Abstract

Objectives

Extrapolation of time-to-event data can be a critical component of cost-effectiveness analysis. This paper contrasts the value of external data on treatment effects as a selection aid in model fitting to the clinical data or for the direct extrapolation of survival.

Methods

We assume the existence of external summary data on both treatment and control and consider two scenarios: availability of external individual patient data (IPD) on the control only; and an absence of external IPD. We describe how the summary data can be used to extrapolate survival or to assess the plausibility of extrapolations of the clinical data. We assess the merit of either approach using a comparison of Cemented and Cementless Total Hip Replacement as a case study. Merit is judged by comparing Incremental Net Benefit (INB) obtained in scenarios with incomplete IPD with that derived from modelling external IPD on both treatment and control.

Results

Measures of fit with the external summary data did not identify survival model specifications which best estimated INB. Addition of external IPD for the control only did not improve estimates of INB. Extrapolation of survival using the external summary data comparing treatment and control improved estimates of INB.

Conclusions

Our case study indicates that summary data comparing treatment and control are more valuable than IPD limited to the control when extrapolating event rates for cost-effectiveness analysis. These data are best exploited in direct extrapolation of event rates, rather than as an aid to select extrapolations based on the clinical data.

Introduction

Accurate quantification of incremental costs and benefits of new medical technologies often requires the extrapolation of event data beyond the time horizon of clinical trials.(1,2) Generally, this is undertaken by parameterisation of the hazard rate for the event of interest with respect to time, allowing extrapolation and prediction of event rates beyond the point of maximum trial follow-up (hereon 't_{max}').(3,4) Selection of the appropriate parameterisation of time is often based on a limited number of standard specifications ranked by measures of model fit such as Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC).(5) Such measures give little confidence that extrapolations are appropriate and there is mounting evidence that inappropriate extrapolations can lead to incorrect inference on cost-effectiveness.(6-9)

In this context it is important to consider any additional data that might inform event rates beyond t_{max}. Frequently, long term survival data are available in disease registries, administrative databases or mortality statistics. A number of publications have addressed the incorporation of such data to inform survival extrapolations.(10) Joint modelling of trial data and individual patient data (IPD) from the external source is typically undertaken using a Bayesian framework under assumptions of proportional hazards, proportional cause specific hazards or additive hazards. Whilst this approach can ensure plausible extrapolations, the external IPD are rarely available on both treatment and control. In the literature to date the treatment effect is estimated from the trial data (11) or informed entirely from an external source such as a hazard ratio derived from meta-analysis.(12-14) There is considerable evidence which demonstrates the sensitivity of cost-effectiveness inference to assumptions on *relative* survival after extrapolation of data.(9,15,16)

In this paper we address the incorporation of external data identifying both treatment and control to improve estimation of *relative* survival, within a likelihood based framework. Such data may be limited to summary statistics from case reports and observational studies, or derived from expert opinion.(17) Recent methodological guidance from the National Institute of Health and Care Excellence (NICE)

Decision Support Unit recommended utilisation of external data to guide model selection on the clinical data or to directly extrapolate survival, but did not elaborate on methods.(18) We illustrate how such data can be used to guide model selection, or to directly extrapolate survival beyond t_{max}. We consider scenarios where external data are available in summary form only and where, in addition, IPD are available for the control. We compare these scenarios with a scenario in which IPD is available on treatment and control which we consider to provide the 'best estimates' of survival. We contrast inference derived from modelling survival in these scenarios with the optimal situation in which IPD are available for both treatment and control in a case study. We conclude with recommendations on the utilization of external data to extrapolate survival.

Methods

Overview

We use a previously published economic evaluation of alternative technologies for Total Hip Replacement (THR) to illustrate the role of external data.(19) Here, IPD on treatment and control, taken from the National Joint Registry of England, Wales, Northern Ireland (NJR), constitutes the clinical or 'trial' data.(20) The external IPD are drawn from a large hospital administrative database, Hospital Episode Statistics (HES).(21) The external data are used to simulate a scenario in which external IPD are available for the control only, and to generate summary survival statistics for treatment and control for six subgroups – men and women aged 60, 70 and 80. We describe how these data might be used to guide survival model selection, or to estimate event rates beyond t_{max}. We apply the best fitting parameterisations of event data to a probabilistic Markov model to generate Incremental Net Benefit (INB) and Cost-Effectiveness Acceptability Curves (CEACs) for Cementless THR compared to Cemented THR. We evaluate how well alternative methods to extrapolate event data perform by comparing the resulting estimates of INB and CEACs with those derived from modelling the external IPD for both treatment and control.

Application of external data

External data in summary form only

We assume summary survival statistics from the external data are available at time $t_1 < t_{max}$ and $t_2 > t_{max}$ by population subgroup *c*. We assume no published Kaplan-Meier survival curves are available; methods are available to reconstruct life table data from Kaplan-Meier curves.(22) These data can be used in two ways: as a selection aid to evaluate survival models fitted to the clinical data; or to directly estimate survival beyond t_{max}. Use of these data as a selection aid requires criteria for ranking models. In the absence of external IPD we cannot calculate the Brier score (23) or Harrell's *c* index (24). Instead, we apply measures based on the sum of the squared error between predicted and observed survival.(4) We calculate two measures of how well *relative* survival at t_2 predicted by each model compares to the observed data. For each model, we calculate the *difference* in predicted survival between treatment and control (d_{pc}) for each population subgroup *c* at time t_2. We calculate the comparable difference (d_{oc}) in observed survival for each subgroup *c* at time t_2 in the external data. We then calculate the mean square error of prediction (MSE) (25) as the sum of the squared difference between predicted (model) differences and observed differences in the external data across each subgroup *c*, and the mean absolute deviation (MAD) (26) as the sum of the absolute (positive) difference between predicted (model) differences and observed differences in the external data for each subgroup *c*.

$$\text{MSE} = \Sigma_c(d_{pc} - d_{oc})^2$$

$$\text{MAD} = \Sigma_c|d_{pc} - d_{oc}|$$

Alternatively, summary survival statistics from the external data can be used to directly estimate survival beyond t_{max}. In the absence of contrary evidence, we assume a constant hazard over time after t_1 as recommended by others.(9,10) The failure probability F_{cd} over each of the *j* periods of length $(\frac{t_2-t_1}{j})$ comprising the interval (t_1,t_2) for each subgroup *c* and therapy *d* can then be written,

$$F_{cd} = 1 - \sqrt[j]{S_{2cd}/S_{1cd}}$$

where S_{1cd} and S_{2cd} are the reported survival at t_1 and t_2 for subgroup c and therapy d respectively, and j is the number of time periods between t_1 and t_2.(27) j would typically be chosen by the analyst to match the time cycle of the model. The resulting probabilities for each subgroup c can be inserted directly into a Markov model to extrapolate survival beyond t_{max}. Or we can exploit the linear relationship between the log cumulative hazard and subgroup covariates in the exponential specification to estimate the failure probability for each subgroup via complementary log-log regression.(28) The latter approach allows estimation of the uncertainty in the failure probability which can be propagated through a probabilistic analysis via the cholesky decomposition of the covariance-correlation matrix.

External IPD available on control

The second scenario assumes that IPD are available for the control therapy, in addition to summary statistics on treatment and control. We follow recommendations to model the combined clinical and external IPD.(14,29) We use a dummy variable to identify patients in the external IPD. Such an approach assumes that the shape of the hazard is the same across populations, and has been previously applied.(14) Differences between populations are assumed to have a proportional impact on the hazard, the log odds of the event or the acceleration of time dependent on model specification.

Again, the summary data can be used in two ways. In order to use the summary data as a model selection aid we calculate MSE and MAD (as above). The summary data can also be utilised in conjunction with models fitted to the combined clinical data and external IPD to directly extrapolate survival with an assumption of proportional hazards for the period beyond t_1. A hazard ratio (HR) can be estimated as,

$$HR = \{ \sum_c \ln(S_{2ct}/S_{1ct}) / \ln(S_{2cc}/S_{1cc}) \}/n_c$$

where S_{1ct} and S_{2ct} are the summary estimates of survival for the treatment subgroup *c* at time t_1 and t_2, respectively, S_{1cc} and S_{2cc} are the corresponding survival estimates for the control, and n_c is the number of subgroups.(28) The mean HR assumes no interaction between treatment effect and subgroup. A crude estimate of the uncertainty around the HR can be derived from the standard deviation of the subgroup estimates of the HR. Where an interaction is indicated separate means can be calculated by subgroup. The hazard for the treatment group beyond t_{max} can then be derived as the product of the HR and the predicted hazard for the control.(11,28) Such an approach necessitates the choice of a proportional hazards model for the combined clinical and external IPD and can introduce bias.(30)

Underlying assumptions

Table 1 summarises the role of external data and the assumptions made in each scenario. In the first scenario only summary data are available beyond t_{max}. Either, one assumes the relationship between the hazard and time after t_{max} can be correctly specified from the clinical data only. Or, one assumes the form of the hazard (i.e. constant) and utilizes the external summary data to estimate it. In the second scenario external IPD for the control therapy are used to specify the relationship between the baseline hazard and survival time beyond t_{max}. Either, one assumes the treatment effect can be adequately specified from the clinical data only. Or, one assumes that hazards are proportional for events occurring after t_1.

Case study

We have published a cost-effectiveness analysis comparing Cemented, Cementless and Hybrid technology for THR in elderly people with Osteoarthritis.(19) The lifetime of each technology is a key consideration in evaluating cost-effectiveness since failure of the THR necessitates costly and invasive revision surgery. The analysis utilized a Markov model of THR built in Excel and parameterised from three large observational databases to quantify the impact of THR failure on costs and outcomes (Figure 1). Quality of life in each health state was estimated from data on THR collected for the Patient Reported Outcome Measures programme.(31) Clinical data on THR lifetime were taken from the NJR, a large clinical database for joint replacement.(20) External data on THR lifetime were drawn from HES, an administrative database used primarily for reimbursement of publicly funded secondary care in England.(20) Additional transition parameters, such as post-operative mortality, were also estimated from HES.

We made some simplifications to facilitate comparisons in this study. Firstly, we compare only Cemented and Cementless THR. Second, we ignore any differences in post-operative quality of life (QOL) or length of stay across THR technologies. Third, we make no distinctions according to the reason for THR failure. Fourth, for clarity and simplicity we report only the results for the youngest age groups for whom extrapolation of THR survival has the largest impact. Markov model parameters are reported in Table S12 in the supplementary material.

For the purpose of this study the NJR constituted the clinical data on event rates (failure of THR). However, we exploited the linkage of data across the NJR and HES to identify THR failures in both datasets. We chose to limit the clinical data to the subset of linked NJR and HES records, and classify THR failure as the first recorded revision operation in either dataset. After exclusions for non-standard diagnoses or non-standard procedures the (linked) NJR data consisted of 76,587 Cemented THR with 964 failures (1.3%), and 48,271 Cementless THR with 924 failures (1.9%) from 2003 to 2009. After exclusions, external (HES) data consisted of 268,757 Cemented THR with 6049 failures (2.3%), and 91,993 Cementless THR with 2164 failures (2.4%) from 1997 to 2009.

Figure 2 provides Kaplan-Meier survival plots for men and women aged 55 to 65 with Cemented and Cementless THR derived from NJR data and from HES. Careful inspection reveals that hazards are not proportional across THR types, especially in men, as indicated by the narrowing or crossing of survival curves. The gradients of the curves fall initially to a minimum at around five years after surgery and then begin to rise, consistent with the classic 'bath-tub' or U-shaped hazard. The U-shape is evident from non-parametric estimates of the hazard function (supplementary material).

Histograms of failures over time in the first year demonstrate elevated hazard rates in the first few weeks after surgery, particularly for Cementless THR. (This can also be seen in the Figure 2 as the sharp early fall in survival.)

Specification of survival models

'Standard' parameterisations of event data, available in many statistical software packages, include: Weibull, Lognormal, Exponential, Gompertz, Loglogistic and Generalized Gamma.(32) Each places restrictions on the shape of the hazard. For the Weibull and Gompertz specifications the hazard is either monotonically increasing or monotonically decreasing. The lognormal and Loglogistic specifications allow a hazard that increases initially and then decreases. Only the Generalized Gamma allows a hazard that decreases initially and then increases. In practice, these functions may be a poor fit with the observed hazard.(9) More flexible specifications are possible including three parameter variations of the Weibull,(33) the Generalized *F*,(16) which nests the Generalized Gamma and mixture models.(34) However, these specifications may require bespoke coding and may encounter problems with convergence.

Two approaches which provide flexibility in modelling the hazard rate whilst remaining easy to implement in standard software packages are modelling the log cumulative hazard using restricted cubic splines,(35) and applying a piecewise constant specification.(36) The latter approach splits survival time into segments which allows the scale of the hazard to vary arbitrarily across segments. Typically piecewise segmentation is imposed on an exponential function which generates a constant

hazard on extrapolation. Replacing the exponential with a Weibull or Gompertz function allows flexibility over the observed time period and extrapolation of an increasing or decreasing hazard.

We selected survival parameterisations from the five standard specifications (Weibull, Lognormal, Gompertz, Loglogistic and Generalized Gamma), and a restricted cubic spline model. In addition, when modelling the combined clinical and external data on treatment and control we included a piecewise segmented model. We included age and sex as covariates, and models were stratified by THR (allowing non-proportional hazards). External IPD were identified using a dummy variable. Interaction terms were included where they led to improvements in model fit. Stata code used in survival modelling is provided in the supplementary material. Selection across specifications was based on measures of internal fit and, where relevant, measures of fit with the external summary data. We tabulate survival at t_1 and t_2, along with 95% confidence intervals, and provide plots of observed survival and predicted survival curves from selected model specifications in the supplementary material, but do not base model selection on them. The plots confirm the challenges in interpreting such visual data without ambiguity.(3) Following the recommendation of Bagust and Beale (9) we excluded the first 65 days follow-up prior to fitting all survival functions due to the highly elevated failure rates in this period. Revisions in the first 65 days were modelled on uncensored data using logistic regression.

Cost-effectiveness analysis

Lifetime costs and quality adjusted life expectancy for men aged 60 and women aged 60 were estimated using the Markov model of THR (figure 1). The model was fully probabilistic and results are reported as the INB of Cementless THR compared to Cemented THR. To calculate INB the incremental health outcome, ΔQ, is multiplied by the threshold willingness to pay value, λ, for a unit gain in health. After subtraction of incremental cost, ΔC, a positive INB indicates the intervention is cost-effective:

$$INB = \Delta Q^{*}\lambda - \Delta C$$

A threshold of £20,000 per QALY, which is commonly applied in the UK, was chosen.(37) To capture uncertainty in parameters each parameter was specified as a random variable, and the probabilistic model was run 1,000 times with parameter values drawn from the specified distributions prior to each run.(38) As the majority of model parameters, including all of the parameters for event rates, were estimated from regression models, we used the Cholesky decomposition of the covariance-correlation matrices to parameterise uncertainty.(39) Reported INB values are mean values derived from 1,000 model runs with 95% credible intervals determined from 2.5th percentile and 97.5th percentile of the distribution of INB values. In addition to reporting INB the model simulations were used to construct CEACs. These curves plot the proportion of simulations in which Cementless THR has a higher INB as the threshold, λ, is varied between zero and £50,000.

Evaluation of the performance of alternative uses of external data

The true survival curves for Cemented and Cementless THR are unknown. To estimate as accurately as possible we fit survival models to the complete clinical and external IPD on treatment *and* control, and apply these in the Markov model. The resulting INB and CEAC are considered the '*best estimates*'. We evaluate the performance of the alternative applications of external summary data (model selection and direct extrapolation) in the two scenarios in which external IPD are limited, by comparison of INB and the resulting inference on cost-effectiveness with the *best estimates*. We also consider how well, in comparison with the *best estimates*, the CEACs capture uncertainty.

Results

Assessment criteria (AIC, BIC, MSE and MAD) for survival models and the resulting INB with 95% credible intervals are reported in Table 2 for both subgroups and for each scenario with respect to external IPD availability.

'Best estimates' of survival

The flexible restricted cubic spline and piecewise constant survival models fitted to the combined clinical and external IPD on both treatment and control provide consistent results - cementless THR is not cost-effective in either subgroup. The CEACs and the 95% credible intervals demonstrate considerable uncertainty in this finding at a threshold of £20,000 per QALY for men aged 60 (supplementary material). Measures of internal fit indicate a much better fit to the data for the flexible models compared to any of the 'standard' models. The flexible models better capture the 'U' shaped hazard revealed by the additional external IPD, as can be seen in the hazard plots (supplementary material). The best fitting 'standard' models, Weibull and Gompertz, generate very different INB and opposing inference on cost-effectiveness.

External data in summary form only

Measures of fit with the summary data (MSE and MAD) and measures of model fit (AIC and BIC) favour the Gompertz specification, which generates the conclusion that Cementless THR is cost-effective in both subgroups. Inference from the remaining models is consistent and in line with inference from the *best estimates* of INB, but the associated CEACs and 95% credible intervals underestimate the uncertainty in the decision for men aged 60. INB is closer to the *best estimates* when extrapolation of survival is undertaken using the external data. Here, model selection for the clinical data has little impact, reflecting the relatively close fit of each of the models to the observed clinical data.

External IPD available on control

Measures of fit with the summary data (MSE and MAD) again favour the Gompertz specification. In contrast, measures of model fit favour the restricted cubic spline model. Neither model generates INB close to the *best estimates*. When an HR estimated from the summary data is used to estimate survival for the treated group beyond t_{max} INB estimates are closer to the *best estimates* for all three of the best fitting models (Weibull, Gompertz, restricted cubic spline) and uncertainty is better captured.

Discussion

In this study we have illustrated how external data might be exploited in the extrapolation of event data beyond maximum trial follow-up, either as a model selection aid or to directly extrapolate survival. In our case study, the value of external data for model selection using the measures MSE and MAD appears low. In both scenarios with incomplete external IPD, MSE and MAD fail to identify the best performing survival specification. There may be superior measures to MSE and MAD or potential improvements through modifications such as weighting by subgroup size. However, our results do not encourage further investigation in this area. In contrast, use of the summary data to directly extrapolate survival improves estimates of INB in each case. This suggests the value of external data, in our case study, lies in its use to directly specify survival or the treatment effect beyond t_{max} rather than as a model selection aid. Further case studies would help to establish the generalizability of this finding. We have considered external data at a single time point beyond t_{max}. However, the methods we applied are generalizable to external summary data at multiple time points, either by assuming a constant hazard between time points or fitting a parametric model using regression.(40) Where such data is taken from multiple sources it may be helpful to combine survival estimates prior to incorporation into the framework suggested here.(41) Our analysis could also be extended to the comparison of multiple treatments. Calculation of F_{cd} from summary statistics for multiple treatments is straightforward. Where external IPD is available on one treatment it would be straightforward to calculate HRs for multiple comparators against that treatment.

Extrapolation of survival by calculation of F_{cd} from summary statistics requires an assumption of a constant hazard. The appropriateness of this assumption depends on the underlying biological mechanism driving the event rate. The hazard for all-cause mortality exhibits a well-known 'bath-tub' shape; extrapolation using age-specific population data is likely to provide better long term prediction (10). However, where disease specific mortality is high, as commonly observed in cancer trials, an assumption of a constant hazard may be reasonable in the absence of contrary evidence. Indeed in this context, Bagust and Beale recommend '…the exponential distribution should be considered the default parametric function for long-term survival projection…'(9) Considerations of the plausibility of extrapolations should be tempered by a focus on parsimony and accurately estimating the *difference* in survival between treatment and control.

Measures of model fit to the clinical data are poor guides to the appropriateness of extrapolations,(9) and do not prioritise relative survival between treatment and control. Hence they are of limited use in selecting model specifications for incremental analysis. Likewise, visual inspection of predicted survival by treatment group is rarely unambiguous. Inclusion of external IPD on the control therapy improves the *overall* accuracy of extrapolated. However, INB is influenced by *relative* survival, and estimation of the treatment effect is still dependent on the clinical data. In our case study, application of a treatment effect beyond t_{max} estimated from summary data on treatment and control improved the estimates of INB for all three models considered. Weighting the estimate of treatment effect by subgroup size may offer further improvements, but the selection of weights is not straightforward and further research is required. Our case study suggests the value of external IPD on the control therapy in isolation may be limited for the estimation of INB. Comparative data on treatment and control is of greater value, even when limited to summary statistics, and the publication of such data should be encouraged.

Modelling survival using restricted cubic splines has been recommended for extrapolation of time to event data in CEA.(10, 42,43) We illustrate an alternative flexible approach – piecewise segmentation of follow-up time in combination with a Gompertz specification. This approach provides flexibility to fit the observed hazard without constraining extrapolation with a constant hazard. Selection of piecewise

segments can be guided by measures of internal fit, or visual inspection of the hazard estimated without parameterisation using a kernel smoother.(44) We would recommend selection of segments to ensure a single segment covers the entire latter period for which the hazard is changing monotonically. The similarity in estimates of INB and the CEACs generated from analysis using the restricted cubic spline and the piecewise Gompertz model provides some reassurance that these models are generating the most appropriate inference. It is essential to fully recognise the impact of structural uncertainty in the selection of survival models and we would urge analysts to report and contrast results across multiple candidate models.

This study is based on real data which was applied to a longstanding clinical debate (cemented vs cementless THR). We made some simplifications, most notably the assumption of similar post-operative QOL across THR types, which increased the impact of survival on cost-effectiveness. Otherwise, the authenticity of the data was maintained. The analysis exploits a well-established Markov model of THR. The case study represents a challenging scenario for extrapolation - hazards are 'U' shaped with minima around five years, beyond which the clinical data was sparse. It is precisely such scenarios in which the judicious use of external data is important. We start from two premises: that all available IPD would be modelled; and that some data on control *and* treatment would be available. The latter might be obtained from expert opinion if unavailable from observational studies or case series reports.(45) We applied the commonly selected survival models and two flexible models that can be easily fitted in Stata, with minimal convergence problems.

Data external to the RCT can be used to directly estimate the relative effect of treatment on long-term survival, if the requisite data are available for both treatment and control regimens. However, the use of such observational data raises the risk of selection bias due to differences in prognostic factors between the treatment and control arms. Analysts must consider carefully the potential for confounding, and ideally adjust for all potential confounders.(46) An advantage of approaches using IPD for extrapolating event rates is that such data allow adjustment for measured confounders using regression. Greater care is needed where summary data are used due to the concern that the selection of patients into either the 'control' or 'treatment' is due to prognostic variables. Unless this is

fully recognised in the analysis, then this use of observational data in the extrapolation may lead to biased estimates of long-term effectiveness even where clinical trial data are available for the short term endpoints. More generally, with any use of observational data it is essential to consider the risk of unobserved confounding and to ensure that methods to adjust for observed confounders are appropriate. These risks of confounding should be explicitly acknowledged and explored through the use of sensitivity analysis wherever possible.(47) However, these risks should be considered in the context of potential concerns around the generalisability and extrapolation of randomised trial data.

Our analysis has some limitations. We assume that the flexible models applied to the combined clinical and external IPD on treatment and control correctly specify survival, and that the resulting INB is the best estimate. Negrin and co-workers took a similar approach in their evaluation of a proposed extension to Bayesian Model Averaging.(48) We have not undertaken simulation work to evaluate the reproducibility of our findings. Whilst it would have been possible to create bootstrap replicates of the clinical and external data, this would have necessitated the automated selection of survival models, which might have downplayed rather than highlighted the sometimes conflicting evidence guiding model selection. There is likely to be some heterogeneity in costs and outcomes between centres undertaking THR which our analysis did not capture. We consider a scenario where external summary data are limited to one time point beyond t_{max}. The diagnostic value of data at multiple time points would be increased. We utilise a clinical dataset, the NJR, that is observational rather than trial based; trial data are likely to consist of fewer observations but with a lower risk of selection bias.

Conclusions

Recent research activity has outlined methods for the inclusion of external data when extrapolating clinical time to event data. Whilst such methods can improve the plausibility of extrapolations, unless the data differentiate treatment and control, their value in estimating relative survival, and hence incremental benefit, may be limited. We have illustrated how summary (aggregate) data on treatment and control can be used either to assess models fitted to the clinical data, or to directly inform the

extrapolation of survival. The value of external summary data as a selection aid appears limited. In contrast, extrapolation of survival using external summary data can improve estimates of cost-effectiveness; it should be considered alongside the extrapolation of clinical data after evaluation of the risk of selection bias. The utility of additional IPD on the control therapy only may be limited for cost-effectiveness inference since relative rather than absolute survival has the largest impact on incremental analysis. Instead, analysts should focus efforts on retrieving any available long term data on survival for treatment and control.

Acknowledgements

We thank the patients and staff of all the hospitals in England, Wales and Northern Ireland who have contributed data to the National Joint Registry. We are grateful to the Healthcare Quality Improvement Partnership (HQIP), the NJR Steering Committee and staff at the NJR Centre for facilitating this work. The views expressed represent those of the authors and do not necessarily reflect those of the National Joint Registry Steering Committee or the Health Quality Improvement Partnership (HQIP) who do not vouch for how the information is presented.

Hospital Episode Statistics (HES) data have been re-used with the permission of the Health and Social Care Information Centre. All rights reserved. Copyright 2012. The HES data linked to NJR data were extracted by Northgate Information Solutions (UK) Ltd.

MP, RG and JvdM were supported by the Department of Health, United Kingdom (Commissioning Development). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. Available from: https://www.nice.org.uk/article/pmg9/resources/non-guidance-guide-to-the-methods-of-technology-appraisal-2013-pdf
2. Tappenden P, Chilcott J, Ward S, et al. Methodological issues in the economic analysis of cancer treatments. Eur J Cancer. 2006;42(17):2867-75.
3. Latimer NR. Survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data inconsistencies, limitations, and a practical guide. Med Decis Making. 2013;33(6):743-54.
4. Ishak KJ, Kreif N, Benedict A, Muszbek N. Overview of parametric survival analysis for health-economic applications. Pharmacoeconomics. 2013;31(8):663-75.
5. Kleinbaum DG, Klein M. Survival analysis: a self-learning text. Springer Science & Business Media; 2006.
6. Gerdtham UG, Zethraeus N. Predicting survival in cost-effectiveness analyses based on clinical trials. Int J Technol Assess. 2003;19(03):507-12.
7. Connock M, Hyde C, Moore D. Cautions regarding the fitting and interpretation of survival curves. Pharmacoeconomics. 2011;29(10):827-37.
8. Kim LG, Thompson SG. Uncertainty and validation of health economic decision models. Health Econ. 2010;19(1):43-55.
9. Bagust A, Beale S. Survival Analysis and Extrapolation Modeling of Time-to-Event Clinical Trial Data for Economic Evaluation An Alternative Approach. Med Decis Making. 2014;34(3):343-51.
10. Jackson C, Stevens J, Ren S, et al. Extrapolating Survival from Randomized Trials Using External Data A Review of Methods. Med Decis Making. 2016 loi:0272989X16639900.
11. Green C, Bryant J, Takeda A, Cooper K, Clegg A, Smith AG, Stephens M. Bortezomib for the treatment of multiple myeloma patients. Health Technology Assessment. 2009;13(1).
12. Williams C, Brunskill S, Altman D, et al. Cost-effectiveness of using prognostic information to select women with breast cancer for adjuvant systemic therapy. Health Technology Assessment. 2006;10(34).
13. Rogowski W, Burch J, Palmer S, et al. The effect of different treatment durations of clopidogrel in patients with non-ST-segment elevation acute coronary syndromes: a systematic review and value of information analysis. Health Technology Assessment. 2009;13(31).
14. Demiris N, Sharples LD. Bayesian evidence synthesis to extrapolate survival estimates in cost-effectiveness studies. Stat Med. 2006;25(11):1960-75.
15. Henriksson M, Epstein DM, Palmer SJ, et al. The cost-effectiveness of an early interventional strategy in non-ST-elevation acute coronary syndrome based on the RITA 3 trial. Heart. 2008;94(6):717-23.
16. Jackson CH, Sharples LD, Thompson SG. Survival models in health economic evaluations: balancing fit and parsimony to improve prediction. The international journal of biostatistics. 2010;6(1).
17. O'Hagan A, Buck CE, Daneshkhah A, et al. Uncertain judgements: eliciting experts' probabilities. John Wiley & Sons; 2006.
18. Latimer N. National Institute for Health and Care Excellence Decision Support Unit. Document 14: Survival analysis for economic evaluations alongside clinical trials – extrapolation with patient-level data. 2011. Available from: http://www.nicedsu.org.uk/NICE%20DSU%20TSD%20Survival%20analysis.updated%20March%202013.v2.pdf [Accessed January 23rd, 2017].
19. Pennington M, Grieve R, Sekhon JS, et al. Cemented, cementless, and hybrid prostheses for total hip replacement: cost effectiveness analysis. BMJ 2013;346:f1026
20. National Joint Registry for England, Wales, Northern Ireland and the Isle of Mann. http://www.njrcentre.org.uk/njrcentre/default.aspx [Accessed January 23rd, 2017].
21. Health and Social Care Information Centre. Hospital Episode Statistics. http://www.hscic.gov.uk/hes [Accessed January 23rd, 2017].

22. Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC medical research methodology. 2012;12(1):9.
23. Brier GW. Verification of forecasts expressed in terms of probability. Monthly weather review. 1950 Jan;78(1):1-3.
24. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361-87.
25. Allen DM. Mean square error of prediction as a criterion for selecting variables. Technometrics. 1971;13(3):469-75.
26. Pham-Gia T, Hung TL. The mean and median absolute deviations. Mathematical and Computer Modelling. 2001;34(7-8):921-36.
27. Kirkwood BR, Sterne JAC Survival analysis: displaying and comparing survival patterns. In:Essential Medical Statistics. 2003;272–286.
28. Perneger TV. Estimating the relative hazard by the ratio of logarithms of event-free proportions. Contemporary clinical trials. 2008;29(5):762-6.
29. Benaglia T, Jackson CH, Sharples LD. Survival extrapolation in the presence of cause specific hazards. Stat Med. 2015;34(5):796-811.
30. Guyot P, Welton NJ, Ouwens MJ, Ades AE. Survival time outcomes in randomized, controlled trials and meta-analyses: the parallel universes of efficacy and cost-effectiveness. Value In Health. 2011;14(5):640-6.
31. Health and Social Care Information Centre. Patient Reported Outcome Measures. http://www.hscic.gov.uk/proms [Accessed January 23rd, 2017].
32. Collett D. Modelling survival data in medical research. CRC press. Boca Raton FL. 2003.
33. Pham H, Lai CD. On recent generalizations of the Weibull distribution. IEEE Transactions on Reliability. 2007;56(3):454-8.
34. Zhang Y. Parametric mixture models in survival analysis with applications. Doctoral dissertation. Temple University, Philadelphia; 2008.
35. Royston P, Lambert PC. Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model. Stata Press books. 2011.
36. Colvert RE, Boardman TJ. Estimation in the piece-wise constant hazard rate model. Commun Stat Theory. 1976;5(11):1013-29.
37. McCabe C, Claxton K, Culyer AJ. The NICE cost-effectiveness threshold. Pharmacoeconomics. 2008;26(9):733-44.
38. Claxton K, Sculpher M, McCabe C, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ. 2005;14(4):339-47.
39. Briggs AH, Claxton K, Sculpher MJ. Decision modelling for health economic evaluation. Handbooks in Health Economic Evaluation; Oxford University Press. 2006.
40. Hoyle MW, Henley W. Improved curve fits to summary survival data: application to economic evaluation of health technologies. BMC medical research methodology. 2011;11(1):139.
41. Combescure C. A review of methods for meta-analysis of aggregated survival data (Doctoral dissertation, University of Geneva). Available at: file:///C:/Users/k1509773/Downloads/unige_43531_attachment01.pdf [Accessed May 3rd, 2017].
42. Davies A, Briggs A, Schneider J, et al. The ends justify the mean: outcome measures for estimating the value of new cancer therapies. Health Outcomes Research in Medicine. 2012 Feb 29;3(1):e25-36.
43. Guyot P, Welton NJ, Beasley M, Ades AE. Extrapolation of trial-based survival curves using external information. Value in Health. 2014;17(7):A326.
44. Silverman BW. Density estimation for statistics and data analysis. CRC press. 1986.
45. Hadorn D, Kvizhinadze G, Collinson L, Blakely T. Use of expert knowledge elicitation to estimate parameters in health economic decision models. Int J Technol Assess Health Care. 2014;30(04):461-8.
46. Faria R, Alava MH, Manca A, Wailoo AJ. National Institute of Health and Care Excellence Decision Support Unit. Technical Support Document 17 The use of observational data to inform estimates of treatment effectiveness in technology appraisal: Methods for comparative

individual patient data. 2015. Available from: http://scharr.dept.shef.ac.uk/nicedsu/technical-support-documents/observational-data-tsd/ [Accessed June 22nd, 2017].
47. Kreif N, Grieve R, Sadique MZ. Statistical methods for cost-effectiveness analyses that use observational data: A critical appraisal tool and review of current practice. Health Econ. 2013;22(4):486-500.
48. Negrín MA, Nam J, Briggs AH. Bayesian Solutions for Handling Uncertainty in Survival Extrapolation. Med Decis Making. 2016. doi:0272989X16650669.

Tables

	Data sources for survival		Assumptions on extrapolation of baseline hazard and treatment effect
	During trial follow-up	Beyond trial follow-up	
External data in summary form only: used for model selection	Clinical data	Clinical data	Hazard and treatment effect correctly specified from clinical data
External data in summary form: used to extrapolate survival	Clinical data	External summary data	Constant hazard beyond t_{max}
External IPD on control: summary data used for model selection	Clinical and external data (control)	External data (control)	Treatment effect correctly specified from clinical data
External IPD on control: summary data used to extrapolate survival	Clinical and external data (control)	External data (control) and summary data	Hazards proportional beyond t_{max}

Table 1 Data sources and assumptions underpinning alternative methods applied to available external data

	Measures of model fit		Measure of fit with external data		INB at £20,000 per QALY (95% credible interval)	
Survival Parameterization	AIC	BIC	MSE ($x10^6$)	MAD ($x10^4$)	Men aged 60	Women aged 60
Models fitted to combined clinical and external data on treatment and control ('best estimates')						
Weibull	81829	81937	429	429	-673 (-1,163 to -172)	-1,424 (-1,959 to -916)
Gompertz	81830	81938	**173**	**288**	1,829 (954 to 2,747)	1,276 (289 to 2,349)
Restricted Cubic Spline[a]	**81640**	**81877**	232	325	-278 (-1,264 to 735)	-1,334 (-2,480 to -256)
Piecewise Gompertz[b]	81651	82008	293	344	-164 (-1,053 to 747)	-1,439 (-2,507 to -437)
Models fitted to clinical data and used to extrapolate survival						
Weibull	18113	18184	737	585	-1,407 (-1,886 to -915)	-1,366 (-1,797 to -912)
Gompertz	**18094**	**18165**	**159**	**269**	250 (-1,104 to 3,058)	348 (-1,003 to 3,446)
Lognormal	18096	18167	451	431	-1,203 (-1,535 to -841)	-1,212 (-1,520 to -884)
Loglogistic	18113	18184	610	474	-1,241 (-1,754 to -718)	-1,267 (-1,763 to -777)
Generalized Gamma	18112	18192	379	373	-1,381 (-2,171 to -648)	-1,358 (-2,116 to -741)
Restricted Cubic Spline[c]	18101	18181	619	506	-1,346 (-1,705 to -983)	-1,294 (-1,685 to -939)
Models fitted to clinical data with extrapolation of survival using the summary external data						
Gompertz	As	above			165 (-1,062 to 1,618)	-1,278 (-2,571 to 105)
Lognormal					218 (-1,017 to 1,758)	-1,287 (-2,564 to 219)
Restricted Cubic Spline					112 (-1,176 to 1,491)	-1,339 (-2,502 to 26)
Models fitted to combined clinical data and external data on control and used to extrapolate survival						
Weibull	77103	77200	550	501	-766 (-1,411 to -145)	-1,634 (-2,228 to -1,109)
Gompertz	77070	77167	**385**	**425**	3,146 (2,095 to 4,290)	2,385 (1,417 to 3,436)
Restricted Cubic Spline[d]	**76953**	**77083**	810	592	-1,129 (-1,949 to -370)	-2,233 (-2,988 to -1,539)
Models fitted to combined clinical data and external data on control with treatment effect beyond t_{max} estimated from summary external data						
Weibull	As	above			-2 (-960 to 752)	-1,410 (-2,678 to -410)
Gompertz					780 (-799 to 2,176)	-1,280 (-3,351 to 312)
Restricted Cubic Spline					339 (-1,005 to 1,365)	-1,409 (-3,143 to -129)

[a]Proportional hazards specification with five degrees of freedom (four knots) for baseline hazard and with age and type of THR modelled as time dependent effects with five degrees of freedom. [b]Piecewise segmentation of time into yearly intervals for the first five years. [c]Proportional hazards specification with two degrees of freedom for baseline hazard and with age and type of THR modelled as time dependent effects with one degree of freedom. [d]Proportional hazards specification with four degrees of freedom for baseline hazard and with age and type of THR modelled as time dependent effects with one degree of freedom.

Table 2 Models optimised in the three scenarios exploiting clinical data only, external IPD on control only, and external data on treatment and control (representing the 'best estimate' of survival). Measures of fit using external summary data to aid model selection, and INB derived from applying the survival model in the CEA using summary data either to aid model selection or to extrapolate survival.

Figures

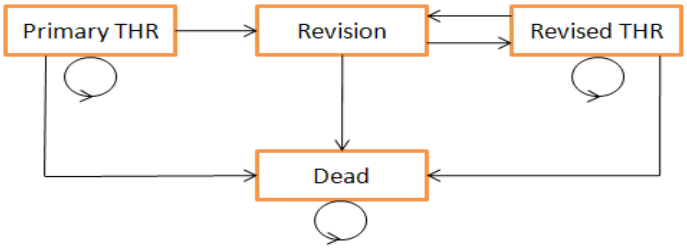


Figure 1. Markov model of total hip replacement.

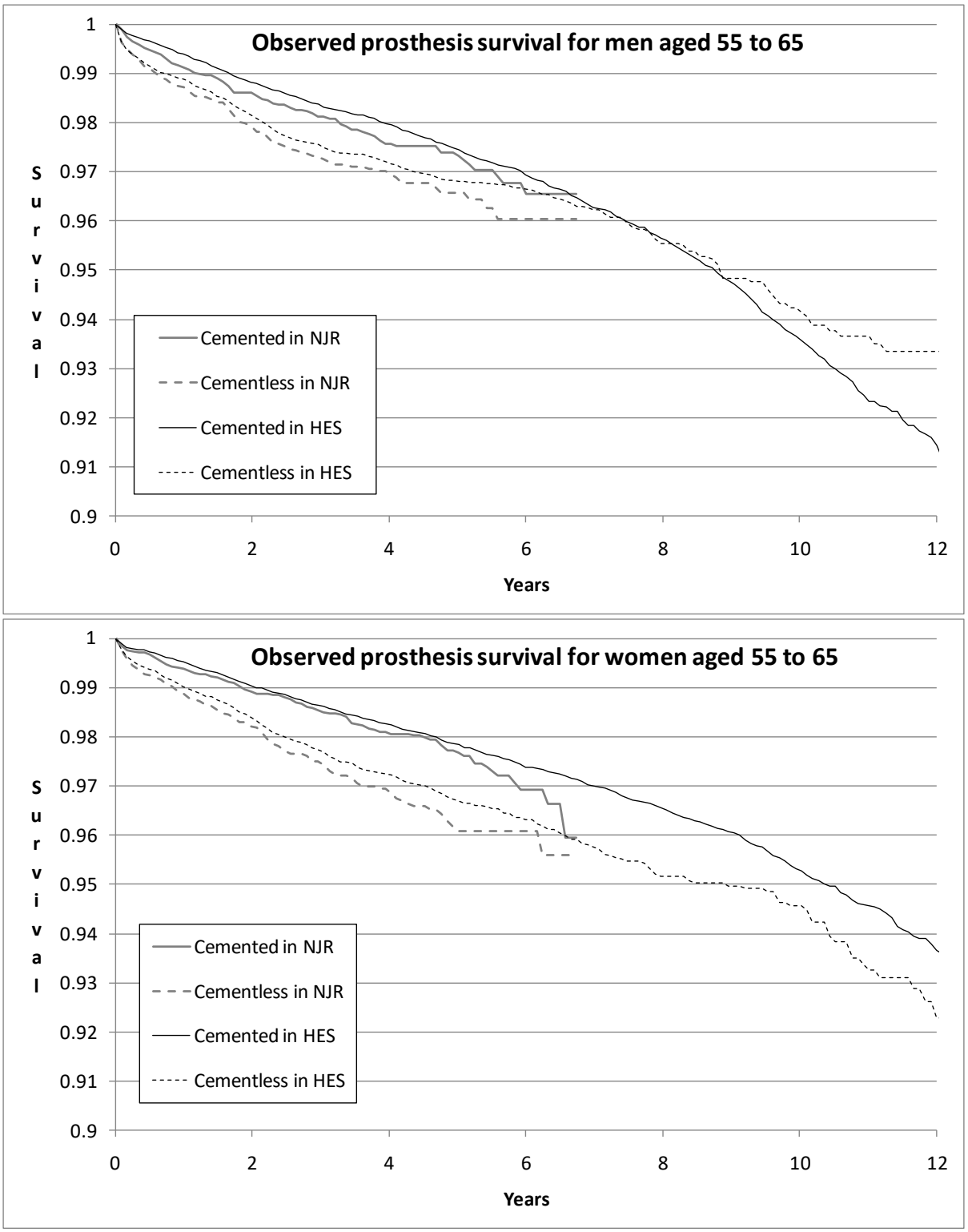


Figure 2. Observed prosthesis survival for men and women aged 55 to 65 recorded in NJR and HES

Supplementary material

Figure S1. For men aged 60, predicted survival from the best survival models fitted to the clinical data only; the clinical data and external IPD on control; the combined clinical and external IPD on treatment and control.

Figure S2. For women aged 60, predicted survival from the best survival models fitted to the clinical data only; the clinical data and external IPD on control; the combined clinical and external IPD on treatment and control.

Figure S3. Hazard function for cemented and cementless THR estimated using the Epanechnikov kernel smoother (non-parametric estimate) from the external data and adjusted for age and sex.

Figure S4. Hazard functions for men and women aged 60 estimated on the combined data for cemented and cementless THR using Weibull, Gompertz, Piecewise Gompertz and restricted cubic spline functions

Figure S5. CEACs generated from survival extrapolations based on the clinical data only

Figure S6. CEACs generated from applying survival models fitted to the clinical data for years 1-5 with simple extrapolation from the external data in summary form

Figure S7. CEACs generated from survival extrapolations based on the clinical data and external IPD on control.

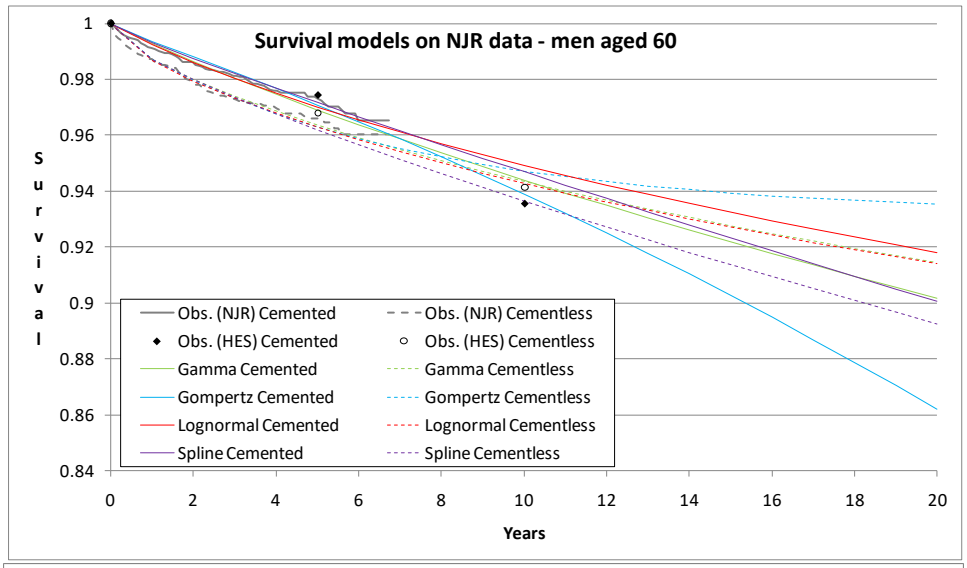
Figure S8. CEACs derived from combining clinical data and external IPD on control with extrapolation of treatment effect beyond 5 years derived from the external data in summary form

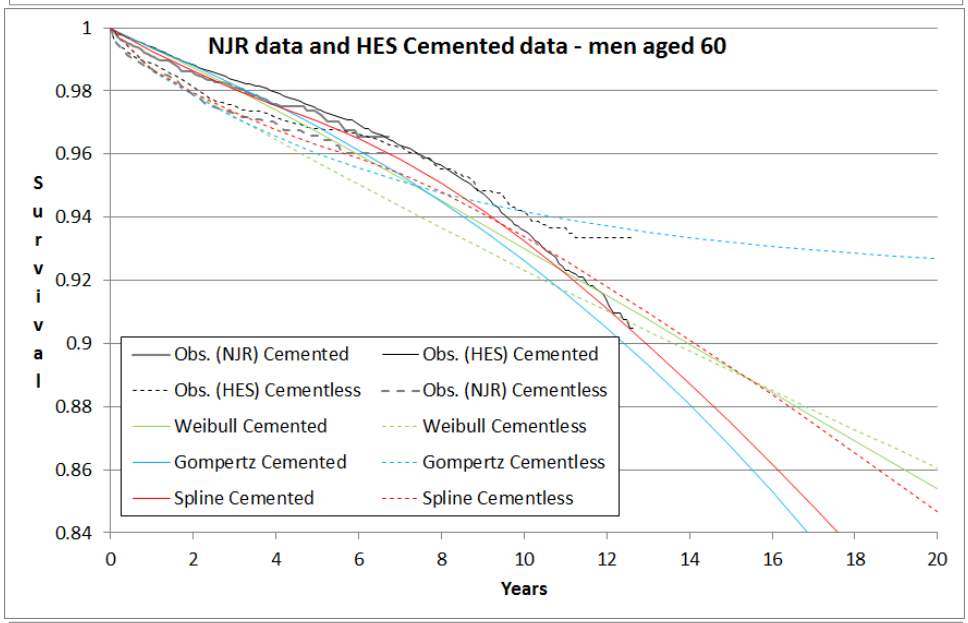
Figure S9. CEACs derived from survival extrapolations based on combined clinical and external IPD (best estimates)

Table S1. Measures of model fit for all models fitted in each scenario

Tables S2 to S11. Survival estimates at 5 and 10 years and 95% confidence intervals for each of the survival models.

Stata code to fit survival models


Survival models on NJR data - men aged 60
Survival
Years
1
0.98
0.96
0.94
0.92
0.9
0.88
0.86
0.84
0
2
4
6
8
10
12
14
16
18
20
Obs. (NJR) Cemented
Obs. (NJR) Cementless
Obs. (HES) Cemented
Obs. (HES) Cementless
Gamma Cemented
Gamma Cementless
Gompertz Cemented
Gompertz Cementless
Lognormal Cemented
Lognormal Cementless
Spline Cemented
Spline Cementless


NJR data and HES Cemented data - men aged 60
Survival
Years
1
0.98
0.96
0.94
0.92
0.9
0.88
0.86
0.84
0
2
4
6
8
10
12
14
16
18
20
Obs. (NJR) Cemented
Obs. (HES) Cemented
Obs. (HES) Cementless
Obs. (NJR) Cementless
Weibull Cemented
Weibull Cementless
Gompertz Cemented
Gompertz Cementless
Spline Cemented
Spline Cementless

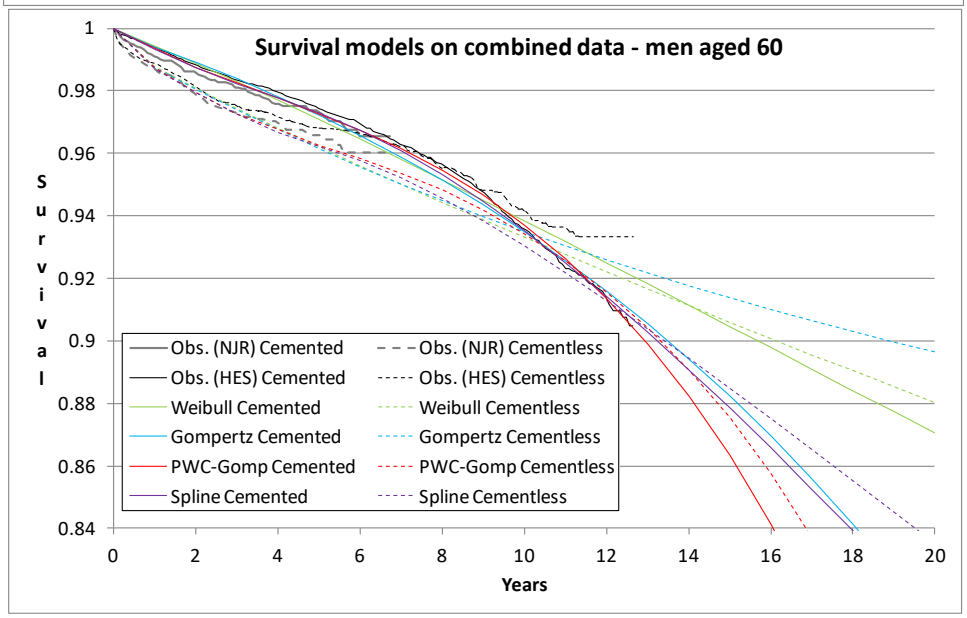
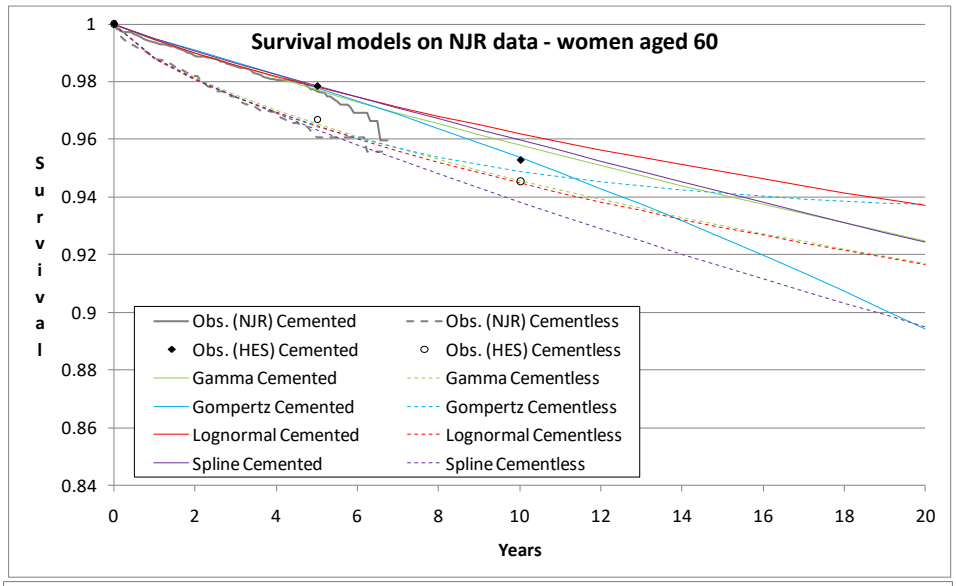
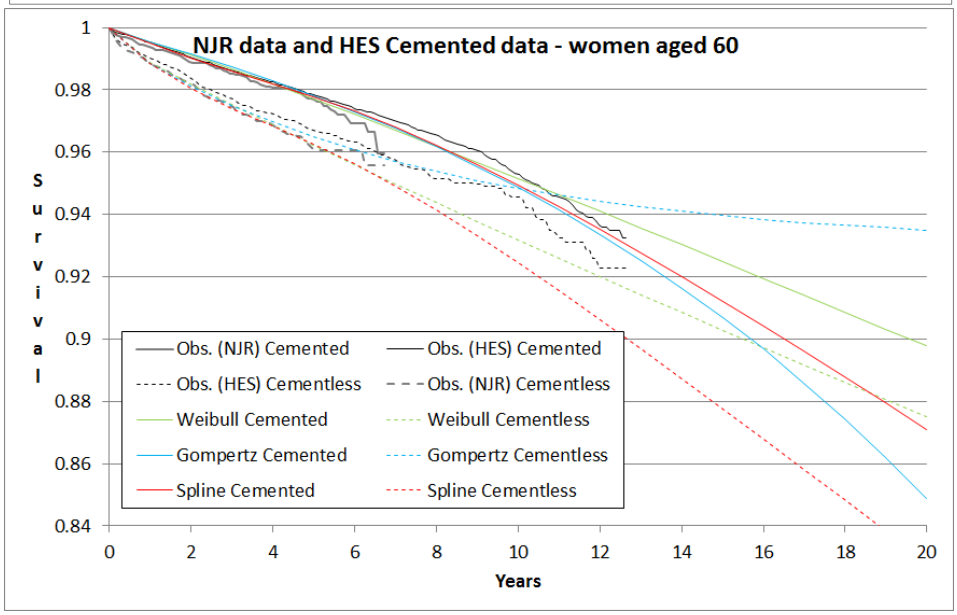

Survival models on combined data - men aged 60
Survival
Years
1
0.98
0.96
0.94
0.92
0.9
0.88
0.86
0.84
0
2
4
6
8
10
12
14
16
18
20
Obs. (NJR) Cemented
Obs. (NJR) Cementless
Obs. (HES) Cemented
Obs. (HES) Cementless
Weibull Cemented
Weibull Cementless
Gompertz Cemented
Gompertz Cementless
PWC-Gomp Cemented
PWC-Gomp Cementless
Spline Cemented
Spline Cementless

Figure S1. Fitted and observed survival, men aged 60.





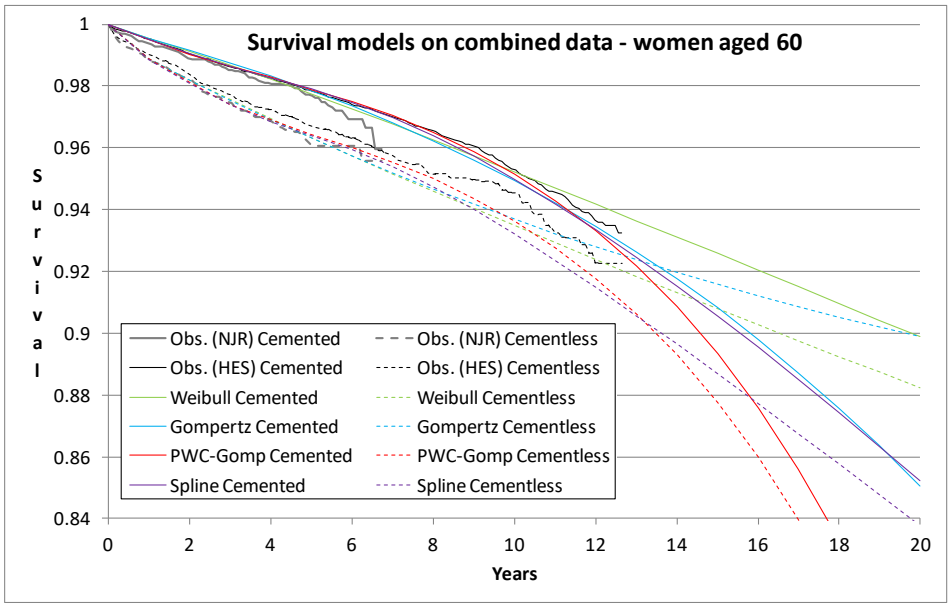


Figure S2. Fitted and observed survival, women aged 60.

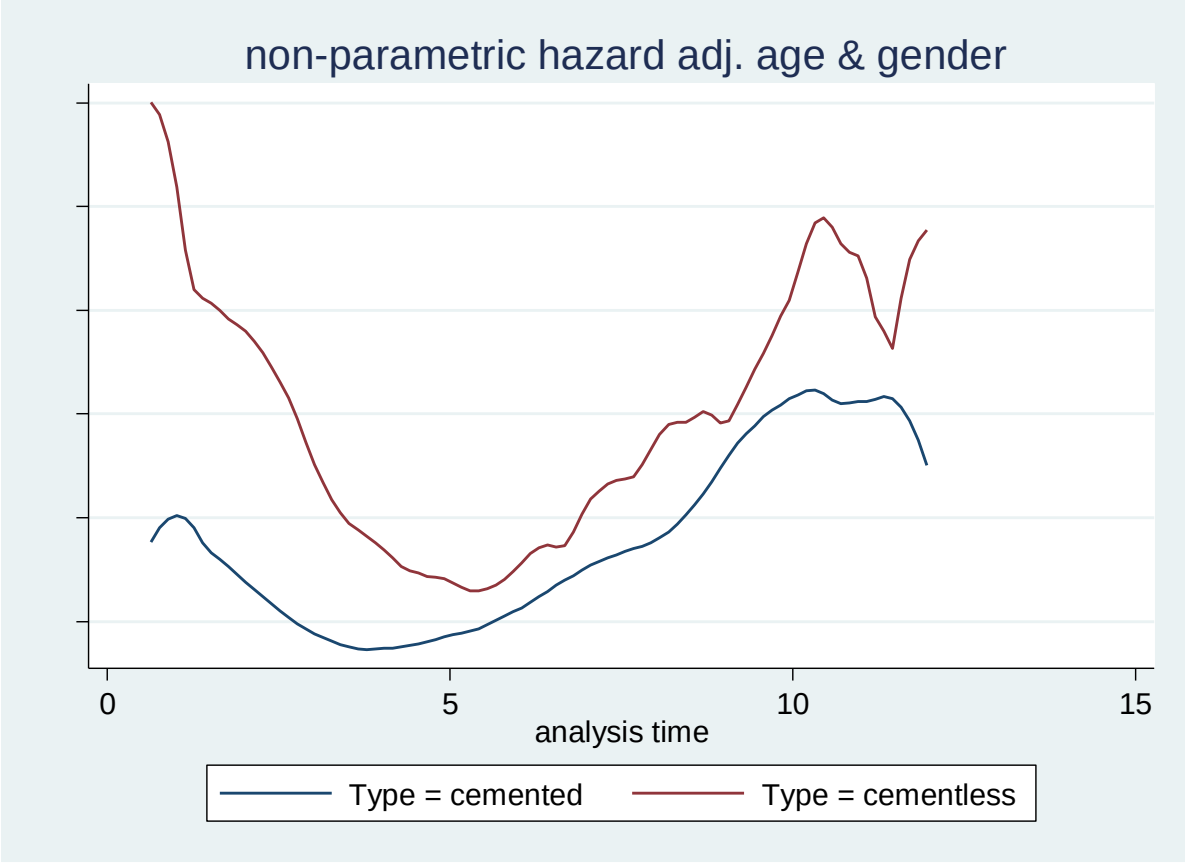


Figure S3. Non-parametric estimate of hazard for cemented and cementless THR

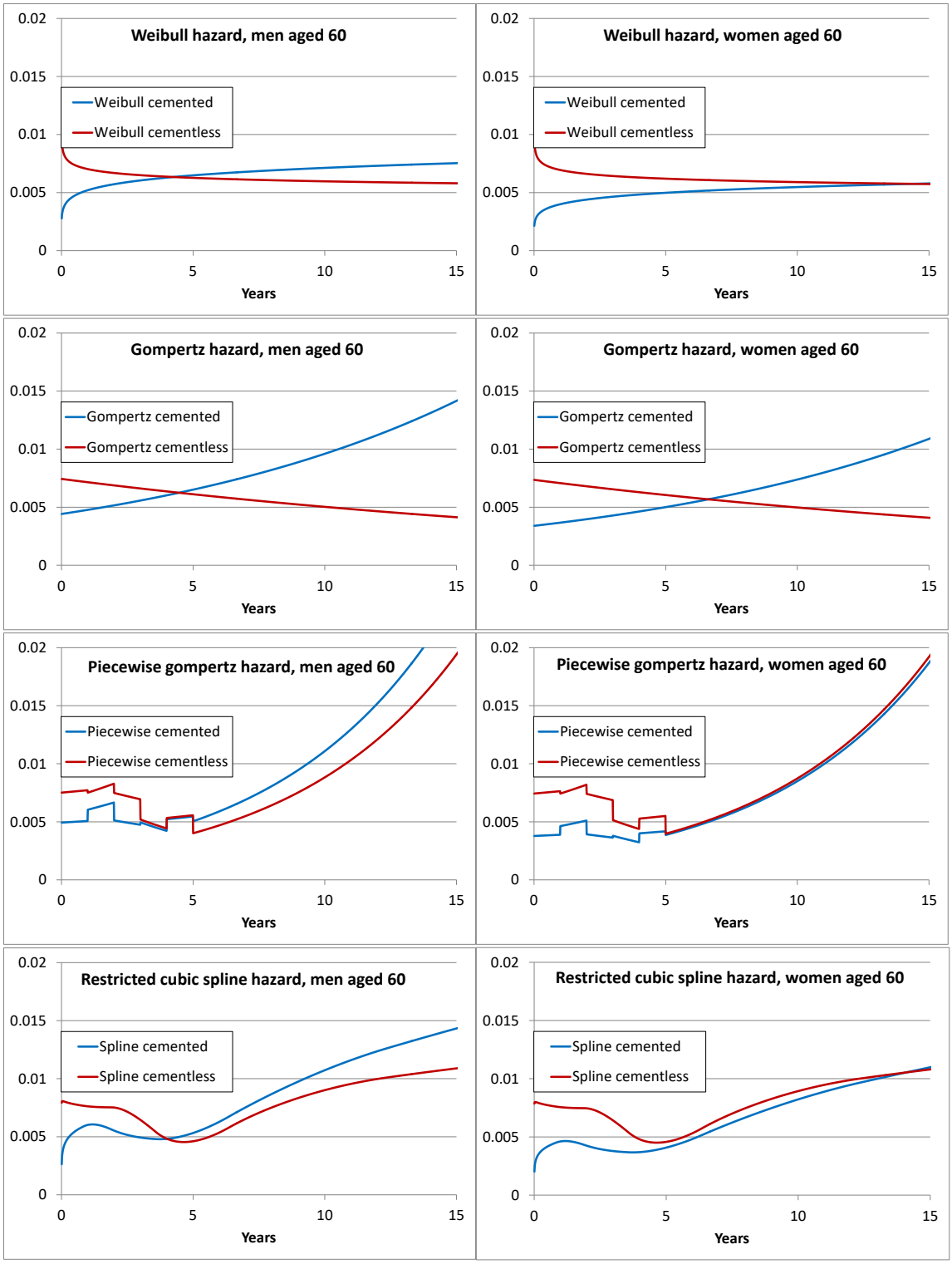


Figure S4. Hazards derived from survival models fitted to the combined data.

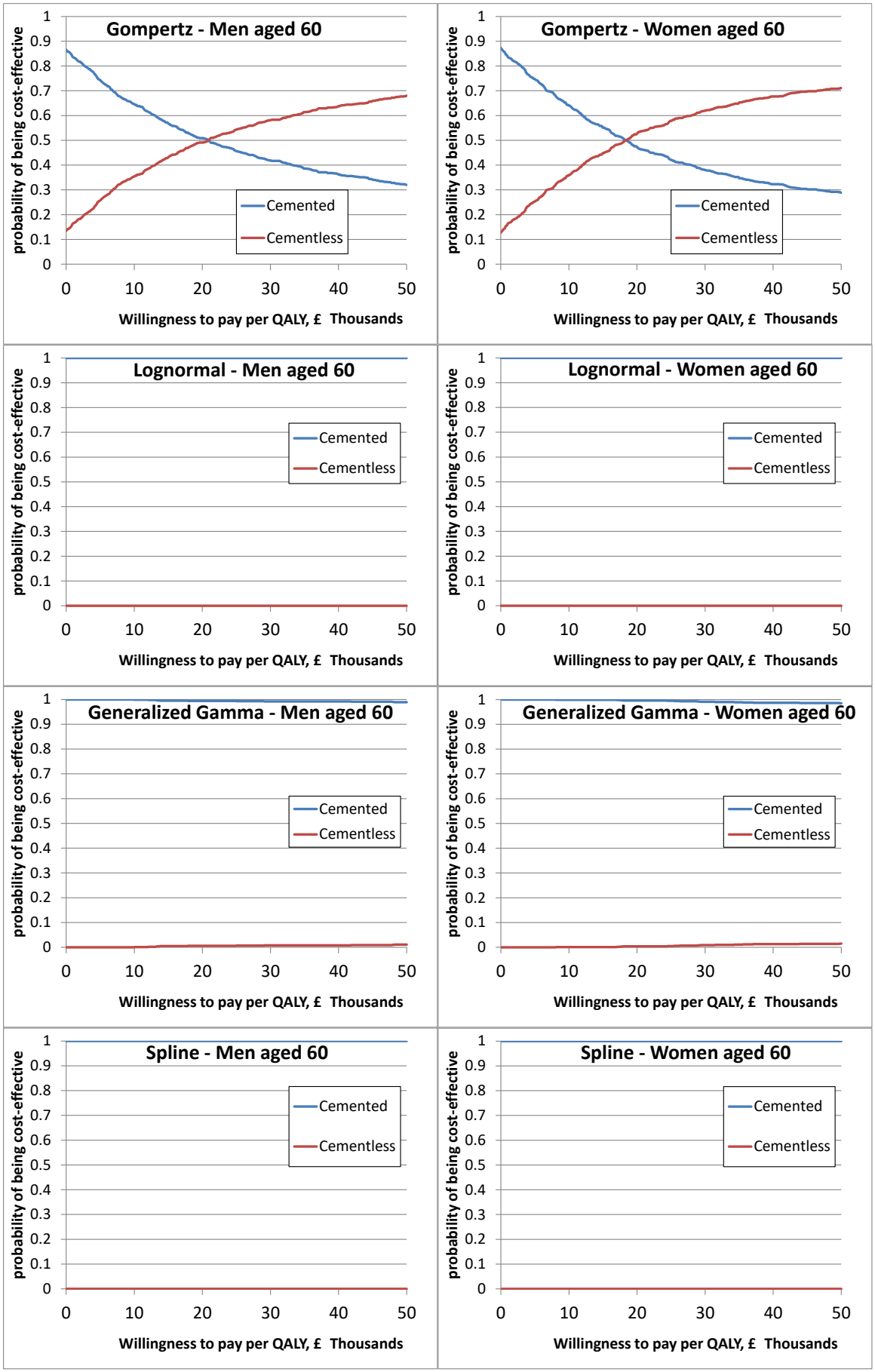


Figure S5. CEACs generated from applying survival models fitted to the clinical data only

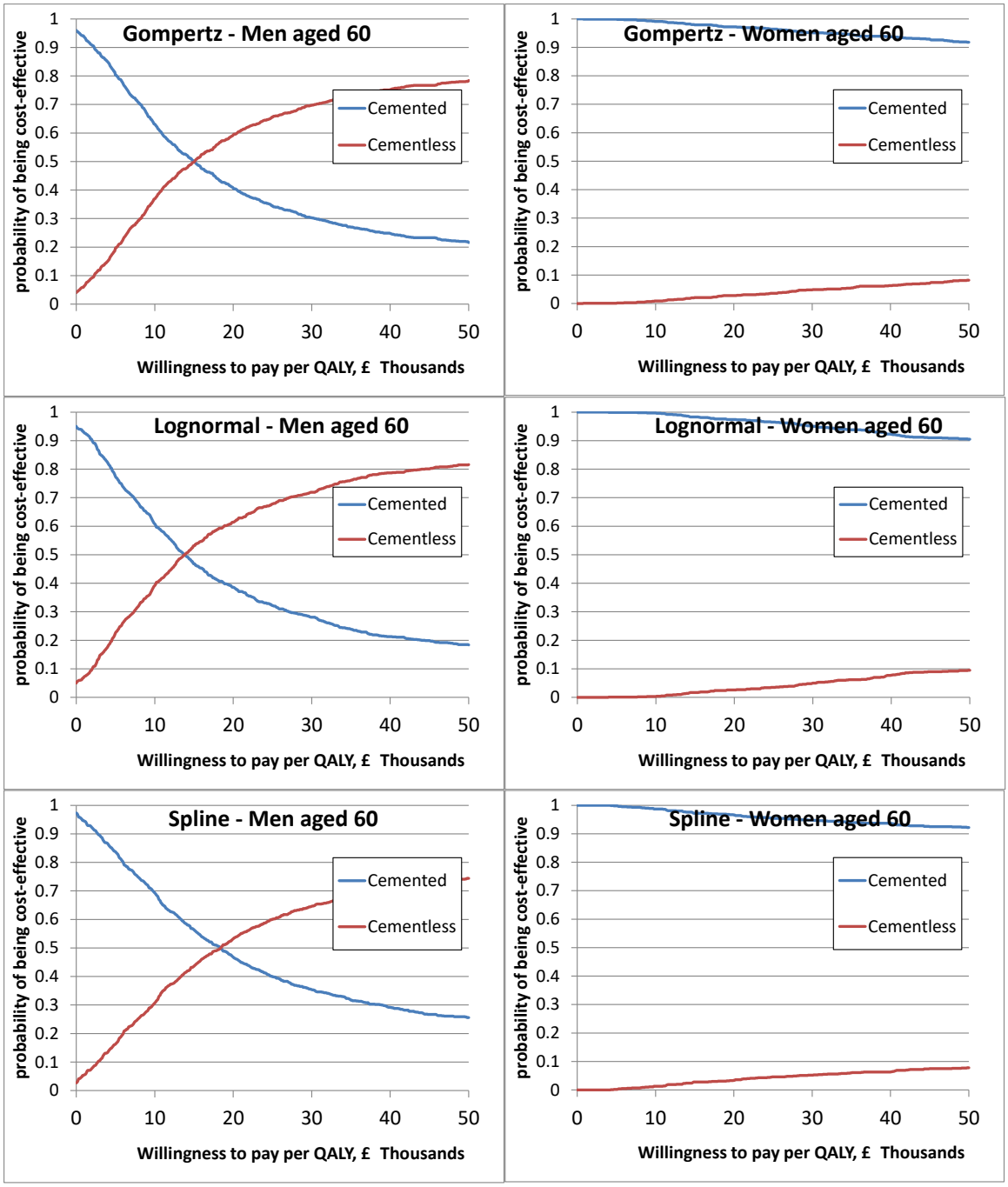


Figure S6. CEACs generated from applying survival models fitted to the clinical data for years 1-5 with simple extrapolation from the external data in summary form

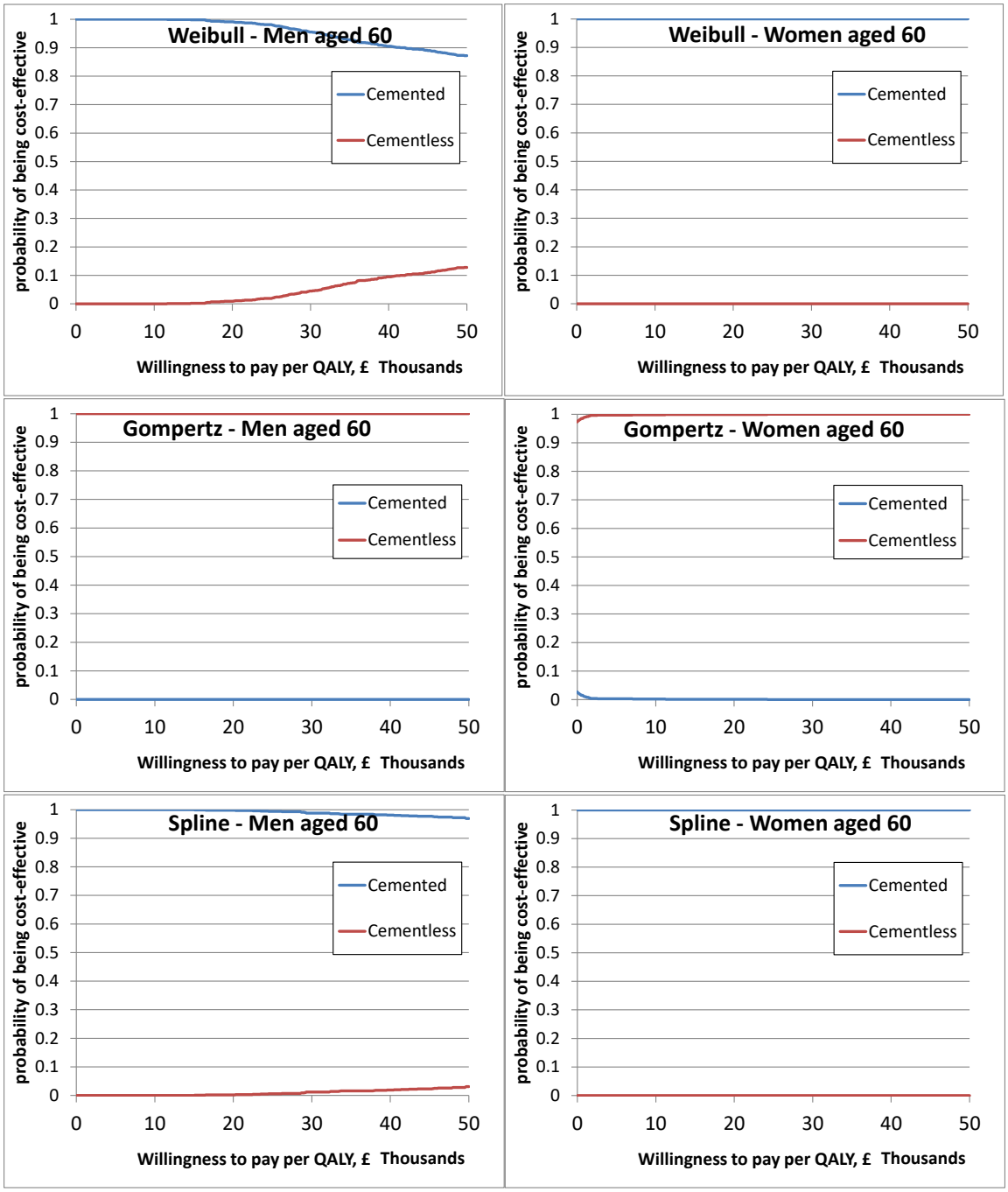


Figure S7. CEACs generated from survival extrapolations based on the clinical data and external IPD on control.

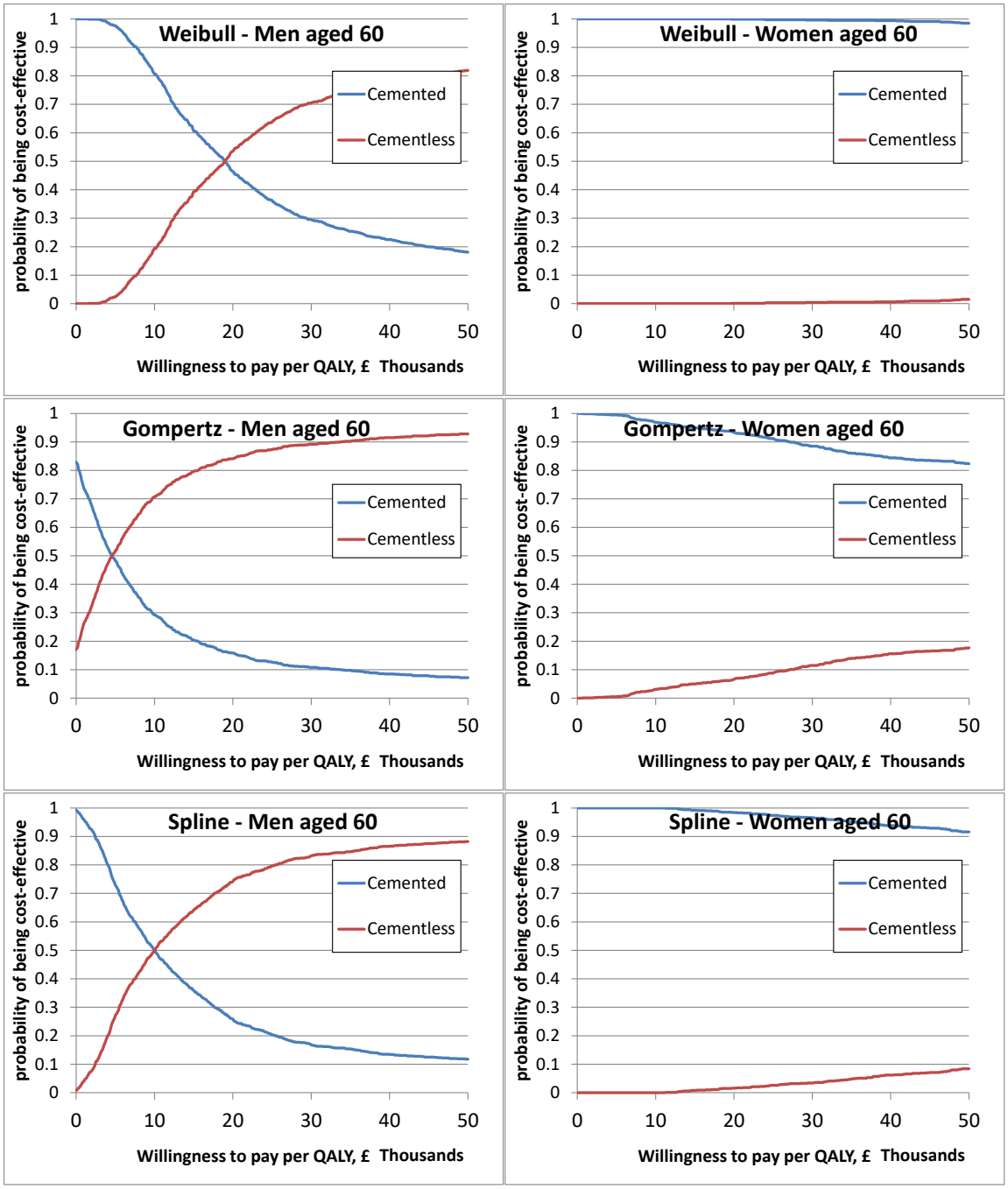


Figure S8. CEACs derived from combining clinical data and external IPD on control with extrapolation of treatment effect beyond 5 years derived from the external data in summary form

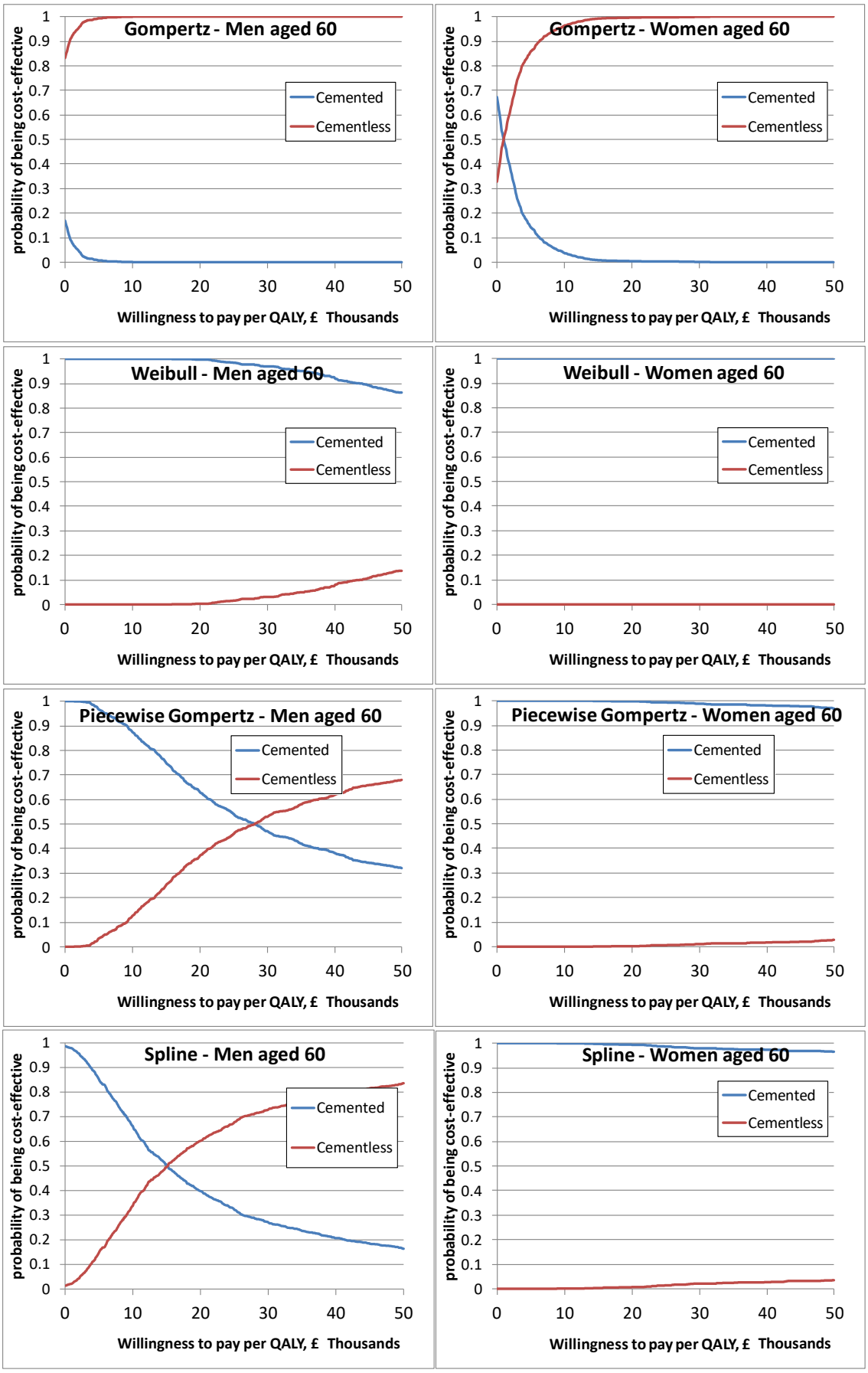


Figure S9. CEACs derived from survival models fitted to the combined clinical and external IPD

	Data available					
	Clinical only		Plus external IPD on control		Clinical and external IPD on treatment and control	
Model	AIC	BIC	AIC	BIC	AIC	BIC
Weibull	18113	18184	77103	77200	81829	81937
Gompertz	**18094**	**18165**	77070	77167	81830	81938
Lognormal	18096	18167	77261	77358	82001	82109
Loglogistic	18113	18184	77119	77217	81849	81957
Generalized Gamma	18112	18192	77228	77336	81837	81966
Restricted Cubic Spline	18101	18181	**76953**	**77083**	**81640**	**81877**
Piecewise Gompertz					81651	82008

Table S1. Measures of model fit for survival models fitted to the available data in each scenario

The following tables report survival estimates at 5 and 10 years with 95% confidence intervals using each of the survival models evaluated in the study. In addition, the absolute difference in survival (cemented – cementless) is reported.

	Weibull	Gompertz	Spline	Lognormal	Loglogistic	Gamma
Cemented 5yr	97.137%	97.086%	97.173%	97.008%	97.991%	96.913%
95%CI	96.736 to 97.451	96.696 to 97.432	96.799 to 97.503	96.535 to 97.378	97.678 to 98.222	96.059 to 97.265
Cemented 10 yr	94.447%	93.897%	94.691%	94.929%	96.137%	94.399%
95%CI	93.452 to 95.205	91.845 to 95.211	93.848 to 95.445	94.063 to 95.592	95.334 to 96.683	92.810 to 95.140
Cementless 5 yr	96.157%	96.299%	96.198%	96.285%	96.932%	96.363%
95%CI	95.700 to 96.573	95.833 to 96.692	95.726 to 96.592	95.825 to 96.663	96.598 to 97.196	95.509 to 96.662
Cementless 10 yr	93.380%	94.716%	93.651%	94.279%	94.868%	94.353%
95%CI	92.430 to 94.185	93.551 to 95.562	92.710 to 94.467	93.513 to 94.896	94.147 to 95.407	92.871 to 94.869
Difference 5 yr	0.980%	0.787%	0.976%	0.723%	1.059%	0.550%
95%CI	0.482 to 1.510	0.305 to 1.296	0.511 to 1.455	0.120 to 1.243	0.745 to 1.392	-0.235 to 1.362
Difference 10 yr	1.067%	-0.818%	1.039%	0.649%	1.269%	0.045%
95%CI	-0.012 to 2.166	-2.753 to 0.772	0.020 to 2.001	-0.387 to 1.532	0.453 to 2.058	-1.546 to 1.462

Table S2. Men aged 60, survival estimated using NJR data only.

	Weibull	Gompertz	Spline	Lognormal	Loglogistic	Gamma
Cemented 5yr	97.827%	97.786%	97.855%	97.816%	98.365%	97.710%
95%CI	97.013 to 97.944	97.470 to 98.030	97.600 to 98.090	97.469 to 98.078	98.114 to 98.533	97.013 to 97.944
Cemented 10 yr	95.787%	95.362%	95.974%	96.203%	96.848%	95.792%
95%CI	94.381 to 96.334	93.856 to 96.336	95.337 to 96.511	95.543 to 96.706	96.229 to 97.264	94.381 to 96.334
Cementless 5 yr	96.308%	96.450%	96.346%	96.458%	97.437%	96.539%
95%CI	95.684 to 96.835	96.064 to 96.838	95.962 to 96.704	96.040 to 96.803	97.132 to 97.666	95.684 to 96.835
Cementless 10 yr	93.569%	94.891%	93.834%	94.486%	95.697%	94.565%
95%CI	93.047 to 95.102	93.834 to 95.672	92.955 to 94.629	93.769 to 95.076	95.067 to 96.164	93.047 to 95.102
Difference 5 yr	1.519%	1.335%	1.509%	1.357%	0.927%	1.171%
95%CI	0.557 to 1.930	0.924 to 1.794	1.127 to 1.878	0.921 to 1.838	0.663 to 1.202	0.557 to 1.930
Difference 10 yr	2.218%	0.471%	2.139%	1.717%	1.151%	1.226%
95%CI	-0.118 to 2.573	-0.954 to 1.810	1.334 to 2.907	0.926 to 2.558	0.507 to 1.746	-0.118 to 2.573

Table S3. Women aged 60, survival estimated using NJR data only.

	Lognormal	Gompertz	Spline
Cemented 5yr	97.008%	97.086%	97.173%
95%CI	96.511 to 97.377	96.671 to 97.440	96.792 to 97.471
Cemented 10 yr	92.938%	93.013%	93.097%
95%CI	90.585 to 94.440	90.807 to 94.514	90.875 to 94.536
Cementless 5 yr	96.285%	96.299%	96.198%
95%CI	95.839 to 96.658	95.841 to 96.684	95.766 to 96.598
Cementless 10 yr	93.963%	93.976%	93.877%
95%CI	92.566 to 94.930	92.569 to 94.874	92.567 to 94.794
Difference 5 yr	0.723%	0.787%	0.976%
95%CI	0.109 to 1.277	0.309 to 1.314	0.492 to 1.455
Difference 10 yr	-1.024%	-0.963%	-0.781%
95%CI	-3.176 to 0.816	-3.125 to 0.664	-2.813 to 0.875

Table S4. Men aged 60, survival estimated using NJR data with extrapolation of survival beyond five years using a constant hazard calculated from HES data

	Lognormal	Gompertz	Spline
Cemented 5yr	97.816%	97.786%	97.855%
95%CI	97.492 to 98.061	97.483 to 98.031	97.585 to 98.081
Cemented 10 yr	94.910%	94.881%	94.949%
95%CI	93.247 to 95.953	93.189 to 96.004	93.269 to 96.037
Cementless 5 yr	96.458%	96.450%	96.346%
95%CI	96.056 to 96.796	96.053 to 96.762	95.922 to 96.707
Cementless 10 yr	93.647%	93.640%	93.539%
95%CI	92.020 to 94.663	91.867 to 94.701	91.850 to 94.594
Difference 5 yr	1.357%	1.335%	1.509%
95%CI	0.933 to 1.804	0.959 to 1.762	1.104 to 1.941
Difference 10 yr	1.263%	1.241%	1.410%
95%CI	-0.423 to 2.977	-0.405 to 2.941	-0.277 to 3.116

Table S5. Women aged 60, survival estimated using NJR data with extrapolation of survival beyond five years using a constant hazard calculated from HES data

	Weibull	Gompertz	Spline
Cemented 5yr	96.699%	96.866%	96.801%
95%CI	96.436 to 96.926	96.598 to 97.100	96.545 to 97.029
Cemented 10 yr	93.013%	92.619%	92.673%
95%CI	92.429 to 93.520	91.953 to 93.206	92.058 to 93.219
Cementless 5 yr	95.740%	96.023%	95.770%
95%CI	95.281 to 96.116	95.614 to 96.402	95.314 to 96.131
Cementless 10 yr	92.328%	94.178%	91.497%
95%CI	91.365 to 93.082	93.205 to 95.004	90.464 to 92.350
Difference 5 yr	0.959%	0.843%	1.031%
95%CI	0.494 to 1.459	0.395 to 1.286	0.592 to 1.509
Difference 10 yr	0.685%	-1.559%	1.176%
95%CI	-0.357 to 1.867	-2.619 to -0.460	0.138 to 2.312

Table S6. Men aged 60, survival estimated using NJR data and individual patient data from HES for cemented THR only

	Weibull	Gompertz	Spline
Cemented 5yr	97.706%	97.814%	97.782%
95%CI	97.530 to 97.869	97.648 to 97.974	97.611 to 97.924
Cemented 10 yr	95.150%	94.859%	94.928%
95%CI	94.767 to 95.529	94.457 to 95.249	94.513 to 95.269
Cementless 5 yr	96.233%	96.491%	96.264%
95%CI	95.904 to 96.502	96.173 to 96.768	95.961 to 96.545
Cementless 10 yr	93.171%	94.841%	92.432%
95%CI	92.418 to 93.778	93.948 to 95.523	91.616 to 93.177
Difference 5 yr	1.473%	1.323%	1.519%
95%CI	1.203 to 1.772	1.049 to 1.618	1.241 to 1.809
Difference 10 yr	1.979%	0.018%	2.496%
95%CI	1.344 to 2.714	-0.746 to 0.932	1.768 to 3.255

Table S7. Women aged 60, survival estimated using NJR data and individual patient data from HES for cemented THR only

	Weibull	Gompertz	Spline
Cemented 5yr	96.699%	96.866%	96.801%
95%CI	96.422 to 96.934	96.633 to 97.087	96.526 to 97.026
Cemented 10 yr	93.013%	92.619%	92.673%
95%CI	92.437 to 93.498	91.983 to 93.206	92.022 to 93.232
Cementless 5 yr	95.740%	96.023%	95.770%
95%CI	95.283 to 96.130	95.605 to 96.372	95.357 to 96.141
Cementless 10 yr	93.644%	93.601%	93.422%
95%CI	92.320 to 94.553	92.116 to 94.634	91.856 to 94.396
Difference 5 yr	0.959%	0.843%	1.031%
95%CI	0.462 to 1.450	0.428 to 1.343	0.595 to 1.506
Difference 10 yr	-0.631%	-0.982%	-0.749%
95%CI	-1.642 to 0.681	-2.089 to 0.561	-1.844 to 0.796

Table S8. Men aged 60, survival estimated using NJR data and individual patient data from HES for cemented THR with a hazard ratio estimated from the summary data for cemented and cementless THR from HES

	Weibull	Gompertz	Spline
Cemented 5yr	97.706%	97.814%	97.782%
95%CI	97.528 to 97.851	97.639 to 97.961	97.610 to 97.943
Cemented 10 yr	95.150%	94.859%	94.928%
95%CI	94.739 to 95.498	94.443 to 95.205	94.502 to 95.301
Cementless 5 yr	96.233%	96.491%	96.264%
95%CI	95.929 to 96.523	96.195 to 96.749	95.974 to 96.541
Cementless 10 yr	93.764%	93.631%	93.507%
95%CI	92.175 to 94.756	91.834 to 94.768	91.731 to 94.574
Difference 5 yr	1.473%	1.323%	1.519%
95%CI	1.195 to 1.761	1.041 to 1.611	1.240 to 1.785
Difference 10 yr	1.386%	1.228%	1.421%
95%CI	0.423 to 2.916	0.103 to 2.959	0.316 to 3.225

Table S9. Women aged 60, s survival estimated using NJR data and individual patient data from HES for cemented THR with a hazard ratio estimated from the summary data for cemented and cementless THR from HES

	Weibull	Gompertz	Spline	Piecewise-Gompertz
Cemented 5yr	97.088%	97.219%	97.275%	97.266%
95%CI	96.850 to 97.315	96.990 to 97.429	97.037 to 97.472	97.033 to 97.475
Cemented 10 yr	93.835%	93.478%	93.519%	93.705%
95%CI	93.275 to 94.320	92.915 to 94.018	92.979 to 94.030	93.165 to 94.167
Cementless 5 yr	96.200%	96.150%	96.231%	96.249%
95%CI	95.852 to 96.498	95.805 to 96.453	95.899 to 96.538	95.886 to 96.544
Cementless 10 yr	93.305%	93.498%	93.030%	93.432%
95%CI	92.619 to 93.901	92.802 to 94.150	92.272 to 93.699	92.627 to 94.048
Difference 5 yr	0.888%	1.069%	1.044%	1.018%
95%CI	0.510 to 1.276	0.706 to 1.438	0.705 to 1.421	0.636 to 1.408
Difference 10 yr	0.530%	-0.020%	0.489%	0.273%
95%CI	-0.269 to 1.372	-0.842 to 0.878	-0.330 to 1.327	-0.518 to 1.198

Table S10. Men aged 60, survival estimated using individual patient data from NJR and HES on cemented and cementless THR

	Weibull	Gompertz	Spline	Piecewise-Gompertz
Cemented 5yr	97.743%	97.848%	97.897%	97.891%
95%CI	97.552 to 97.907	97.670 to 98.013	97.727 to 98.059	97.712 to 98.049
Cemented 10 yr	95.218%	94.940%	94.987%	95.132%
95%CI	94.830 to 95.581	94.493 to 95.364	94.557 to 95.387	94.675 to 95.483
Cementless 5 yr	96.337%	96.311%	96.384%	96.409%
95%CI	95.990 to 96.652	95.978 to 96.581	96.074 to 96.673	96.107 to 96.704
Cementless 10 yr	93.469%	93.683%	93.205%	93.617%
95%CI	92.815 to 94.077	92.953 to 94.255	92.450 to 93.882	92.962 to 94.251
Difference 5 yr	1.406%	1.537%	1.513%	1.482%
95%CI	1.052 to 1.759	1.237 to 1.887	1.201 to 1.826	1.164 to 1.796
Difference 10 yr	1.748%	1.257%	1.782%	1.515%
95%CI	1.010 to 2.475	0.552 to 2.075	0.975 to 2.618	0.749 to 2.288

Table S11. Women aged 60, survival estimated using individual patient data from NJR and HES on cemented and cementless THR

Stata code to fit each of the survival models used in the study is reported below. The variable age is the patient's age, Male is a dummy for male sex, Unc is a dummy for Cementless THR, HES is a dummy for observations from the HES dataset. [Parameterization] is a place holder for the specification of the survival function: wei (weibull), gom (gompertz), lnorm (lognormal), llog (loglogistic).

Models fitted to the NJR data only:

```
stset time, failure(revised) origin(time 65) scale(365) id(id)
gen MaleUnc = Male*Unc
set more off
streg age Male Unc MaleUnc, d([Parameterization]) anc(age Unc)
matrix a = e(V)
matrix b=cholesky(a)
matrix list b

streg age Male Unc MaleUnc, d(ggamma) anc(age Unc) anc2(Unc)
matrix a = e(V)
matrix b=cholesky(a)
matrix list b
set more off

stpm2 age Male Unc MaleUnc, df(2) tvc(age Unc) dftvc(1) /*
*/scale(hazard) noorthog
matrix a = e(V)
matrix b = a["xb:age" . . "xb:_cons", "xb:age" . . "xb:_cons"]
matrix c = cholesky(b)
matrix list c
```

Models fitted to the NJR data and HES data on cemented THR only:

```
stset time, failure(revised) origin(time 65) scale(365) id(id)
gen MaleUnc = Male*Unc
set more off
streg age Male Unc MaleUnc HES, d([Parameterization]) anc(age Unc)
matrix a = e(V)
matrix b=cholesky(a)
matrix list b

streg age Male Unc MaleUnc HES, d(ggamma) anc(Unc) anc2(Unc)
matrix a = e(V)
matrix b=cholesky(a)
matrix list b
set more off

stpm2 age Male Unc MaleUnc HES, df(4) tvc(age Unc) dftvc(1) /*
*/scale(hazard) noorthog
matrix a = e(V)
matrix b = a["xb:age" . . "xb:_cons", "xb:age" . . "xb:_cons"]
matrix c = cholesky(b)
matrix list c
```

Models fitted to the NJR data and HES data on cemented and cementless THR:

```
stset time, failure(revised) origin(time 65) scale(365) id(id)
gen MaleUnc = Male*Unc
set more off
streg age Male Unc MaleUnc HES, d([Parameterization]) anc(age Unc)
matrix a = e(V)
matrix b=cholesky(a)
matrix list b

streg age Male Unc MaleUnc HES, d(ggamma) anc(age Unc) anc2(Unc)
matrix a = e(V)
matrix b=cholesky(a)
matrix list b
set more off

stpm2 age Male Unc MaleUnc HES, df(5) tvc(age Unc) dftvc(5) /*
*/scale(hazard) noorthog
matrix a = e(V)
matrix b = a["xb:age" . . "xb:_cons", "xb:age" . . "xb:_cons"]
matrix c = cholesky(b)
matrix list c

stsplit Duration, at(0.8219178 1.8219178 2.8219178 3.8219178
4.8219178)
gen year1 = 0
gen year2 = 0
gen year3 = 0
gen year4 = 0
gen year5 = 0

replace year1 = 1 if _t<=float(0.8219178)
replace year2 = 1 if _t>float(0.8219178) & _t<=float(1.8219178)
replace year3 = 1 if _t>float(1.8219178) & _t<=float(2.8219178)
replace year4 = 1 if _t>float(2.8219178) & _t<=float(3.8219178)
replace year5 = 1 if _t>float(3.8219178) & _t<=float(4.8219178)

gen year1Unc = year1*Unc
gen year2Unc = year2*Unc
gen year3Unc = year3*Unc
gen year4Unc = year4*Unc
gen year5Unc = year5*Unc

gen ageYear1 = age*year1
gen ageYear2 = age*year2
gen ageYear3 = age*year3
gen ageYear4 = age*year4
gen ageYear5 = age*year5

streg age Male Unc MaleUnc year1 year2 year3 year4 year5 year1Unc/*
*/ year2Unc year3Unc year4Unc year5Unc HES, d(gom) anc(age year1/*
*/ year2 year3 year4 year5 ageYear1 ageYear2 ageYear3 ageYear4/*
*/ ageYear5) nohr
matrix a = e(V)
matrix b=cholesky(a)
matrix list b
```