0.27	-0.23

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Average	Low	High	High	Lognormal	-0.36	-1.64	-0.83	-0.12	-0.18	0.74	-0.27	-0.23
Average	Low	High	High	Weibull	-0.30	-1.64	-0.76	0.01	-0.04	0.75	-0.25	-0.20
Average	Low	High	Low	Gompertz	0.85	-3.72	-2.24	0.27	0.83	9.10	NA	NA
Average	Low	High	Low	Log-logistic	0.38	-3.79	-2.71	-0.31	0.16	8.53	NA	NA
Average	Low	High	Low	Lognormal	0.29	-3.76	-2.76	-0.40	0.05	8.29	NA	NA
Average	Low	High	Low	Weibull	0.54	-3.80	-2.57	-0.10	0.45	8.71	0.53	NA
Average	Low	High	Medium	Gompertz	0.14	-2.97	-1.48	0.30	0.44	4.67	-0.11	NA
Average	Low	High	Medium	Log-logistic	-0.02	-2.99	-1.77	-0.03	0.01	4.67	-0.04	NA
Average	Low	High	Medium	Lognormal	-0.03	-2.99	-1.79	-0.06	-0.02	4.67	-0.02	NA
Average	Low	High	Medium	Weibull	0.04	-2.99	-1.67	0.10	0.19	4.71	-0.09	NA
Average	Low	Low	High	Gompertz	0.08	-0.06	-0.07	0.56	0.36	-0.07	-0.06	-0.06
Average	Low	Low	High	Log-logistic	-0.11	-0.14	-0.14	-0.06	-0.11	-0.14	-0.14	-0.07
Average	Low	Low	High	Lognormal	-0.06	-0.08	-0.08	0.01	-0.06	-0.09	-0.07	NA
Average	Low	Low	High	Weibull	0.03	-0.06	-0.07	0.35	0.18	-0.07	-0.07	-0.07
Average	Low	Low	Low	Gompertz	1.66	0.23	0.24	2.24	3.20	2.74	1.28	NA
Average	Low	Low	Low	Log-logistic	-0.04	0.54	-0.67	0.02	0.36	-0.91	0.39	NA
Average	Low	Low	Low	Lognormal	-0.19	0.61	-0.66	-0.10	0.07	-0.86	NA	NA
Average	Low	Low	Low	Weibull	1.09	0.74	0.03	1.60	2.51	0.18	1.47	NA
Average	Low	Low	Medium	Gompertz	0.34	0.05	0.01	1.01	1.01	-0.04	-0.02	NA
Average	Low	Low	Medium	Log-logistic	-0.24	-0.21	-0.38	-0.05	-0.13	-0.42	-0.27	NA
Average	Low	Low	Medium	Lognormal	-0.11	-0.08	-0.26	0.05	-0.05	-0.30	-0.05	NA
Average	Low	Low	Medium	Weibull	0.17	0.10	-0.06	0.62	0.56	-0.14	-0.06	NA
Average	Low	Medium	High	Gompertz	0.07	-0.09	-0.09	0.54	0.37	-0.07	-0.09	-0.07
Average	Low	Medium	High	Log-logistic	-0.13	-0.19	-0.19	-0.02	-0.11	-0.14	-0.15	-0.09
Average	Low	Medium	High	Lognormal	-0.08	-0.13	-0.13	0.03	-0.07	-0.11	-0.08	NA
Average	Low	Medium	High	Weibull	0.02	-0.09	-0.09	0.34	0.18	-0.09	-0.09	-0.07
Average	Low	Medium	Low	Gompertz	1.38	-0.16	-0.08	1.87	2.72	2.91	1.03	NA
Average	Low	Medium	Low	Log-logistic	-0.34	0.03	-0.86	-0.04	0.23	-1.08	NA	NA
Average	Low	Medium	Low	Lognormal	-0.36	0.11	-0.84	-0.11	0.06	-1.03	NA	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Average	Low	Medium	Low	Weibull	0.69	0.20	-0.32	1.19	1.92	0.07	1.06	NA
Average	Low	Medium	Medium	Gompertz	0.25	-0.16	-0.15	0.94	0.95	0.05	-0.10	NA
Average	Low	Medium	Medium	Log-logistic	-0.28	-0.33	-0.49	0.01	-0.10	-0.50	-0.25	NA
Average	Low	Medium	Medium	Lognormal	-0.18	-0.23	-0.40	0.08	-0.05	-0.42	-0.04	NA
Average	Low	Medium	Medium	Weibull	0.09	-0.09	-0.21	0.61	0.56	-0.25	-0.11	NA
Average	Medium	High	High	Gompertz	-0.85	-3.26	-1.34	-0.50	-0.34	0.88	-0.73	-0.63
Average	Medium	High	High	Log-logistic	-0.85	-3.31	-1.24	-0.45	-0.31	0.84	-0.64	NA
Average	Medium	High	High	Lognormal	-0.79	-3.23	-1.23	-0.47	-0.37	0.84	-0.57	-0.52
Average	Medium	High	High	Weibull	-0.65	-3.13	-1.02	-0.14	0.06	0.93	-0.60	NA
Average	Medium	High	Low	Gompertz	-0.51	-7.33	-4.13	-1.02	0.66	8.47	0.30	NA
Average	Medium	High	Low	Log-logistic	-0.47	-7.42	-4.02	-0.97	0.54	8.81	0.23	NA
Average	Medium	High	Low	Lognormal	-0.59	-7.28	-4.29	-1.25	0.07	8.70	0.51	NA
Average	Medium	High	Low	Weibull	0.25	-7.15	-2.99	0.04	1.68	9.57	0.32	NA
Average	Medium	High	Medium	Gompertz	-0.61	-5.87	-2.63	-0.45	0.29	5.44	-0.44	NA
Average	Medium	High	Medium	Log-logistic	-0.56	-5.94	-2.47	-0.38	0.28	5.39	-0.26	NA
Average	Medium	High	Medium	Lognormal	-0.57	-5.80	-2.55	-0.48	0.06	5.37	-0.02	NA
Average	Medium	High	Medium	Weibull	-0.17	-5.68	-1.87	0.21	0.98	5.66	-0.32	NA
Average	Medium	Low	High	Gompertz	-0.46	-1.17	-1.00	0.24	0.42	-0.48	-0.80	NA
Average	Medium	Low	High	Log-logistic	-0.69	-1.87	-1.34	-0.27	-0.10	0.26	-0.84	NA
Average	Medium	Low	High	Lognormal	-0.61	-1.88	-1.29	-0.38	-0.32	0.62	-0.43	NA
Average	Medium	Low	High	Weibull	0.16	-1.12	-0.38	0.94	1.20	0.78	-0.45	NA
Average	Medium	Low	Low	Gompertz	1.66	-1.52	-1.11	1.93	4.95	4.08	NA	NA
Average	Medium	Low	Low	Log-logistic	0.10	-3.61	-2.90	0.12	2.60	3.10	1.26	NA
Average	Medium	Low	Low	Lognormal	-1.75	-3.94	-4.26	-1.62	0.08	0.98	NA	NA
Average	Medium	Low	Low	Weibull	3.07	-3.34	0.05	3.17	5.74	9.71	NA	NA
Average	Medium	Low	Medium	Gompertz	0.05	-1.67	-1.44	1.21	2.57	0.46	-0.83	NA
Average	Medium	Low	Medium	Log-logistic	-0.77	-3.18	-2.54	-0.07	0.90	1.64	-1.40	NA
Average	Medium	Low	Medium	Lognormal	-0.77	-3.37	-2.89	-0.62	-0.09	2.41	-0.06	NA
Average	Medium	Low	Medium	Weibull	1.54	-2.35	-0.19	2.34	3.64	6.17	-0.36	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Average	Medium	Medium	High	Gompertz	-0.52	-1.42	-1.10	0.10	0.28	-0.18	-0.79	NA
Average	Medium	Medium	High	Log-logistic	-0.70	-2.01	-1.35	-0.31	-0.14	0.43	-0.81	NA
Average	Medium	Medium	High	Lognormal	-0.64	-2.01	-1.32	-0.41	-0.35	0.70	-0.47	NA
Average	Medium	Medium	High	Weibull	0.01	-1.39	-0.55	0.73	0.98	0.83	-0.53	NA
Average	Medium	Medium	Low	Gompertz	0.89	-2.38	-1.86	1.19	4.04	3.46	NA	NA
Average	Medium	Medium	Low	Log-logistic	-0.32	-4.09	-3.21	-0.15	2.22	3.64	NA	NA
Average	Medium	Medium	Low	Lognormal	-1.11	-4.33	-4.36	-1.63	0.06	1.91	1.70	NA
Average	Medium	Medium	Low	Weibull	2.65	-3.80	-0.45	2.69	5.16	9.62	NA	NA
Average	Medium	Medium	Medium	Gompertz	-0.20	-2.25	-1.81	0.81	2.07	1.05	-1.05	NA
Average	Medium	Medium	Medium	Log-logistic	-0.70	-3.48	-2.61	-0.16	0.77	2.59	-1.29	NA
Average	Medium	Medium	Medium	Lognormal	-0.72	-3.64	-2.94	-0.66	-0.12	3.14	-0.11	NA
Average	Medium	Medium	Medium	Weibull	1.23	-2.75	-0.56	1.95	3.20	6.07	-0.54	NA
Old	High	High	High	Gompertz	-2.75	-2.79	-2.79	-2.67	-2.69	-2.78	-2.80	NA
Old	High	High	High	Log-logistic	-3.00	-3.03	-3.03	-2.91	-2.93	-3.03	-3.05	NA
Old	High	High	High	Lognormal	-3.09	-3.12	-3.13	-3.01	-3.03	-3.12	-3.15	NA
Old	High	High	High	Weibull	-2.89	-2.92	-2.93	-2.81	-2.83	-2.92	-2.94	NA
Old	High	High	Medium	Gompertz	-2.48	-2.88	-2.78	-2.27	-1.85	-2.44	-2.64	NA
Old	High	High	Medium	Log-logistic	-2.66	-3.12	-2.99	-2.50	-2.09	-2.62	-2.82	-2.50
Old	High	High	Medium	Lognormal	-2.75	-3.23	-3.06	-2.59	-2.19	-2.60	-2.85	NA
Old	High	High	Medium	Weibull	-2.57	-3.02	-2.87	-2.38	-1.96	-2.49	-2.73	NA
Old	High	Low	High	Gompertz	-2.89	-2.92	-2.89	-2.74	-2.75	-3.01	-3.04	NA
Old	High	Low	High	Log-logistic	-2.86	-2.89	-2.87	-2.72	-2.72	-2.98	-3.00	NA
Old	High	Low	High	Lognormal	-2.83	-2.86	-2.84	-2.69	-2.70	-2.95	-2.95	NA
Old	High	Low	High	Weibull	-2.86	-2.88	-2.86	-2.71	-2.72	-2.98	-3.00	NA
Old	High	Low	Medium	Gompertz	-2.22	-2.13	-2.38	-2.02	-1.53	-2.94	-2.30	NA
Old	High	Low	Medium	Log-logistic	-2.23	-2.10	-2.40	-2.02	-1.52	-3.02	-2.31	NA
Old	High	Low	Medium	Lognormal	-2.27	-2.08	-2.46	-2.06	-1.54	-3.14	-2.35	NA
Old	High	Low	Medium	Weibull	-2.21	-2.10	-2.37	-2.00	-1.51	-2.95	-2.31	NA
Old	High	Medium	High	Gompertz	-3.08	-3.11	-3.09	-2.94	-2.95	-3.18	-3.19	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Old	High	Medium	High	Log-logistic	-3.08	-3.11	-3.10	-2.95	-2.96	-3.19	-3.19	NA
Old	High	Medium	High	Lognormal	-3.07	-3.10	-3.08	-2.94	-2.95	-3.17	-3.15	NA
Old	High	Medium	High	Weibull	-3.07	-3.10	-3.08	-2.94	-2.94	-3.17	-3.18	NA
Old	High	Medium	Medium	Gompertz	-2.50	-2.47	-2.68	-2.28	-1.79	-3.13	-2.63	NA
Old	High	Medium	Medium	Log-logistic	-2.54	-2.47	-2.74	-2.33	-1.83	-3.23	-2.67	NA
Old	High	Medium	Medium	Lognormal	-2.62	-2.50	-2.83	-2.39	-1.87	-3.37	-2.74	NA
Old	High	Medium	Medium	Weibull	-2.53	-2.47	-2.72	-2.31	-1.81	-3.19	-2.66	NA
Old	Low	High	High	Gompertz	-0.63	-0.66	-0.63	-0.52	-0.57	-0.74	-0.66	NA
Old	Low	High	High	Log-logistic	-0.68	-0.68	-0.68	-0.60	-0.66	-0.74	-0.71	NA
Old	Low	High	High	Lognormal	-0.68	-0.68	-0.68	-0.60	-0.66	-0.73	-0.71	NA
Old	Low	High	High	Weibull	-0.67	-0.68	-0.67	-0.57	-0.63	-0.74	-0.70	-0.70
Old	Low	High	Low	Gompertz	-0.23	-1.84	-1.04	-0.19	0.02	1.94	-0.30	NA
Old	Low	High	Low	Log-logistic	-0.41	-1.84	-1.27	-0.42	-0.24	1.63	-0.32	NA
Old	Low	High	Low	Lognormal	-0.39	-1.84	-1.26	-0.41	-0.23	1.67	-0.27	NA
Old	Low	High	Low	Weibull	-0.32	-1.86	-1.18	-0.31	-0.11	1.82	-0.29	NA
Old	Low	High	Medium	Gompertz	-0.45	-1.29	-0.75	-0.29	-0.25	0.35	-0.50	NA
Old	Low	High	Medium	Log-logistic	-0.54	-1.30	-0.86	-0.42	-0.41	0.30	-0.53	NA
Old	Low	High	Medium	Lognormal	-0.53	-1.30	-0.86	-0.42	-0.41	0.31	-0.52	NA
Old	Low	High	Medium	Weibull	-0.50	-1.30	-0.82	-0.37	-0.34	0.34	-0.52	NA
Old	Low	Low	High	Gompertz	-0.08	-0.16	-0.16	0.15	0.04	-0.15	-0.15	-0.15
Old	Low	Low	High	Log-logistic	-0.17	-0.19	-0.17	-0.13	-0.14	-0.18	-0.18	NA
Old	Low	Low	High	Lognormal	-0.15	-0.17	-0.15	-0.11	-0.13	-0.17	-0.17	NA
Old	Low	Low	High	Weibull	-0.10	-0.17	-0.15	0.08	-0.01	-0.15	-0.15	-0.15
Old	Low	Low	Low	Gompertz	0.55	0.05	0.03	0.92	1.32	0.60	0.37	NA
Old	Low	Low	Low	Log-logistic	-0.11	0.33	-0.51	0.02	0.30	-0.73	NA	NA
Old	Low	Low	Low	Lognormal	-0.19	0.33	-0.55	-0.09	0.12	-0.74	NA	NA
Old	Low	Low	Low	Weibull	0.43	0.41	-0.06	0.77	1.22	-0.18	0.41	NA
Old	Low	Low	Medium	Gompertz	0.04	-0.08	-0.11	0.35	0.36	-0.14	-0.13	NA
Old	Low	Low	Medium	Log-logistic	-0.23	-0.21	-0.31	-0.10	-0.12	-0.34	-0.26	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Old	Low	Low	Medium	Lognormal	-0.18	-0.16	-0.27	-0.07	-0.10	-0.30	-0.18	NA
Old	Low	Low	Medium	Weibull	-0.01	-0.04	-0.14	0.23	0.22	-0.19	-0.14	NA
Old	Low	Medium	High	Gompertz	-0.14	-0.19	-0.19	0.06	-0.03	-0.20	-0.20	-0.20
Old	Low	Medium	High	Log-logistic	-0.20	-0.22	-0.21	-0.15	-0.18	-0.22	-0.22	-0.22
Old	Low	Medium	High	Lognormal	-0.19	-0.21	-0.19	-0.15	-0.17	-0.21	-0.21	-0.21
Old	Low	Medium	High	Weibull	-0.15	-0.19	-0.19	0.00	-0.07	-0.19	-0.19	-0.19
Old	Low	Medium	Low	Gompertz	0.32	-0.22	-0.19	0.68	1.04	0.49	0.13	NA
Old	Low	Medium	Low	Log-logistic	-0.27	-0.02	-0.63	-0.07	0.17	-0.82	NA	NA
Old	Low	Medium	Low	Lognormal	-0.25	0.00	-0.66	-0.13	0.06	-0.83	0.09	NA
Old	Low	Medium	Low	Weibull	0.14	0.02	-0.32	0.46	0.82	-0.31	0.17	NA
Old	Low	Medium	Medium	Gompertz	-0.05	-0.23	-0.22	0.24	0.26	-0.17	-0.20	NA
Old	Low	Medium	Medium	Log-logistic	-0.26	-0.29	-0.38	-0.11	-0.13	-0.39	-0.27	NA
Old	Low	Medium	Medium	Lognormal	-0.23	-0.25	-0.35	-0.10	-0.12	-0.36	-0.21	NA
Old	Low	Medium	Medium	Weibull	-0.11	-0.18	-0.25	0.14	0.14	-0.27	-0.20	NA
Old	Medium	High	High	Gompertz	-1.37	-1.38	-1.39	-1.28	-1.34	-1.39	-1.40	-1.40
Old	Medium	High	High	Log-logistic	-1.34	-1.35	-1.36	-1.25	-1.31	-1.38	-1.37	-1.37
Old	Medium	High	High	Lognormal	-1.29	-1.30	-1.30	-1.21	-1.28	-1.33	-1.32	-1.32
Old	Medium	High	High	Weibull	-1.28	-1.31	-1.31	-1.17	-1.23	-1.33	-1.31	-1.30
Old	Medium	High	Low	Gompertz	-0.98	-2.97	-1.82	-0.90	-0.40	1.13	-0.95	NA
Old	Medium	High	Low	Log-logistic	-0.90	-3.02	-1.74	-0.84	-0.40	1.40	-0.80	NA
Old	Medium	High	Low	Lognormal	-0.95	-2.98	-1.79	-0.88	-0.49	1.40	NA	NA
Old	Medium	High	Low	Weibull	-0.72	-2.96	-1.43	-0.55	-0.07	1.58	-0.87	NA
Old	Medium	High	Medium	Gompertz	-1.11	-2.20	-1.45	-0.94	-0.78	-0.11	-1.17	NA
Old	Medium	High	Medium	Log-logistic	-1.06	-2.22	-1.38	-0.89	-0.75	-0.04	-1.09	NA
Old	Medium	High	Medium	Lognormal	-1.03	-2.17	-1.36	-0.88	-0.77	0.00	-1.00	NA
Old	Medium	High	Medium	Weibull	-0.96	-2.16	-1.23	-0.73	-0.56	0.05	-1.10	NA
Old	Medium	Low	High	Gompertz	-0.95	-1.00	-1.00	-0.81	-0.87	-1.00	-1.00	NA
Old	Medium	Low	High	Log-logistic	-0.97	-0.99	-1.01	-0.87	-0.90	-0.99	-1.01	-1.00
Old	Medium	Low	High	Lognormal	-0.95	-0.95	-0.97	-0.88	-0.92	-0.95	-0.98	-0.98

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Old	Medium	Low	High	Weibull	-0.94	-0.99	-1.01	-0.79	-0.86	-0.99	-1.00	NA
Old	Medium	Low	Low	Gompertz	-0.20	-0.94	-1.02	-0.05	0.83	-0.18	0.14	NA
Old	Medium	Low	Low	Log-logistic	-0.56	-1.50	-1.36	-0.35	0.40	-0.14	-0.39	NA
Old	Medium	Low	Low	Lognormal	-1.00	-1.68	-1.95	-0.94	-0.34	-0.88	-0.21	NA
Old	Medium	Low	Low	Weibull	0.03	-1.73	-0.65	0.24	0.95	1.46	-0.08	NA
Old	Medium	Low	Medium	Gompertz	-0.69	-1.00	-0.98	-0.32	-0.01	-0.84	-0.97	NA
Old	Medium	Low	Medium	Log-logistic	-0.83	-1.33	-1.17	-0.54	-0.30	-0.63	-0.98	NA
Old	Medium	Low	Medium	Lognormal	-0.85	-1.39	-1.27	-0.68	-0.53	-0.51	-0.71	NA
Old	Medium	Low	Medium	Weibull	-0.49	-1.27	-0.76	-0.17	0.11	0.07	-0.93	NA
Old	Medium	Medium	High	Gompertz	-1.02	-1.06	-1.07	-0.90	-0.94	-1.06	-1.07	NA
Old	Medium	Medium	High	Log-logistic	-1.04	-1.06	-1.08	-0.95	-0.98	-1.06	-1.09	NA
Old	Medium	Medium	High	Lognormal	-1.01	-1.02	-1.04	-0.96	-0.99	-1.02	-1.05	NA
Old	Medium	Medium	High	Weibull	-1.00	-1.04	-1.07	-0.86	-0.92	-1.04	-1.06	NA
Old	Medium	Medium	Low	Gompertz	-0.45	-1.25	-1.22	-0.22	0.61	-0.19	NA	NA
Old	Medium	Medium	Low	Log-logistic	-0.70	-1.75	-1.52	-0.51	0.20	-0.08	-0.53	NA
Old	Medium	Medium	Low	Lognormal	-1.00	-1.87	-1.95	-0.94	-0.37	-0.54	-0.31	NA
Old	Medium	Medium	Low	Weibull	-0.12	-1.87	-0.81	0.10	0.80	1.35	-0.31	NA
Old	Medium	Medium	Medium	Gompertz	-0.82	-1.20	-1.13	-0.48	-0.19	-0.85	-1.06	NA
Old	Medium	Medium	Medium	Log-logistic	-0.90	-1.46	-1.25	-0.64	-0.42	-0.58	-1.04	NA
Old	Medium	Medium	Medium	Lognormal	-0.90	-1.51	-1.32	-0.74	-0.60	-0.43	-0.79	NA
Old	Medium	Medium	Medium	Weibull	-0.61	-1.41	-0.89	-0.30	-0.03	0.00	-1.00	NA
Young	High	High	High	Gompertz	-1.40	-5.09	-2.70	-1.16	0.10	1.70	-1.22	NA
Young	High	High	High	Log-logistic	-1.61	-5.33	-2.59	-1.44	-0.25	1.58	-1.60	NA
Young	High	High	High	Lognormal	-1.50	-5.47	-2.46	-1.50	-0.41	1.57	NA	-0.73
Young	High	High	High	Weibull	-1.28	-5.17	-2.51	-1.20	0.03	1.74	-1.37	-0.49
Young	High	High	Medium	Gompertz	-2.42	-8.42	-5.80	-2.46	1.24	2.05	-1.15	NA
Young	High	High	Medium	Log-logistic	-2.83	-9.18	-5.37	-2.74	0.65	2.52	NA	NA
Young	High	High	Medium	Lognormal	-3.10	-9.83	-5.41	-3.07	0.07	2.76	NA	NA
Young	High	High	Medium	Weibull	-2.63	-8.84	-5.44	-2.53	0.98	2.65	NA	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Young	High	Low	High	Gompertz	0.57	1.37	0.32	1.20	2.90	-2.96	NA	NA
Young	High	Low	High	Log-logistic	-1.62	-0.48	-2.06	-0.95	0.92	-5.51	NA	NA
Young	High	Low	High	Lognormal	-2.53	-1.87	-3.47	-2.24	-0.46	-5.48	-1.64	NA
Young	High	Low	High	Weibull	-1.01	-0.12	-1.41	-0.37	1.43	-4.51	-1.07	NA
Young	High	Low	Medium	Gompertz	0.30	2.37	NA	0.32	3.93	-5.43	NA	NA
Young	High	Low	Medium	Log-logistic	-2.34	1.07	NA	-2.48	1.84	-9.77	NA	NA
Young	High	Low	Medium	Lognormal	-5.66	-0.56	NA	-6.19	-0.92	-14.98	NA	NA
Young	High	Low	Medium	Weibull	-1.53	1.18	NA	-1.57	2.47	-8.19	NA	NA
Young	High	Medium	High	Gompertz	0.47	1.56	0.24	1.35	3.24	-3.61	0.02	NA
Young	High	Medium	High	Log-logistic	-1.87	-0.95	-2.42	-1.20	0.69	-5.46	NA	NA
Young	High	Medium	High	Lognormal	-2.51	-2.28	-3.51	-2.28	-0.55	-4.74	-1.69	NA
Young	High	Medium	High	Weibull	-1.30	-0.39	-1.76	-0.57	1.32	-4.86	-1.55	NA
Young	High	Medium	Medium	Gompertz	-4.37	NA	NA	NA	NA	-4.37	NA	NA
Young	High	Medium	Medium	Log-logistic	-2.66	0.45	NA	-2.68	1.81	-10.24	NA	NA
Young	High	Medium	Medium	Lognormal	-14.22	NA	NA	NA	NA	-14.22	NA	NA
Young	High	Medium	Medium	Weibull	-1.56	0.95	NA	-1.51	2.70	-8.39	NA	NA
Young	Low	High	High	Gompertz	-0.04	-2.39	-0.98	0.45	0.39	2.15	-0.05	0.13
Young	Low	High	High	Log-logistic	-0.21	-2.44	-1.22	0.09	-0.03	2.22	-0.10	0.02
Young	Low	High	High	Lognormal	-0.20	-2.42	-1.21	0.10	-0.03	2.24	-0.08	0.02
Young	Low	High	High	Weibull	-0.12	-2.42	-1.11	0.27	0.17	2.25	-0.08	0.06
Young	Low	High	Low	Gompertz	1.55	-4.68	-3.05	0.46	1.19	14.48	0.91	NA
Young	Low	High	Low	Log-logistic	1.08	-4.73	-3.58	-0.35	0.21	13.69	1.24	NA
Young	Low	High	Low	Lognormal	1.09	-4.71	-3.59	-0.37	0.19	13.67	1.36	NA
Young	Low	High	Low	Weibull	1.31	-4.79	-3.41	0.01	0.70	14.22	1.16	NA
Young	Low	High	Medium	Gompertz	0.60	-3.87	-2.09	0.60	0.74	8.18	0.04	NA
Young	Low	High	Medium	Log-logistic	0.33	-3.92	-2.50	0.08	0.09	8.14	0.12	NA
Young	Low	High	Medium	Lognormal	0.36	-3.89	-2.49	0.09	0.09	8.18	0.19	NA
Young	Low	High	Medium	Weibull	0.48	-3.91	-2.33	0.34	0.43	8.29	0.03	NA
Young	Low	Low	High	Gompertz	0.12	-0.05	-0.06	0.67	0.43	-0.06	-0.05	-0.05

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Young	Low	Low	High	Log-logistic	-0.13	-0.15	-0.15	-0.05	-0.11	-0.14	-0.14	NA
Young	Low	Low	High	Lognormal	-0.05	-0.07	-0.07	0.04	-0.05	-0.08	-0.06	NA
Young	Low	Low	High	Weibull	0.06	-0.05	-0.06	0.40	0.20	-0.06	-0.06	NA
Young	Low	Low	Low	Gompertz	2.28	0.24	0.29	2.79	3.99	4.48	1.86	NA
Young	Low	Low	Low	Log-logistic	0.04	0.58	-0.68	0.06	0.38	-0.93	0.80	NA
Young	Low	Low	Low	Lognormal	0.13	0.63	-0.68	-0.11	0.04	-0.88	1.77	NA
Young	Low	Low	Low	Weibull	1.32	0.77	0.03	1.87	2.94	0.40	1.90	NA
Young	Low	Low	Medium	Gompertz	0.41	0.07	0.02	1.22	1.19	-0.03	-0.01	NA
Young	Low	Low	Medium	Log-logistic	-0.26	-0.22	-0.40	-0.04	-0.15	-0.44	-0.29	NA
Young	Low	Low	Medium	Lognormal	-0.11	-0.07	-0.27	0.08	-0.05	-0.31	-0.03	NA
Young	Low	Low	Medium	Weibull	0.21	0.12	-0.05	0.72	0.63	-0.13	-0.04	NA
Young	Low	Medium	High	Gompertz	0.12	-0.08	-0.07	0.71	0.48	-0.05	-0.07	-0.05
Young	Low	Medium	High	Log-logistic	-0.12	-0.21	-0.21	0.01	-0.11	-0.13	-0.15	-0.06
Young	Low	Medium	High	Lognormal	-0.06	-0.13	-0.13	0.08	-0.06	-0.09	-0.06	NA
Young	Low	Medium	High	Weibull	0.05	-0.07	-0.08	0.44	0.24	-0.07	-0.07	-0.05
Young	Low	Medium	Low	Gompertz	1.87	-0.17	-0.09	2.33	3.37	4.46	1.32	NA
Young	Low	Medium	Low	Log-logistic	-0.14	0.00	-0.93	-0.03	0.25	-1.12	0.96	NA
Young	Low	Medium	Low	Lognormal	-0.38	0.11	-0.88	-0.11	0.04	-1.07	NA	NA
Young	Low	Medium	Low	Weibull	0.96	0.22	-0.32	1.48	2.35	0.50	1.52	NA
Young	Low	Medium	Medium	Gompertz	0.36	-0.16	-0.14	1.20	1.16	0.15	-0.08	NA
Young	Low	Medium	Medium	Log-logistic	-0.30	-0.36	-0.53	0.04	-0.12	-0.55	-0.27	NA
Young	Low	Medium	Medium	Lognormal	-0.17	-0.24	-0.42	0.13	-0.05	-0.43	0.00	NA
Young	Low	Medium	Medium	Weibull	0.13	-0.08	-0.21	0.75	0.65	-0.25	-0.10	NA
Young	Medium	High	High	Gompertz	-0.73	-5.30	-1.85	-0.21	0.11	3.25	-0.39	NA
Young	Medium	High	High	Log-logistic	-0.66	-5.39	-1.66	-0.10	0.18	3.20	-0.21	NA
Young	Medium	High	High	Lognormal	-0.66	-5.24	-1.67	-0.16	0.05	3.15	-0.07	NA
Young	Medium	High	High	Weibull	-0.30	-5.01	-1.24	0.44	0.83	3.35	-0.17	NA
Young	Medium	High	Low	Gompertz	-0.50	-10.09	-6.13	-1.46	1.09	14.07	NA	NA
Young	Medium	High	Low	Log-logistic	-0.17	-10.10	-5.96	-1.32	0.94	14.63	0.81	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Young	Medium	High	Low	Lognormal	-0.31	-9.85	-6.25	-1.73	0.19	14.42	1.35	NA
Young	Medium	High	Low	Weibull	0.80	-9.67	-4.49	0.14	2.59	15.41	NA	NA
Young	Medium	High	Medium	Gompertz	-0.46	-8.43	-4.09	-0.50	0.64	9.89	-0.24	NA
Young	Medium	High	Medium	Log-logistic	-0.39	-8.52	-3.84	-0.37	0.66	9.75	-0.02	NA
Young	Medium	High	Medium	Lognormal	-0.42	-8.32	-3.96	-0.54	0.30	9.65	0.38	NA
Young	Medium	High	Medium	Weibull	0.27	-8.04	-2.86	0.60	1.82	10.21	-0.13	NA
Young	Medium	Low	High	Gompertz	-0.41	-1.66	-1.36	0.72	1.01	-0.18	-0.98	NA
Young	Medium	Low	High	Log-logistic	-0.70	-2.87	-1.99	-0.09	0.16	1.61	-1.01	NA
Young	Medium	Low	High	Lognormal	-0.39	-2.82	-1.86	-0.17	-0.12	2.36	-0.17	0.02
Young	Medium	Low	High	Weibull	0.95	-1.40	-0.15	2.17	2.64	2.65	-0.19	NA
Young	Medium	Low	Low	Gompertz	2.85	-2.11	-1.38	2.99	7.46	7.27	NA	NA
Young	Medium	Low	Low	Log-logistic	0.41	-4.85	-3.88	0.38	3.92	6.48	NA	NA
Young	Medium	Low	Low	Lognormal	-1.08	-5.27	-5.63	-2.17	0.11	2.57	3.91	NA
Young	Medium	Low	Low	Weibull	5.52	-3.95	0.56	5.30	9.22	16.52	5.48	NA
Young	Medium	Low	Medium	Gompertz	0.36	-2.31	-1.98	1.97	4.01	1.57	-1.09	NA
Young	Medium	Low	Medium	Log-logistic	-0.79	-4.42	-3.62	0.05	1.46	3.70	-1.92	NA
Young	Medium	Low	Medium	Lognormal	-0.62	-4.58	-3.98	-0.68	0.02	5.20	0.31	NA
Young	Medium	Low	Medium	Weibull	3.13	-2.88	-0.01	4.06	6.10	11.45	0.09	NA
Young	Medium	Medium	High	Gompertz	-0.41	-2.08	-1.53	0.55	0.85	0.65	-0.91	NA
Young	Medium	Medium	High	Log-logistic	-0.61	-3.07	-1.96	-0.06	0.20	2.12	-0.85	NA
Young	Medium	Medium	High	Lognormal	-0.39	-3.04	-1.89	-0.18	-0.11	2.61	-0.15	0.04
Young	Medium	Medium	High	Weibull	0.73	-1.87	-0.44	1.82	2.29	2.87	-0.30	NA
Young	Medium	Medium	Low	Gompertz	2.07	-3.25	-2.22	2.17	6.42	7.21	NA	NA
Young	Medium	Medium	Low	Log-logistic	0.04	-5.50	-4.37	-0.08	3.32	6.83	NA	NA
Young	Medium	Medium	Low	Lognormal	-1.80	-5.74	-5.70	-2.08	0.16	4.33	NA	NA
Young	Medium	Medium	Low	Weibull	4.82	-4.71	-0.33	4.44	8.19	16.50	NA	NA
Young	Medium	Medium	Medium	Gompertz	0.08	-3.12	-2.53	1.39	3.25	2.75	-1.22	NA
Young	Medium	Medium	Medium	Log-logistic	-0.60	-4.85	-3.76	-0.07	1.26	5.50	-1.65	NA
Young	Medium	Medium	Medium	Lognormal	-0.50	-4.92	-4.05	-0.68	0.04	6.37	0.26	NA
Young	Medium	Medium	Medium	Weibull	2.61	-3.54	-0.61	3.42	5.33	11.26	-0.21	NA

Table 97:

Mean bias of overall mean survival – converging hazards models

Scenario characteristics						*Distributions used for extrapolation*						
Age	**Survival**	**Heterogeneity**	**Information**	**Distribution**	**Mean of method**	**Exponential**	**Weibull**	**Log-logistic**	**Lognormal**	**Gompertz**	**Gen. Gamma**	**Gen. F**
Average	High	High	High	Gompertz	1.06	-0.44	0.66	1.46	2.07	1.70	0.92	NA
Average	High	High	High	Log-logistic	1.10	-0.24	0.88	1.41	1.89	1.67	0.99	NA
Average	High	High	High	Lognormal	1.05	-0.21	0.92	1.34	1.75	1.57	0.93	NA
Average	High	High	High	Weibull	1.09	-0.32	0.81	1.45	1.99	1.70	0.94	NA
Average	High	High	Medium	Gompertz	0.49	-2.67	-1.05	1.18	3.38	1.62	NA	NA
Average	High	High	Medium	Log-logistic	0.48	-2.71	-0.58	1.08	2.95	1.67	NA	NA
Average	High	High	Medium	Lognormal	0.50	-2.89	-0.37	1.02	2.65	2.08	NA	NA
Average	High	High	Medium	Weibull	0.76	-2.66	-0.61	1.25	3.24	1.92	1.40	NA
Average	High	Low	High	Gompertz	2.21	3.06	2.44	2.81	3.29	-0.38	2.06	NA
Average	High	Low	High	Log-logistic	1.59	2.59	1.76	2.25	2.84	-1.26	1.34	NA
Average	High	Low	High	Lognormal	1.25	2.09	1.19	1.74	2.37	-1.07	1.21	NA
Average	High	Low	High	Weibull	1.79	2.70	1.96	2.41	2.97	-0.86	1.53	NA
Average	High	Low	Medium	Gompertz	1.64	3.91	NA	2.46	4.06	-3.88	NA	NA
Average	High	Low	Medium	Log-logistic	0.98	3.73	NA	1.81	3.72	-5.35	NA	NA
Average	High	Low	Medium	Lognormal	-0.39	3.27	NA	0.30	2.90	-8.01	NA	NA
Average	High	Low	Medium	Weibull	1.12	3.72	NA	1.90	3.77	-4.91	NA	NA
Average	High	Medium	High	Gompertz	2.21	3.15	2.42	2.91	3.52	-0.55	1.80	NA
Average	High	Medium	High	Log-logistic	1.33	2.24	1.41	1.97	2.63	-1.21	0.96	NA
Average	High	Medium	High	Lognormal	1.09	1.72	0.95	1.52	2.15	-0.81	1.00	1.07
Average	High	Medium	High	Weibull	1.55	2.43	1.64	2.19	2.84	-0.97	1.14	NA
Average	High	Medium	Medium	Gompertz	2.02	4.15	NA	2.94	4.62	-3.61	NA	NA
Average	High	Medium	Medium	Log-logistic	0.72	3.40	NA	1.57	3.65	-5.74	NA	NA
Average	High	Medium	Medium	Lognormal	-0.55	2.68	NA	0.15	2.76	-7.78	NA	NA
Average	High	Medium	Medium	Weibull	1.20	3.55	NA	1.99	3.95	-4.68	NA	NA
Average	Low	High	High	Gompertz	0.32	-1.53	-0.20	0.91	0.90	1.12	0.47	0.56
Average	Low	High	High	Log-logistic	0.20	-1.55	-0.39	0.65	0.58	1.20	0.44	0.50

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Average	Low	High	High	Lognormal	0.20	-1.55	-0.40	0.63	0.56	1.22	0.44	0.50
Average	Low	High	High	Weibull	0.27	-1.54	-0.31	0.78	0.74	1.23	0.45	0.53
Average	Low	High	Low	Gompertz	1.28	-3.71	-2.06	1.08	1.76	9.35	NA	NA
Average	Low	High	Low	Log-logistic	0.77	-3.77	-2.60	0.40	0.98	8.84	NA	NA
Average	Low	High	Low	Lognormal	0.67	-3.74	-2.65	0.29	0.84	8.59	NA	NA
Average	Low	High	Low	Weibull	1.02	-3.79	-2.43	0.66	1.33	9.00	1.36	NA
Average	Low	High	Medium	Gompertz	0.66	-2.94	-1.20	1.13	1.30	5.03	0.61	NA
Average	Low	High	Medium	Log-logistic	0.47	-2.95	-1.54	0.74	0.80	5.08	0.73	NA
Average	Low	High	Medium	Lognormal	0.47	-2.96	-1.56	0.71	0.77	5.09	0.76	NA
Average	Low	High	Medium	Weibull	0.55	-2.96	-1.41	0.90	1.03	5.11	0.66	NA
Average	Low	Low	High	Gompertz	0.13	-0.04	-0.05	0.69	0.45	-0.05	-0.05	-0.05
Average	Low	Low	High	Log-logistic	-0.08	-0.12	-0.12	-0.01	-0.08	-0.11	-0.11	-0.03
Average	Low	Low	High	Lognormal	-0.04	-0.06	-0.06	0.06	-0.04	-0.07	-0.05	NA
Average	Low	Low	High	Weibull	0.06	-0.04	-0.05	0.42	0.22	-0.05	-0.05	-0.05
Average	Low	Low	Low	Gompertz	1.93	0.26	0.28	2.70	3.86	2.94	1.56	NA
Average	Low	Low	Low	Log-logistic	0.02	0.59	-0.66	0.09	0.44	-0.91	0.56	NA
Average	Low	Low	Low	Lognormal	-0.16	0.65	-0.65	-0.06	0.11	-0.85	NA	NA
Average	Low	Low	Low	Weibull	1.28	0.78	0.06	1.89	2.96	0.23	1.78	NA
Average	Low	Low	Medium	Gompertz	0.42	0.08	0.03	1.23	1.21	-0.02	0.00	NA
Average	Low	Low	Medium	Log-logistic	-0.22	-0.18	-0.37	0.00	-0.11	-0.41	-0.26	NA
Average	Low	Low	Medium	Lognormal	-0.09	-0.05	-0.25	0.11	-0.03	-0.29	-0.02	NA
Average	Low	Low	Medium	Weibull	0.22	0.13	-0.05	0.74	0.64	-0.13	-0.04	NA
Average	Low	Medium	High	Gompertz	0.14	-0.07	-0.06	0.72	0.51	-0.04	-0.06	-0.04
Average	Low	Medium	High	Log-logistic	-0.08	-0.16	-0.17	0.06	-0.06	-0.10	-0.10	-0.02
Average	Low	Medium	High	Lognormal	-0.04	-0.10	-0.11	0.11	-0.02	-0.07	-0.04	NA
Average	Low	Medium	High	Weibull	0.06	-0.06	-0.07	0.46	0.26	-0.06	-0.06	-0.04
Average	Low	Medium	Low	Gompertz	1.65	-0.14	-0.04	2.32	3.33	3.12	1.32	NA
Average	Low	Medium	Low	Log-logistic	-0.29	0.06	-0.86	0.06	0.33	-1.07	NA	NA
Average	Low	Medium	Low	Lognormal	-0.33	0.15	-0.84	-0.04	0.12	-1.02	NA	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Average	Low	Medium	Low	Weibull	0.87	0.24	-0.30	1.47	2.32	0.14	1.35	NA
Average	Low	Medium	Medium	Gompertz	0.36	-0.14	-0.12	1.20	1.19	0.11	-0.07	NA
Average	Low	Medium	Medium	Log-logistic	-0.24	-0.30	-0.47	0.11	-0.05	-0.49	-0.21	NA
Average	Low	Medium	Medium	Lognormal	-0.14	-0.21	-0.39	0.17	0.00	-0.40	0.01	NA
Average	Low	Medium	Medium	Weibull	0.15	-0.06	-0.19	0.78	0.69	-0.23	-0.09	NA
Average	Medium	High	High	Gompertz	0.52	-2.77	0.10	1.14	1.36	1.92	0.90	1.00
Average	Medium	High	High	Log-logistic	0.44	-2.87	0.20	1.17	1.37	1.79	0.97	NA
Average	Medium	High	High	Lognormal	0.49	-2.84	0.14	1.08	1.23	1.75	1.00	1.05
Average	Medium	High	High	Weibull	0.64	-2.70	0.43	1.48	1.75	1.88	0.98	NA
Average	Medium	High	Low	Gompertz	0.49	-7.27	-3.41	0.50	2.50	8.87	1.75	NA
Average	Medium	High	Low	Log-logistic	0.50	-7.37	-3.31	0.52	2.32	9.16	1.68	NA
Average	Medium	High	Low	Lognormal	0.32	-7.24	-3.71	0.13	1.72	9.04	1.96	NA
Average	Medium	High	Low	Weibull	1.28	-7.10	-2.10	1.66	3.58	9.90	1.74	NA
Average	Medium	High	Medium	Gompertz	0.56	-5.72	-1.55	1.21	2.11	6.10	1.21	NA
Average	Medium	High	Medium	Log-logistic	0.59	-5.81	-1.36	1.26	2.06	5.98	1.40	NA
Average	Medium	High	Medium	Lognormal	0.52	-5.69	-1.55	1.07	1.74	5.93	1.61	NA
Average	Medium	High	Medium	Weibull	1.00	-5.56	-0.66	1.90	2.83	6.23	1.26	NA
Average	Medium	Low	High	Gompertz	0.31	-0.75	-0.48	1.35	1.62	0.30	-0.17	NA
Average	Medium	Low	High	Log-logistic	0.21	-1.51	-0.68	0.88	1.13	1.34	0.09	NA
Average	Medium	Low	High	Lognormal	0.28	-1.54	-0.65	0.72	0.81	1.68	0.65	NA
Average	Medium	Low	High	Weibull	1.17	-0.70	0.46	2.31	2.68	1.90	0.34	NA
Average	Medium	Low	Low	Gompertz	2.69	-1.10	-0.43	3.43	7.06	4.51	NA	NA
Average	Medium	Low	Low	Log-logistic	0.96	-3.43	-2.47	1.36	4.38	3.52	2.42	NA
Average	Medium	Low	Low	Lognormal	-1.22	-3.82	-4.11	-0.80	1.29	1.36	NA	NA
Average	Medium	Low	Low	Weibull	4.22	-3.17	1.08	4.99	7.98	10.21	NA	NA
Average	Medium	Low	Medium	Gompertz	0.95	-1.32	-1.02	2.57	4.31	1.26	-0.13	NA
Average	Medium	Low	Medium	Log-logistic	0.06	-2.98	-2.18	1.14	2.39	2.62	-0.63	NA
Average	Medium	Low	Medium	Lognormal	0.02	-3.21	-2.63	0.41	1.10	3.38	1.09	NA
Average	Medium	Low	Medium	Weibull	2.68	-2.10	0.72	4.05	5.65	7.26	0.51	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Average	Medium	Medium	High	Gompertz	0.35	-1.00	-0.50	1.26	1.54	0.80	-0.01	NA
Average	Medium	Medium	High	Log-logistic	0.25	-1.63	-0.63	0.90	1.14	1.54	0.20	NA
Average	Medium	Medium	High	Lognormal	0.28	-1.67	-0.61	0.73	0.83	1.76	0.66	NA
Average	Medium	Medium	High	Weibull	1.04	-0.97	0.33	2.11	2.47	1.98	0.35	NA
Average	Medium	Medium	Low	Gompertz	1.85	-2.05	-1.25	2.61	6.05	3.90	NA	NA
Average	Medium	Medium	Low	Log-logistic	0.49	-3.94	-2.80	1.09	3.97	4.10	NA	NA
Average	Medium	Medium	Low	Lognormal	-0.43	-4.22	-4.20	-0.77	1.31	2.32	3.00	NA
Average	Medium	Medium	Low	Weibull	3.76	-3.66	0.52	4.47	7.35	10.13	NA	NA
Average	Medium	Medium	Medium	Gompertz	0.72	-1.95	-1.37	2.16	3.77	1.97	-0.29	NA
Average	Medium	Medium	Medium	Log-logistic	0.18	-3.29	-2.22	1.09	2.28	3.61	-0.42	NA
Average	Medium	Medium	Medium	Lognormal	0.10	-3.48	-2.64	0.42	1.11	4.12	1.09	NA
Average	Medium	Medium	Medium	Weibull	2.34	-2.53	0.31	3.61	5.14	7.12	0.37	NA
Old	High	High	High	Gompertz	-0.54	-0.53	-0.54	-0.48	-0.55	-0.54	-0.57	NA
Old	High	High	High	Log-logistic	-0.67	-0.65	-0.66	-0.64	-0.73	-0.67	-0.69	NA
Old	High	High	High	Lognormal	-0.73	-0.70	-0.71	-0.70	-0.80	-0.72	-0.75	NA
Old	High	High	High	Weibull	-0.61	-0.59	-0.60	-0.57	-0.65	-0.61	-0.64	NA
Old	High	High	Medium	Gompertz	-0.35	-0.69	-0.58	0.02	0.40	-0.56	-0.66	NA
Old	High	High	Medium	Log-logistic	-0.48	-0.81	-0.67	-0.16	0.18	-0.69	-0.73	-0.46
Old	High	High	Medium	Lognormal	-0.50	-0.88	-0.67	-0.22	0.08	-0.62	-0.71	NA
Old	High	High	Medium	Weibull	-0.39	-0.75	-0.58	-0.06	0.29	-0.57	-0.69	NA
Old	High	Low	High	Gompertz	-0.09	-0.10	0.05	0.05	-0.13	-0.12	-0.28	NA
Old	High	Low	High	Log-logistic	-0.08	-0.09	0.05	0.06	-0.11	-0.14	-0.27	NA
Old	High	Low	High	Lognormal	-0.08	-0.08	0.05	0.06	-0.10	-0.14	-0.23	NA
Old	High	Low	High	Weibull	-0.08	-0.09	0.06	0.06	-0.11	-0.13	-0.26	NA
Old	High	Low	Medium	Gompertz	0.30	0.71	0.43	0.63	0.77	-0.94	0.17	NA
Old	High	Low	Medium	Log-logistic	0.29	0.75	0.41	0.65	0.81	-1.03	0.17	NA
Old	High	Low	Medium	Lognormal	0.24	0.77	0.35	0.64	0.84	-1.22	0.09	NA
Old	High	Low	Medium	Weibull	0.31	0.75	0.43	0.65	0.81	-0.97	0.17	NA
Old	High	Medium	High	Gompertz	-0.36	-0.36	-0.26	-0.25	-0.41	-0.38	-0.49	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Old	High	Medium	High	Log-logistic	-0.38	-0.39	-0.29	-0.27	-0.42	-0.41	-0.48	NA
Old	High	Medium	High	Lognormal	-0.38	-0.39	-0.31	-0.29	-0.43	-0.41	-0.46	NA
Old	High	Medium	High	Weibull	-0.36	-0.37	-0.28	-0.26	-0.41	-0.39	-0.48	NA
Old	High	Medium	Medium	Gompertz	0.01	0.38	0.09	0.36	0.55	-1.13	-0.19	NA
Old	High	Medium	Medium	Log-logistic	-0.05	0.37	0.01	0.32	0.54	-1.26	-0.26	NA
Old	High	Medium	Medium	Lognormal	-0.14	0.34	-0.11	0.26	0.53	-1.49	-0.38	NA
Old	High	Medium	Medium	Weibull	-0.03	0.37	0.03	0.33	0.56	-1.21	-0.27	NA
Old	Low	High	High	Gompertz	-0.04	-0.36	-0.03	0.23	0.18	-0.22	-0.01	NA
Old	Low	High	High	Log-logistic	-0.08	-0.36	-0.08	0.13	0.08	-0.19	-0.05	NA
Old	Low	High	High	Lognormal	-0.07	-0.36	-0.08	0.13	0.08	-0.18	-0.04	NA
Old	Low	High	High	Weibull	-0.05	-0.36	-0.06	0.18	0.12	-0.19	-0.03	-0.03
Old	Low	High	Low	Gompertz	0.35	-1.77	-0.64	0.68	0.99	2.39	0.46	NA
Old	Low	High	Low	Log-logistic	0.16	-1.76	-0.97	0.39	0.66	2.14	0.47	NA
Old	Low	High	Low	Lognormal	0.18	-1.76	-0.95	0.41	0.68	2.19	0.54	NA
Old	Low	High	Low	Weibull	0.27	-1.78	-0.82	0.55	0.85	2.31	0.52	NA
Old	Low	High	Medium	Gompertz	0.16	-1.15	-0.21	0.54	0.63	0.92	0.24	NA
Old	Low	High	Medium	Log-logistic	0.08	-1.14	-0.38	0.39	0.44	0.92	0.23	NA
Old	Low	High	Medium	Lognormal	0.08	-1.13	-0.38	0.39	0.44	0.94	0.25	NA
Old	Low	High	Medium	Weibull	0.12	-1.14	-0.30	0.47	0.53	0.95	0.24	NA
Old	Low	Low	High	Gompertz	0.08	-0.04	-0.05	0.47	0.34	-0.05	-0.05	-0.05
Old	Low	Low	High	Log-logistic	-0.03	-0.06	-0.06	0.05	0.00	-0.06	-0.07	NA
Old	Low	Low	High	Lognormal	-0.02	-0.05	-0.05	0.07	0.00	-0.06	-0.05	NA
Old	Low	Low	High	Weibull	0.04	-0.04	-0.05	0.33	0.20	-0.05	-0.05	-0.05
Old	Low	Low	Low	Gompertz	1.00	0.22	0.21	1.62	2.23	0.92	0.78	NA
Old	Low	Low	Low	Log-logistic	0.07	0.58	-0.45	0.26	0.64	-0.69	NA	NA
Old	Low	Low	Low	Lognormal	-0.04	0.57	-0.50	0.10	0.36	-0.71	NA	NA
Old	Low	Low	Low	Weibull	0.81	0.66	0.08	1.35	2.02	-0.04	0.81	NA
Old	Low	Low	Medium	Gompertz	0.27	0.06	0.02	0.79	0.82	-0.03	-0.01	NA
Old	Low	Low	Medium	Log-logistic	-0.11	-0.09	-0.24	0.08	0.03	-0.27	-0.17	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Old	Low	Low	Medium	Lognormal	-0.06	-0.03	-0.19	0.12	0.05	-0.22	-0.07	NA
Old	Low	Low	Medium	Weibull	0.17	0.10	-0.04	0.55	0.53	-0.11	-0.04	NA
Old	Low	Medium	High	Gompertz	0.07	-0.05	-0.05	0.43	0.32	-0.05	-0.05	-0.04
Old	Low	Medium	High	Log-logistic	-0.03	-0.07	-0.08	0.09	0.03	-0.06	-0.06	-0.04
Old	Low	Medium	High	Lognormal	-0.02	-0.06	-0.06	0.09	0.02	-0.06	-0.05	-0.04
Old	Low	Medium	High	Weibull	0.04	-0.05	-0.06	0.31	0.20	-0.05	-0.05	-0.05
Old	Low	Medium	Low	Gompertz	0.74	-0.08	-0.02	1.35	1.89	0.83	0.50	NA
Old	Low	Medium	Low	Log-logistic	-0.09	0.17	-0.58	0.22	0.53	-0.78	NA	NA
Old	Low	Medium	Low	Lognormal	-0.05	0.19	-0.61	0.12	0.36	-0.80	0.45	NA
Old	Low	Medium	Low	Weibull	0.49	0.21	-0.21	1.00	1.54	-0.17	0.56	NA
Old	Low	Medium	Medium	Gompertz	0.21	-0.09	-0.08	0.71	0.75	0.00	-0.05	NA
Old	Low	Medium	Medium	Log-logistic	-0.11	-0.15	-0.29	0.15	0.09	-0.30	-0.13	NA
Old	Low	Medium	Medium	Lognormal	-0.07	-0.12	-0.26	0.16	0.10	-0.27	-0.05	NA
Old	Low	Medium	Medium	Weibull	0.10	-0.04	-0.14	0.51	0.51	-0.17	-0.07	NA
Old	Medium	High	High	Gompertz	-0.19	-0.46	-0.16	-0.01	-0.06	-0.28	-0.16	-0.16
Old	Medium	High	High	Log-logistic	-0.20	-0.49	-0.15	-0.01	-0.07	-0.33	-0.16	-0.16
Old	Medium	High	High	Lognormal	-0.19	-0.49	-0.13	-0.01	-0.08	-0.32	-0.15	-0.15
Old	Medium	High	High	Weibull	-0.14	-0.46	-0.10	0.07	0.02	-0.30	-0.10	-0.10
Old	Medium	High	Low	Gompertz	0.01	-2.69	-0.87	0.50	1.13	1.81	0.19	NA
Old	Medium	High	Low	Log-logistic	0.06	-2.80	-0.79	0.52	1.08	2.02	0.35	NA
Old	Medium	High	Low	Lognormal	-0.06	-2.78	-0.90	0.43	0.95	2.01	NA	NA
Old	Medium	High	Low	Weibull	0.27	-2.74	-0.35	0.87	1.45	2.17	0.24	NA
Old	Medium	High	Medium	Gompertz	0.05	-1.68	-0.24	0.47	0.68	0.91	0.17	NA
Old	Medium	High	Medium	Log-logistic	0.05	-1.76	-0.19	0.48	0.67	0.90	0.22	NA
Old	Medium	High	Medium	Lognormal	0.04	-1.76	-0.23	0.44	0.60	0.90	0.29	NA
Old	Medium	High	Medium	Weibull	0.16	-1.71	0.01	0.66	0.88	0.96	0.16	NA
Old	Medium	Low	High	Gompertz	0.17	0.03	0.03	0.47	0.45	0.03	0.03	NA
Old	Medium	Low	High	Log-logistic	0.13	-0.07	0.03	0.35	0.35	0.09	0.06	0.07
Old	Medium	Low	High	Lognormal	0.11	-0.09	0.02	0.26	0.25	0.10	0.09	0.09

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Old	Medium	Low	High	Weibull	0.24	0.02	0.15	0.55	0.49	0.14	0.08	NA
Old	Medium	Low	Low	Gompertz	1.04	0.16	0.03	1.56	2.68	0.50	1.30	NA
Old	Medium	Low	Low	Log-logistic	0.50	-0.82	-0.53	1.08	2.09	0.52	0.68	NA
Old	Medium	Low	Low	Lognormal	-0.14	-1.15	-1.50	0.16	1.05	-0.32	0.93	NA
Old	Medium	Low	Low	Weibull	1.26	-1.10	0.67	1.92	2.74	2.23	1.12	NA
Old	Medium	Low	Medium	Gompertz	0.56	0.03	0.07	1.24	1.72	0.22	0.09	NA
Old	Medium	Low	Medium	Log-logistic	0.31	-0.58	-0.26	0.85	1.23	0.53	0.10	NA
Old	Medium	Low	Medium	Lognormal	0.22	-0.75	-0.50	0.58	0.84	0.64	0.52	NA
Old	Medium	Low	Medium	Weibull	0.83	-0.43	0.57	1.44	1.81	1.42	0.20	NA
Old	Medium	Medium	High	Gompertz	0.10	-0.08	-0.03	0.36	0.34	0.00	-0.02	NA
Old	Medium	Medium	High	Log-logistic	0.04	-0.17	-0.05	0.25	0.24	0.01	-0.02	NA
Old	Medium	Medium	High	Lognormal	0.02	-0.19	-0.05	0.17	0.16	0.02	0.01	NA
Old	Medium	Medium	High	Weibull	0.15	-0.10	0.06	0.44	0.40	0.06	0.01	NA
Old	Medium	Medium	Low	Gompertz	0.72	-0.34	-0.25	1.32	2.40	0.48	NA	NA
Old	Medium	Medium	Low	Log-logistic	0.32	-1.17	-0.75	0.87	1.83	0.61	0.52	NA
Old	Medium	Medium	Low	Lognormal	-0.13	-1.40	-1.47	0.19	1.03	0.07	0.82	NA
Old	Medium	Medium	Low	Weibull	1.06	-1.31	0.42	1.73	2.56	2.12	0.85	NA
Old	Medium	Medium	Medium	Gompertz	0.38	-0.30	-0.14	1.02	1.45	0.25	-0.02	NA
Old	Medium	Medium	Medium	Log-logistic	0.21	-0.78	-0.35	0.72	1.07	0.59	0.03	NA
Old	Medium	Medium	Medium	Lognormal	0.16	-0.92	-0.54	0.51	0.75	0.72	0.43	NA
Old	Medium	Medium	Medium	Weibull	0.66	-0.66	0.37	1.25	1.61	1.30	0.10	NA
Young	High	High	High	Gompertz	1.62	-2.38	0.57	2.11	3.35	4.07	2.00	NA
Young	High	High	High	Log-logistic	1.69	-1.99	1.12	2.11	3.14	3.89	1.90	NA
Young	High	High	High	Lognormal	1.83	-1.86	1.37	2.11	2.98	3.75	NA	2.63
Young	High	High	High	Weibull	1.91	-2.10	1.02	2.23	3.36	4.08	2.01	2.76
Young	High	High	Medium	Gompertz	0.00	-6.50	-3.29	0.57	4.47	3.35	1.39	NA
Young	High	High	Medium	Log-logistic	-0.13	-6.73	-2.23	0.62	3.96	3.76	NA	NA
Young	High	High	Medium	Lognormal	-0.32	-7.20	-2.09	0.36	3.37	3.93	NA	NA
Young	High	High	Medium	Weibull	-0.04	-6.61	-2.53	0.71	4.27	3.93	NA	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Young	High	Low	High	Gompertz	4.73	5.92	4.96	5.58	6.71	0.50	NA	NA
Young	High	Low	High	Log-logistic	2.53	4.10	2.54	3.46	4.87	-2.31	NA	NA
Young	High	Low	High	Lognormal	1.61	2.69	1.03	2.11	3.55	-2.11	2.37	NA
Young	High	Low	High	Weibull	3.18	4.46	3.19	4.02	5.33	-1.15	3.22	NA
Young	High	Low	Medium	Gompertz	3.29	6.71	NA	3.83	7.06	-4.44	NA	NA
Young	High	Low	Medium	Log-logistic	0.70	5.44	NA	1.04	5.17	-8.84	NA	NA
Young	High	Low	Medium	Lognormal	-2.62	3.88	NA	-2.81	2.65	-14.19	NA	NA
Young	High	Low	Medium	Weibull	1.53	5.56	NA	1.99	5.75	-7.20	NA	NA
Young	High	Medium	High	Gompertz	4.60	6.13	4.80	5.71	7.13	-0.49	4.31	NA
Young	High	Medium	High	Log-logistic	2.23	3.60	2.06	3.12	4.64	-2.29	NA	NA
Young	High	Medium	High	Lognormal	1.59	2.23	0.91	2.00	3.42	-1.27	2.25	NA
Young	High	Medium	High	Weibull	2.81	4.16	2.73	3.76	5.25	-1.67	2.60	NA
Young	High	Medium	Medium	Gompertz	-3.34	NA	NA	NA	NA	-3.34	NA	NA
Young	High	Medium	Medium	Log-logistic	0.41	4.88	NA	0.85	5.23	-9.31	NA	NA
Young	High	Medium	Medium	Lognormal	-13.37	NA	NA	NA	NA	-13.37	NA	NA
Young	High	Medium	Medium	Weibull	1.53	5.38	NA	2.09	6.07	-7.41	NA	NA
Young	Low	High	High	Gompertz	0.41	-2.36	-0.67	1.12	1.06	2.45	0.52	0.73
Young	Low	High	High	Log-logistic	0.23	-2.40	-0.95	0.72	0.58	2.54	0.49	0.64
Young	Low	High	High	Lognormal	0.24	-2.38	-0.94	0.73	0.58	2.57	0.52	0.64
Young	Low	High	High	Weibull	0.33	-2.38	-0.82	0.92	0.82	2.57	0.51	0.68
Young	Low	High	Low	Gompertz	1.95	-4.68	-2.97	1.14	1.97	14.65	1.60	NA
Young	Low	High	Low	Log-logistic	1.45	-4.72	-3.53	0.22	0.85	13.89	1.99	NA
Young	Low	High	Low	Lognormal	1.46	-4.71	-3.54	0.20	0.83	13.87	2.12	NA
Young	Low	High	Low	Weibull	1.71	-4.78	-3.34	0.64	1.44	14.41	1.90	NA
Young	Low	High	Medium	Gompertz	1.00	-3.86	-1.94	1.30	1.46	8.42	0.63	NA
Young	Low	High	Medium	Log-logistic	0.72	-3.91	-2.39	0.71	0.72	8.41	0.75	NA
Young	Low	High	Medium	Lognormal	0.74	-3.88	-2.38	0.72	0.72	8.46	0.82	NA
Young	Low	High	Medium	Weibull	0.87	-3.90	-2.20	1.01	1.12	8.55	0.64	NA
Young	Low	Low	High	Gompertz	0.14	-0.04	-0.05	0.75	0.47	-0.05	-0.05	-0.05

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Young	Low	Low	High	Log-logistic	-0.11	-0.14	-0.15	-0.03	-0.10	-0.13	-0.13	NA
Young	Low	Low	High	Lognormal	-0.04	-0.06	-0.06	0.07	-0.04	-0.07	-0.05	NA
Young	Low	Low	High	Weibull	0.08	-0.04	-0.05	0.44	0.22	-0.05	-0.05	NA
Young	Low	Low	Low	Gompertz	2.48	0.26	0.31	3.14	4.49	4.61	2.11	NA
Young	Low	Low	Low	Log-logistic	0.08	0.60	-0.68	0.09	0.42	-0.92	0.95	NA
Young	Low	Low	Low	Lognormal	0.18	0.65	-0.68	-0.09	0.05	-0.88	2.04	NA
Young	Low	Low	Low	Weibull	1.45	0.79	0.04	2.06	3.24	0.44	2.15	NA
Young	Low	Low	Medium	Gompertz	0.46	0.08	0.03	1.36	1.30	-0.02	0.00	NA
Young	Low	Low	Medium	Log-logistic	-0.25	-0.21	-0.40	-0.02	-0.14	-0.44	-0.28	NA
Young	Low	Low	Medium	Lognormal	-0.09	-0.06	-0.26	0.11	-0.04	-0.30	-0.02	NA
Young	Low	Low	Medium	Weibull	0.23	0.13	-0.05	0.79	0.67	-0.13	-0.04	NA
Young	Low	Medium	High	Gompertz	0.16	-0.07	-0.06	0.83	0.55	-0.03	-0.06	-0.03
Young	Low	Medium	High	Log-logistic	-0.10	-0.20	-0.20	0.05	-0.09	-0.11	-0.13	-0.03
Young	Low	Medium	High	Lognormal	-0.04	-0.12	-0.12	0.12	-0.04	-0.08	-0.04	NA
Young	Low	Medium	High	Weibull	0.07	-0.06	-0.07	0.51	0.27	-0.06	-0.06	-0.04
Young	Low	Medium	Low	Gompertz	2.06	-0.16	-0.07	2.66	3.82	4.60	1.54	NA
Young	Low	Medium	Low	Log-logistic	-0.09	0.02	-0.92	0.02	0.30	-1.11	1.16	NA
Young	Low	Medium	Low	Lognormal	-0.36	0.12	-0.88	-0.07	0.08	-1.07	NA	NA
Young	Low	Medium	Low	Weibull	1.09	0.23	-0.31	1.67	2.62	0.55	1.76	NA
Young	Low	Medium	Medium	Gompertz	0.42	-0.15	-0.12	1.38	1.31	0.19	-0.07	NA
Young	Low	Medium	Medium	Log-logistic	-0.28	-0.36	-0.53	0.09	-0.10	-0.55	-0.25	NA
Young	Low	Medium	Medium	Lognormal	-0.15	-0.23	-0.41	0.18	-0.03	-0.43	0.03	NA
Young	Low	Medium	Medium	Weibull	0.16	-0.07	-0.20	0.85	0.71	-0.24	-0.09	NA
Young	Medium	High	High	Gompertz	0.40	-5.07	-0.64	1.31	1.71	3.97	1.14	NA
Young	Medium	High	High	Log-logistic	0.45	-5.18	-0.44	1.40	1.75	3.85	1.31	NA
Young	Medium	High	High	Lognormal	0.40	-5.06	-0.52	1.28	1.55	3.77	1.41	NA
Young	Medium	High	High	Weibull	0.80	-4.82	-0.02	1.94	2.41	4.00	1.27	NA
Young	Medium	High	Low	Gompertz	0.25	-10.07	-5.64	-0.14	2.76	14.35	NA	NA
Young	Medium	High	Low	Log-logistic	0.67	-10.08	-5.48	-0.01	2.56	14.87	2.15	NA

Age	Survival	Heterogeneity	Information	Distribution	Mean of method	Exponential	Weibull	Log-logistic	Lognormal	Gompertz	Gen. Gamma	Gen. F
Young	Medium	High	Low	Lognormal	0.47	-9.84	-5.86	-0.53	1.67	14.65	2.71	NA
Young	Medium	High	Low	Weibull	1.61	-9.65	-3.84	1.59	4.34	15.64	NA	NA
Young	Medium	High	Medium	Gompertz	0.52	-8.37	-3.34	0.96	2.29	10.33	1.23	NA
Young	Medium	High	Medium	Log-logistic	0.58	-8.47	-3.04	1.09	2.28	10.15	1.46	NA
Young	Medium	High	Medium	Lognormal	0.50	-8.28	-3.25	0.84	1.82	10.03	1.87	NA
Young	Medium	High	Medium	Weibull	1.26	-7.99	-1.96	2.13	3.53	10.60	1.28	NA
Young	Medium	Low	High	Gompertz	0.11	-1.48	-1.11	1.57	1.93	0.39	-0.62	NA
Young	Medium	Low	High	Log-logistic	-0.03	-2.71	-1.63	0.84	1.15	2.47	-0.32	NA
Young	Medium	Low	High	Lognormal	0.31	-2.68	-1.49	0.72	0.78	3.16	0.71	0.95
Young	Medium	Low	High	Weibull	1.71	-1.22	0.36	3.31	3.91	3.55	0.31	NA
Young	Medium	Low	Low	Gompertz	3.66	-1.93	-0.93	4.23	9.33	7.59	NA	NA
Young	Medium	Low	Low	Log-logistic	1.05	-4.78	-3.61	1.40	5.47	6.79	NA	NA
Young	Medium	Low	Low	Lognormal	-0.55	-5.22	-5.56	-1.57	1.05	2.84	5.14	NA
Young	Medium	Low	Low	Weibull	6.53	-3.88	1.31	6.91	11.32	16.88	6.67	NA
Young	Medium	Low	Medium	Gompertz	1.01	-2.17	-1.79	3.05	5.47	2.14	-0.63	NA
Young	Medium	Low	Medium	Log-logistic	-0.18	-4.34	-3.46	1.01	2.68	4.40	-1.38	NA
Young	Medium	Low	Medium	Lognormal	-0.03	-4.52	-3.87	0.12	0.95	5.90	1.26	NA
Young	Medium	Low	Medium	Weibull	4.02	-2.78	0.54	5.51	7.90	12.21	0.75	NA
Young	Medium	Medium	High	Gompertz	0.21	-1.89	-1.21	1.47	1.86	1.45	-0.40	NA
Young	Medium	Medium	High	Log-logistic	0.11	-2.91	-1.53	0.93	1.26	2.98	-0.06	NA
Young	Medium	Medium	High	Lognormal	0.35	-2.89	-1.47	0.76	0.85	3.40	0.79	1.02
Young	Medium	Medium	High	Weibull	1.51	-1.69	0.12	2.99	3.58	3.77	0.30	NA
Young	Medium	Medium	Low	Gompertz	2.84	-3.11	-1.82	3.37	8.23	7.52	NA	NA
Young	Medium	Medium	Low	Log-logistic	0.67	-5.44	-4.13	0.92	4.84	7.15	NA	NA
Young	Medium	Medium	Low	Lognormal	-1.39	-5.69	-5.62	-1.42	1.16	4.61	NA	NA
Young	Medium	Medium	Low	Weibull	5.75	-4.66	0.37	5.99	10.21	16.85	NA	NA
Young	Medium	Medium	Medium	Gompertz	0.76	-3.00	-2.33	2.46	4.67	3.43	-0.69	NA
Young	Medium	Medium	Medium	Log-logistic	0.05	-4.77	-3.58	0.93	2.51	6.24	-1.00	NA
Young	Medium	Medium	Medium	Lognormal	0.12	-4.86	-3.92	0.18	1.03	7.06	1.25	NA
Young	Medium	Medium	Medium	Weibull	3.48	-3.45	-0.07	4.84	7.08	11.99	0.49	NA