Using quadratic programming to reconstruct data from published survival and competing risks analyses

Andrew C. Titman*

School of Mathematical Sciences

Lancaster University, UK

Abstract

The ability to retrieve pseudo-individual patient data (IPD) from published survival study results is important to facilitate meta-analysis, evidence synthesis or secondary data analyses for the purpose of decision modelling for cost effectiveness analysis. While established methods exist for retrieving pseudo-IPD from Kaplan–Meier plots, these algorithms are not easily extendable to other types of survival data, nor do they allow all available information to be incorporated. An optimization-based approach is proposed where the task of reconstructing the IPD is formulated as a quadratic program (QP) with linear constraints. The method easily allows auxiliary information such as marked censoring times. Moreover, the same approach can be used to reconstruct patient-level competing risks survival data from published cumulative incidence functions. In simulation, the QP-based method is shown to outperform existing algorithms particularly when data on numbers at risk and marked censoring times are available. The methods are illustrated through reconstruction of data from a published study on patients with advanced stage follicular lymphoma.

Keywords: survival analysis; pseudo-individual patient data; competing risks analysis; quadratic programming

1 Introduction

It is often necessary or desirable to obtain approximate individual-level patient data (IPD) from a published study for the purposes of meta-analysis or evidence synthesis. For instance, meta-analysis procedures that use IPD may have superior properties to those relying only on summary level data.[1] Similarly, a published paper may not provide the required estimate or summary statistic for a particular analysis but clearly this could be computed were the IPD available. The availability of IPD also facilitates

*email: a.titman@lancaster.ac.uk

parametric modelling of the data which can be important for survival extrapolation in cost-effectiveness modelling. Nevertheless, it is often not possible to obtain the original data for reasons of patient or commercial confidentiality, or the process of obtaining the original data is too time consuming. In the context of time-to-event outcomes, the Kaplan–Meier estimate of the survivor function, which is almost always reported, conveys a substantial amount of information particularly if information on the number of patients at risk at some time points is also given. This has led to several proposals for algorithms to extract summary statistics or reconstruct *pseudo individual patient data* (pseudo-IPD) from the Kaplan–Meier curve.[2–5]

By far the most commonly used method for generating pseudo-IPD is that of Guyot *et al*, referred to as the iKM algorithm.[6] This involves eight distinct steps, where initial estimates of the number at risks, number of observed events and number censored at each time point are made and then subsequently refined. More recently Liu et al[7] proposed a modified iKM algorithm which improves the robustness and stability of the original iKM algorithm and is implemented within the *IPDfromKM* package in R.[8] Separately, Rogula et al[9] have proposed an algorithm of a similar nature to iKM which uses the information from the marked censoring times, or 'tick' marks, often displayed on published Kaplan–Meier curves. In simulation they show that this approach usually performs better than iKM when these tick marks are available. However, their method does not attempt to use information of total number of events or numbers at risk at intermediate times to refine the estimate.

While these existing algorithms have been shown to perform well in a wide range of situations, the underlying algorithms are somewhat *ad hoc* in nature which makes their modification to accommodate additional information, or their extension to other types of event history data, difficult.

Typically, the time points and survival decrements obtained through digitization will be subject to varying levels of error or omission depending on the image quality and the digitization method used. In contrast, it is usually reasonable to assume that the reported numbers of patients at risk and total number of events will be accurate. However, while the iKM and modified iKM algorithms use the information, they do not necessarily guarantee that the pseudo-IPD obtained will agree in terms of number of patients at risk and total number of events. Conversely, the method of Rogula *et al* does not use this information at all.

Many time-to-event endpoints are subject to competing risks, meaning that the appropriate quantity of interest should be cumulative incidence function (CIF) with respect to the event(s) of interest.[10] While non-parametric estimates of the CIFs of a similar nature to the Kaplan–Meier estimates can be produced, and are routinely reported in studies involving competing risks, there is currently no established method of reconstructing IPD from such estimates.

This paper presents a general framework for reconstructing IPD by expressing the procedure as a quadratic programming optimization problem. The remainder of the paper is as follows. Section 2 presents the quadratic programming framework in the context of reconstructing IPD from Kaplan–Meier curves. In Section 3, the approach is extended to provide a method to extract IPD from published cumulative incidence distribution curves in competing risks analyses. The methods are illustrated on published results from a study into long-term outcomes of patients with advanced follicular lymphoma in Section 4, and assessed via simulation in Section 5. The paper concludes with a discussion including consideration of potential extensions of the approach.

2 Quadratic programming framework

The overall aim of reconstructing the IPD is to find a set of data points that will return an estimate of the survival or cumulative incidence curve as close as possible to the observed estimate. As such, it can be viewed as an optimization procedure. In what follows, the problem is posed in terms of a quadratic program. Quadratic programming refers to optimization problems involving quadratic objective functions, for which there are efficient and well-established algorithms for solving quadratic programs with linear constraints.[11] Such problems can be defined as finding the vector $\mathbf{x}$ of length p that will

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\mathbf{x}^{'}\mathbf{Q}\mathbf{x} + \mathbf{q}^{'}\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} \end{aligned} \tag{2.1}$$

where $\mathbf{Q}$ is a $p \times p$ symmetric matrix, $\mathbf{q}$ is a vector of length p, $\mathbf{A}$ is an $m \times p$ matrix and $\mathbf{b}$ is a vector of length m.

We look to frame the problem of finding the number of events and number of patients censored at individual time points as a linearly constrained optimization problem, where the objective is to minimize the squared discrepancy between the observed decrement of the survivor function and the decrement implied by completed data. For competing risks data, the objective becomes the minimization of the squared discrepancy between the increments of the cumulative incidence curve and the increments implied by the completed data. Information provided on the number of patients at risk and total number of events is used to define linear constraints for the quadratic program.

2.1 Kaplan–Meier survival data

First consider the case discussed by previous authors of reconstructing IPD from a published Kaplan–Meier curve and the number of patients at risk at certain time points. In this case, the available information consists of extracted survival curve estimates $\{s_i, i = 1, \ldots, N_s\}$ at corresponding times

$\{t_i, i = 1, \ldots, N_s\}$ (representing co-ordinates from the digitized Kaplan–Meier curve) and numbers at risk $\{R_j, j = 0, 1, \ldots, N_r\}$ at corresponding times $\{\tau_j, j = 0, 1, \ldots, N_r\}$ where $\tau_0 = 0$ and $R_0 = N$ is the total number of patients at risk at time 0 (taken from information typically given below the time axis of the plot) where usually $N_r \ll N_s$. We assume that at any given time t_i, an unknown number of events, d_i, occurred and an unknown number of patients, c_i, were censored. Since a good reconstruction would ensure a close match between the 'observed' curve defined by s_i and the 'fitted' curve defined through the d_i and c_i, the aim is to minimize some measure of this discrepancy. The definition of the Kaplan–Meier estimator ensures that for all i we have the relationship

$$\frac{s_i}{s_{i-1}} = 1 - \frac{d_i}{r_i}$$

where $r_i = N - \sum_{j<i}(d_j + c_j)$ is the number at risk and $s_0 = 1$. Let $o_i = 1 - s_i/s_{i-1}$ then under a perfect reconstruction

$$o_i(N - \sum_{j<i}(d_j + c_j)) - d_i = 0 \tag{2.2}$$

for all $i = 1, \ldots, N_s$. This motivates using a criterion

$$W = \sum_{i=1}^{N_s} \left(o_i\{N - \sum_{j<i}(\tilde{d}_j + \tilde{c}_j)\} - \tilde{d}_i \right)^2 \tag{2.3}$$

to judge the fit of a given reconstruction with estimated event counts $\tilde{d}_j$ and $\tilde{c}_j$. This particular formulation is useful since each term to be squared in (2.3) is linear in the variables to be optimized ($\tilde{d}_i$ and $\tilde{c}_i$) and hence W is a quadratic form. Given the information on patients at risk, we have equality constraints of the form

$$\sum_{i \in \mathcal{C}_j} (\tilde{d}_i + \tilde{c}_i) = R_j - R_{j-1}, \text{ for } j = 1, \ldots, N_r \tag{2.4}$$

where $\mathcal{C}_j = \{i : \tau_{j-1} \leq t_i < \tau_j\}$. Note that, in some cases there may be no uncensored events between τ_{j-1} and τ_j, but nevertheless $R_j - R_{j-1} > 0$. To ensure that the equality constraints can be met in these cases, we include $\{\tau_1, \ldots, \tau_{N_r}\}$ in the set of candidate times at which a patient could be censored. This assumption differs from that of the iKM and modified iKM algorithms where only the decrement points are used to determine the censoring times. If information on the total number of events, N_E, is available a further equality constraint

$$\sum_{i=1}^{N_s} \tilde{d}_i = N_E,$$

should be included. In addition, since the Kaplan–Meier curve can only decrease at times corresponding to at least one observed event, we have inequality constraints $\tilde{d}_i \geq 1$ for $i \in \{j : o_j > 0\}$ and $\tilde{c}_i \geq 0$ for $i = 1, \ldots, N_s$.

Hence, ignoring the further condition that the $\tilde{d}_i$ and $\tilde{c}_i$ are integers, the minimization problem can be expressed as a quadratic program with linear constraints for which it is straightforward to find the

solution with standard software. For instance, in **R** the function *solve.QP* within the package *quadprog* will perform such optimization.[12] The explicit form of the quadratic program in terms of the quantities given in (2.1) is given in Appendix A.

Any proper reconstruction of the data requires that $\tilde{d}_i$ and $\tilde{c}_i$ are integers. If these integer constraints are imposed, then the formal minimization problem becomes substantially more difficult. The problem is then a mixed integer quadratic program (MIQP) which can be solved, for instance, via either branch and bound or outer-approximation algorithms.[13,14] Currently most software that can accommodate MIQP is commercial, although IBM-CPLEX[15] is available free for academic use and can be used within **R** using the *Rcplex*[16] or *cplexAPI*[17] packages. The main practical issue is that a MIQP is an NP-hard problem[18] for which no polynomial time algorithm is available and hence the computation time required to find the optimum can become prohibitive, particularly for curves with a large number of identified points.

For practical purposes, one can specify the maximum allowable computation time for the MIQP solver, with software giving the best solution found in that time, rather than the global optimum. Alternatively, it is straightforward to devise a heuristic that maps the continuous solution to an approximate integer solution. In principle, one can either first round the number of observed events, d_i, and then determine the censored observations c_i or instead round the censored observations first. Since for most time points $d_i \ll r_i$, the overall Kaplan-Meier estimate is more heavily influenced by changes in d_i than c_i. Hence, an approach that rounds the d_i first is considered. Let $\tilde{d}_i$ be the continuous solution for the number of events d_i. Then an integer estimate can be obtained through midpoint rounding of the cumulative sum of the observed events,

$$\hat{d}_i = \left\lfloor \frac{1}{2} + \sum_{j=1}^{i} \tilde{d}_j \right\rfloor - \left\lfloor \frac{1}{2} + \sum_{j=1}^{i-1} \tilde{d}_j \right\rfloor.$$

Since the continuous solution still constrains the aggregate number of events to be integer, the solutions often involve a total of $m + k$ events over m potential time points. The effect of the rounding scheme is to distribute the k events across these m times. However, the rounding potentially causes the $\hat{d}_i$ to violate the period constraints in (2.4). Let $\tilde{c}_i$ be the continuous solution for the number of censoring events then the equality constraints can be retained if, for each $i \in \mathcal{C}_j$, the number of censoring events are scaled by

$$V_j = \frac{(R_j - R_{j-1})}{\sum_{i \in \mathcal{C}_j} (\hat{d}_i + \tilde{c}_i)}, \; j = 1, \ldots, N_r.$$

Hence an integer solution can be obtained by taking

$$\hat{c}_i = \left\lfloor \frac{1}{2} + \sum_{k=1}^{i} V_{j(k)} \tilde{c}_k \right\rfloor - \left\lfloor \frac{1}{2} + \sum_{k=1}^{i-1} V_{j(k)} \tilde{c}_k \right\rfloor, \tag{2.5}$$

where $j(k) = \{j : k \in \mathcal{C}_j\}$ denotes the period within which time point k lies.

Incorporating information on censoring times

A particular advantage of the quadratic programming approach is the ease with which other information can be incorporated into the reconstruction. Published Kaplan–Meier curves often include marked censoring times, or 'tick' marks, to indicate the times at which at least one patient was censored. If it can be assumed that the digitization captures all of these points then it is desirable to constrain the data reconstruction to only permit censoring events at those time points. Conversely, unless the tick point coincides with a decrement in the Kaplan–Meier curve, it is also sensible to constrain the number of survival events at those times to be 0. This can be achieved by augmenting the existing set of time points, t_i, to include the times of the tick marks. The value of $\tilde{d}_i$ at any time point added will be constrained to 0. Let $\mathcal{T}^*$ be the set of tick marks, then for $\mathcal{C}^* = \{i : t_i \in \mathcal{T}^*, o_i = 0\}$ and $\mathcal{S}^* = \{i : t_i \notin \mathcal{T}^*)\}$ we add the additional constraints

$$\tilde{c}_j = 0 \text{ for } j \in \mathcal{S}^*, \text{ and } \tilde{d}_j = 0, \text{ for } j \in \mathcal{C}^*,$$

which is equivalent to removing those variables from the optimization. Note also that, since in the continuous solution any j not corresponding to a tick mark would have $\tilde{c}_j = 0$, this property would be retained after the scaling taken in (2.5). Depending on the confidence in the identification of the tick marks, one can make the further constraint that $\tilde{c}_j \geq 1$ for $j \in \mathcal{C}^*$, i.e. that all tick marks correspond to a distinct censoring time.

2.2 Avoiding indeterminable solutions

The resulting quadratic program may not necessarily have a positive definite objective matrix, $\mathbf{Q}$, implying there is not a unique solution, even to the continuous problem. A notable, though not unique, reason for this issue is if tick marks are available and there are two or more tick marks between event times. Since the Kaplan–Meier estimate only depends on the numbers at risk at the event times, any allocation of the same number of censored events in those tick marks will result in the same decrements of the Kaplan–Meier estimate and hence the same objective value for the quadratic program. One possible solution to this issue would be to collapse together such points within the optimization, including some constraint that the number of censored events in the group should be at least the number of tick marks, and then use some heuristic to allocate events within groups of points (e.g. allocating equally).

Alternatively, one can instead add an additional term $\epsilon(\sum_i c_i^2)$ to the objective in (2.3) for some suitably small value ϵ - which leads to a preference for solutions with fewer censoring ties in cases where there is no, or virtually no, difference between solutions. This is similar to the iKM algorithm's default assumption that censoring events are uniformly distributed across decrement points. Considerations of the size for ϵ are mainly to ensure it is large enough such that the resultant objective matrix is deemed

to be numerically invertible. If the approximate optimization method, based on integer rounding the continuous solutions is used, the resulting solution is not sensitive to the choice of ϵ and a value of $\epsilon = 0.001$ is a reasonable default choice. This latter approach of regularizing the optimization appears to perform better at reconstructing the pattern of censoring for both real and simulated datasets, and has the advantage of always ensuring the objective is positive-definite.

3 Competing risks data

Often a time-to-event outcome may be subject to competing risks, for instance cohort studies into time-to-diagnosis of cancer would have death before cancer diagnosis as a competing risk. Such data can be characterized by the time-to-event T and the corresponding event indicator $D \in \{1, \ldots, m\}$ representing which of m possible causes occurred. The primary quantity of interest is usually the cumulative incidence function for the cause of interest, defined as $F_j(t) = P(T \leq t, D = j)$. Note that, unlike in the standard survival case, in the presence of competing risks the cumulative incidence will not generally correspond to one minus the Kaplan–Meier curve (i.e. computed by treating other events as censoring), but is instead be a function of all the individual cause-specific hazards (see for instance, Putter et al (2007)[10]).

The Aalen-Johansen estimate of $F_j(t)$ is given by

$$F_j(t) = \sum_{i:t_i \leq t} \frac{d_{ij}}{r_i} \hat{S}(t_i-) \tag{3.1}$$

where

$$\hat{S}(t_i-) = \prod_{k:t_i<t} (1 - \frac{d_k}{r_k}) \tag{3.2}$$

is the Kaplan–Meier estimate of all-cause survival just before time t_i and $d_k = \sum_j d_{kj}$ is the total number of events at time t_k.

Suppose that for a given study the non-parametric cumulative incidence curves for each of m competing risks are available at a series of time points, such that there are observations f_{ij} at time t_i for $i = 1, \ldots, N_S$ and $j = 1, \ldots, m$. Note that the assumption that $\hat{F}_j(t)$ is constant between observed points allows f_{ij} to be defined for all j even if the individual digitization of $\hat{F}_j(t)$ did not include an observation at t_i.

From (3.1) and (3.2) we can note that

$$\frac{f_{ij} - f_{i-1j}}{1 - \sum_j f_{i-1j}} = \frac{d_{ij}}{r_i}$$

where $f_{0j} = 0$. Using a similar argument to Section 2.1, taking

$$o_{ij} = \frac{f_{ij} - f_{i-1j}}{1 - \sum_j f_{i-1j}} \tag{3.3}$$

provides a set of identities $o_{ij}r_i = d_{ij}$ for $i = 1, \ldots, N_S$ and $j = 1, \ldots, m$ and thus a quadratic form for the objective function can be obtained by

$$W = \sum_{i=1}^{N_S} \sum_{j=1}^{m} \left(o_{ij} \left\{ N - \sum_{k<i} \sum_{l=1}^{m} d_{kl} - \sum_{k<i} c_k \right\} - d_{ij} \right)^2 . \tag{3.4}$$

As in the Kaplan–Meier case, we would look to add an additional term $\epsilon(\sum_{i=1}^{N_S} c_i^2)$ to (3.4) to ensure the objective matrix is positive definite. Note that (3.4) reduces to (2.3) when there is a single risk (and hence $F_1(t) \equiv 1 - S(t)$).

The information on patients at risk defines equality constraints of the form

$$\sum_{i \in \mathcal{C}_j} \sum_{k=1}^{m} \{d_{ik} + c_i\} = R_j - R_{j-1}, \ j = 1, \ldots, N_r$$

where $\mathcal{C}_j = \{i : \tau_{j-1} \leq t_i < \tau_j\}$. If available, information on the total number of events of each type, N_{Ej}, could be incorporated through m additional equality constraints

$$\sum_{i=1}^{N_S} d_{ij} = N_{Ej}, \text{ for } j = 1, \ldots, m.$$

Alternatively, if only the total number of events of all types is known, an equality constraint, $\sum_{i=1}^{N_S} \sum_{j}^{m} d_{ij} = N_E$, can be given. Since $\hat{F}_j(t)$ can only increase at times corresponding to an observed event of type j, we can also impose the inequality constraints $d_{ij} \geq 1$ for $i \in \{k : o_{kj} > 0\}$.

As in the case of Kaplan–Meier survival data, either an 'exact' MIQP can be solved or else the method may proceed by first solving the continuous QP and then applying a heuristic to approximate the integer solution. Specifically, for continuous solution $\tilde{d}_{ij}$ the integer solution is found by taking

$$\hat{d}_{ij} = \left\lfloor \frac{1}{2} + \sum_{k=1}^{i} \tilde{d}_{kj} \right\rfloor - \left\lfloor \frac{1}{2} + \sum_{k=1}^{i-1} \tilde{d}_{kj} \right\rfloor .$$

The estimated censoring events are then found in the same way as in Section 2.1, by using the midpoint rounding in (2.5) with scaling

$$V_j = \frac{(R_j - R_{j-1})}{\sum_{i \in \mathcal{C}_j} (\sum_{k=1}^{m} \hat{d}_{ik} + \tilde{c}_i)}, \ j = 1, \ldots, N_r.$$

Note that if only one risk is of interest, provided the all-cause Kaplan–Meier curve and the cumulative incidence function for the risk of interest are reported, it is possible to recover the sufficient statistics to either perform a Cox proportional cause-specific hazard[19] or Fine-Gray competing risks analysis[20] with respect to the cause of interest.

4 Illustrative example: Advanced follicular lymphoma study

As an illustrative example, the results from a retrospective study on long-term outcomes of patients with advanced stage follicular lymphoma (FL) are considered.[21] The authors followed a cohort of 286 stage

III-IVA FL patients seen between 2000 and 2011 who followed a 'watch and wait' management strategy. A particular endpoint of interest was progression-free survival (PFS) which was defined as the time from diagnosis to lymphoma treatment or death from any cause.

4.1 Progression-free survival data

In the original paper, the Kaplan–Meier estimate of progression free-survival among the 'watch and wait' management group is presented, including information on the number of patients at risk at 20-month intervals and tick marks to indicate the times at which patients were censored. Digitization of the curves was performed by using WebPlotDigitizer,[22] identifying 113 distinct decrement points and 80 tick marks via the manual extraction method.

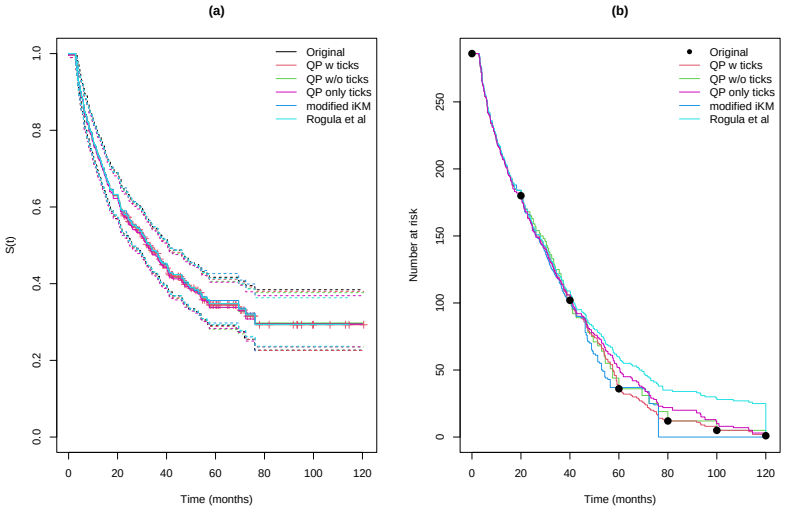

Figure 1: Progression-free survival estimates (panel (a)) and estimated number at risk (panel b) from reconstructed pseudo-IPD using different methods.

The proposed QP method is compared with the corresponding reconstructions using the modified iKM algorithm and the algorithm of Rogula *et al.* Figure 1 shows the resulting Kaplan–Meier estimates and associated pointwise 95% confidence intervals (in each case these are computed assuming the sampling distribution of $\log \hat{S}(t)$ is approximately normally distribution). All methods give reasonably good agreement with respect to the implied Kaplan–Meier curve. Moreover, all methods except those not using the persons at risk information give good agreement with the confidence intervals, with deviations occurring later in follow-up. There is more disagreement regarding the estimated numbers at risk. Notably, the modified iKM algorithm censors individuals at the mid-point between observed event times, or at the final event time. However, since the last uncensored event occurs at around 76 months, all remaining patients are incorrectly censored at this time, which contradicts the persons at risk informa-

tion provided. The method of Rogula *et al* uses the information from the marked censoring times, but not the information of numbers at risk. As a consequence, the estimated number at risk at later time points is greatly over-estimated. The corresponding QP estimate using only the marked censoring times also over-estimates the number at risk at later times, but to a much lesser degree. The total number of events was not reported and there is some disagreement across methods with 177, 178, 181, 172 and 182 estimated for QP with ticks, QP w/o ticks, QP ticks only, modified iKM and Rogula *et al*, respectively.

Often the purpose of obtaining pseudo-IPD is to facilitate survival extrapolation for cost-effectiveness analysis in order to assist the calculation of life expectancy or healthy life expectancy, which usually involves fitting appropriate parametric models. While often an unavoidable part of cost-effectiveness analysis, in practice survival extrapolation requires careful consideration. For instance, any extrapolation model should be clinically plausible, the sensitivity of conclusions to modelling assumptions should be assessed, and external information should be used in the extrapolation, where possible.[23] Here we consider a simplified process where a Royston-Parmar flexible spline model[24] with one internal knot is fitted to the reconstructed pseudo-IPD based on each method. This models the log cumulative hazard as a natural cubic spline with respect to log-time such that

$$\log\{-\log S(t)\} = \gamma_0 + \gamma_1 \log t + \gamma_2 \left\{ (\log t - k_1)_+^3 - \frac{k_2 - k_1}{k_2 - k_0}(\log t - k_0)_+^3 - \frac{k_1 - k_0}{k_2 - k_0}(\log t - k_2)_+^3 \right\}, \quad (4.1)$$

where k_0 and k_2 are the lower and upper boundary knot points and k_1 is the internal knot point, all measured with respect to log-time and $(x)_+ = \max(0, x)$. When $\gamma_2 = 0$, the parameters γ_0 and γ_1 correspond to the log-rate and shape in a Weibull model, with $\gamma_2 \neq$ determining the degree of deviation away from the Weibull trend.

For the QP method without the information on marked censoring times, we note that, although the QP procedure assigns censoring times at either decrement times or the times at which numbers of patients at risk were reported, the published data would not change if the censoring in fact occurred anywhere in the interval $(t_i, t_{i+1}]$. As a consequence, when fitting parametric models the censoring times are taken at the midpoint of this interval, i.e. $0.5(t_i + t_{i+1})$. This approach matches how the modified iKM algorithm assigns censoring times. All models are fitted using the *flexsurv* package in **R**.[25] For comparability of the parameters, the same set of knot points (including boundary knots) are used across all fits. Specifically, we use the default method of choosing knot points in *flexsurv* to choose the knots based on the QP with ticks reconstruction which gives knots at the logged values of $t = (0.05, 14.67, 76.16)$ corresponding to the minimum, median and maximum observed event time in the pseudo-IPD based on the QP reconstruction using all the available information.

Table 1 gives the parameter estimates and their corresponding standard errors for each of the methods. There is close agreement between the estimated parameters for the QP methods. Figure 2 gives the

Table 1: Comparison of parameter estimates for a Royston-Parmar flexible spline model with one internal knot point. The parameters are defined in (4.1)

	Point estimates			
Parameter	QP w ticks	QP w/o ticks	modified iKM	Rogula
R γ_0	-4.8283	-4.8164	-5.3570	-5.4805
γ_1	2.4841	2.4734	2.9415	3.1617
γ_2	0.0693	0.0689	0.0869	0.0976
	Standard errors			
Parameter	QP w ticks	QP w/o ticks	modified iKM	Rogula
γ_0	0.4032	0.4020	0.4540	0.4569
γ_1	0.3607	0.3589	0.4045	0.3906
γ_2	0.0147	0.0146	0.0164	0.0153

corresponding estimates of survival and the estimated hazard functions. The QP estimates with or without tick marks are virtually indistinguishable. The modified iKM's pseudo-IPD gives a slightly higher survival estimate (and lower hazard estimate) in the tail which is similar to that given by the QP estimate using just the marked censoring times, while the estimate based on Rogula *et al*'s method is noticeably higher.

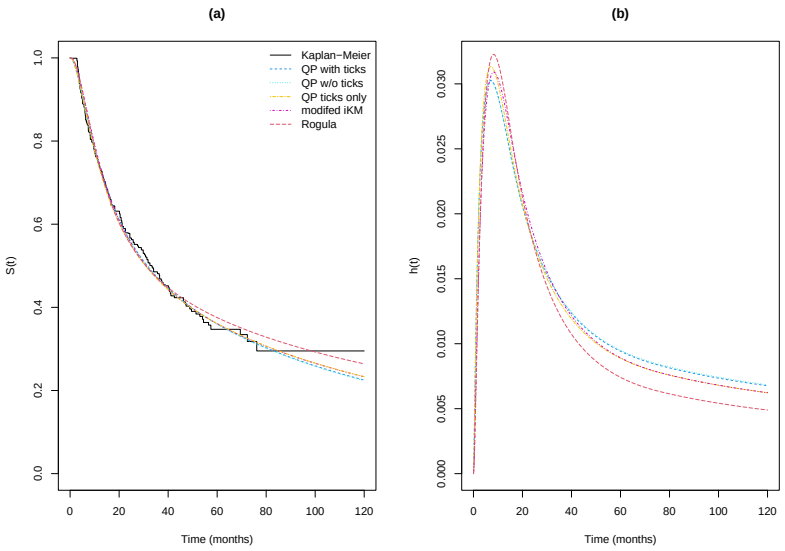

Figure 2: Comparison of Royston-Parmar flexible spline model fitted survivor functions (panel (a)) and hazard functions (panel (b)) based on pseudo-IPD generated via different methods.

For a health economic evaluation it may be necessary to determine the restricted mean survival for these patients up to say 25 years (300 months). Based on extrapolating the Royston-Parmar model fits, while the model based on the pseudo-IPD from the QP with all data gives an estimate of 6.37 years of progression-free survival, the modified iKM gives 6.56 years and Rogula et al's method gives 7.25 years.

In Section S2 of the Supplementary Material, the reported progression-free survival for the two arms of the KEYNOTE-177 trial is used to provide a comparison of the reconstruction methods with respect to their ability to replicate the reported hazard ratio.

4.2 Competing risks analysis

Among patients treated with 'watch and wait' management, a competing risks analysis was performed with respect to the two component outcomes that defined progression-free survival; 'Treatment or follicular lymphoma related death' and 'death from unrelated cause before progression'. The cumulative incidence functions are presented including the number of patients at risk at 20 month intervals and tick marks indicating the timing of censoring. Pointwise 95% confidence intervals for the cumulative incidence of treatment or follicular lymphoma related death are also given.

As before, digitization of the curves was performed by using WebPlotDigitizer,[22] identifying 84 and 12 distinct increment points for the two curves, along with 78 tick marks.

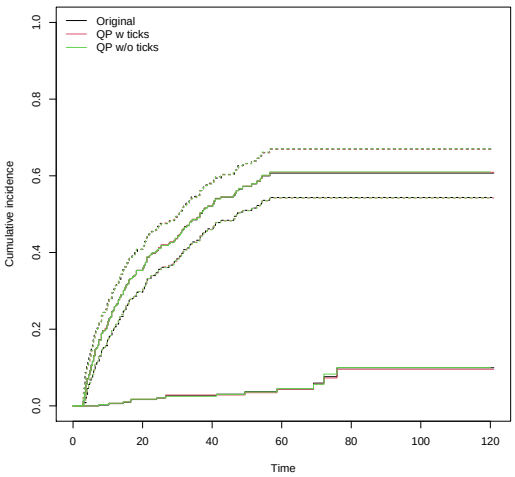

Figure 3: Reconstructed cumulative incidence estimates from the FL study. The upper curve corresponds to treatment or FL-related death and shows the 95% confidence interval. The lower curve corresponds to death from an unrelated cause before progression.

Figure 3 presents the original digitized and the reconstructed estimates of the cumulative incidence functions. The reconstructions both with and without the information on censoring times give good

agreement with the digitized cumulative incidence curve for 'Treatment or follicular lymphoma related death'. Both reconstructions have some level of disagreement for the cumulative incidence of death from other causes before progression. This is likely to be due to quality of the digitization - the original curve is dashed making identification of step points more difficult.

5 Simulations

In this section, three sets of simulations are presented, where the first two sets aim to compare the proposed methods to existing methods with respect to reconstructing IPD for Kaplan–Meier curves, and the last set assesses the proposed methods ability to reconstruct IPD from published cumulative incidence curves.

5.1 Comparison with the modified iKM and Rogula *et al* methods

The first set of simulations aim to compare the proposed approach for reconstructing IPD for Kaplan–Meier curves with existing methods. Specifically, the focus is on assessing each method's ability to replicate the results of analyses on the underlying true IPD. A secondary aim is to assess the impact of additional information, such as the marked censoring times, total number of events and reporting of number of patients at risk, on the quality of the reconstruction. A further aim is to assess whether the exact MIQP approach yields more accurate reconstructions than the heuristic QP method based on rounding.

Data generating mechanisms

The underlying survival times are generated from a Weibull distribution with hazard function $h(t) = \lambda^\alpha \alpha t^{\alpha-1}$ with a shape parameter $\alpha = 0.8$ and rate parameter $\lambda = 0.2$. A shape parameter of $\alpha = 0.8$ is chosen to induce a higher rate of events close to time 0 leading to a higher rate of tied events in the coarsened data. The censoring distribution is taken to be independent Unif$(2, 8)$, resulting in a censoring rate of around 39% (i.e. around 61 uncensored deaths per dataset). A total sample size of 125 patients is assumed. To emulate the level of accuracy that might be observed in practice (for instance reporting to the nearest day or week), the continuous-time event times are coarsened by rounding up to the next multiple of 0.05. The true IPD is assumed to be reported at this level of resolution. In addition, to emulate the level of error that might occur through digitization, the values of $\hat{S}(t)$ are rounded to three decimal places when supplied to the algorithms.

Different scenarios are generated by varying the level of auxiliary information available in addition to the Kaplan–Meier curve itself for a given simulated dataset. In the full information case, it is assumed

that the number at risk is reported at $t = 0, 1, 2, \ldots, 8$, that the total number of events is given and the set of marked censoring times are available. To provide a direct comparison with the modified iKM algorithm, a scenario where the marked censoring times are omitted is considered. Similarly, to compare with the Rogula *et al* method, a scenario is included where only the total sample size and the marked censoring times are used. An intermediate scenario where the marked censoring times, total sample size and total events are known, but not intermediate numbers at risk is also considered for the proposed QP method.

Estimands and performance measures

The comparator methods are assessed with respect to their ability to replicate the actual Kaplan–Meier estimate, $\hat{S}(t)$, and also their ability to replicate the actual cumulative persons at risk, $Y(t)$, across the total follow-up period. In addition, the maximum likelihood estimates of the parametric Weibull survival model are assessed for each method.

The performance with respect to the replication of $\hat{S}(t)$ and $Y(t)$ is assessed with respect to the integrated absolute discrepancy compared to the respective quantities in the true IPD, leading to two measures:

$$\Delta_S = \int_0^\tau |\hat{\hat{S}}(t) - \hat{S}(t)|dt$$

and

$$\Delta_Y = \int_0^\tau |\hat{Y}(t) - Y(t)|dt,$$

where $\hat{\hat{S}}(t)$ and $\hat{Y}(t)$ are the implied Kaplan–Meier estimate and implied number at risk based on the reconstructed pseudo-IPD. In each case, τ is taken to be the maximum follow-up time in the true data. A perfect reconstruction would give $\Delta_S = 0$ and $\Delta_Y = 0$.

The performance of the estimates of the Weibull model parameters are assessed with respect to their bias and root-mean squared error, where these are defined in terms of the discrepancy with the corresponding estimate from the true IPD (not the discrepancy from the true population parameter).

Methods

For the full information scenario, and the scenario with all information except the marked censoring times, the exact MIQP method is applied, using IBM ILOG CPLEX, but with a maximum computation time limited to 60 seconds for the optimized, and absolute and relative convergence tolerances of 1×10^{-8} and 1×10^{-7}, respectively. In cases where the objective function for the solution of the MIQP optimization at termination was worse than for the approximate method, the approximate solution was used instead.

Table 2: Accuracy of different method of reconstructing pseudo-IPD from simulated Kaplan–Meier curve data in the base case ($N = 125$, $S(t)$ points rounded to 3 decimal places. 20 possible values of t per time unit. nd = number of deaths given.

	With ticks		Without ticks			Without at risk data		
Measure	MIQP	QP	MIQP	QP	mod iKM	QP ticks + nd	QP ticks	Rogula
$\Delta(S)$	0.0018	0.0026	0.0049	0.0092	0.0422	0.0036	0.0204	0.0192
$\Delta(Y)$	1.0568	1.1045	9.3705	10.2064	23.4115	1.9093	17.7187	24.6555
RMSE($\log(\lambda)$)	0.0022	0.0019	0.0052	0.0070	0.0419	0.0035	0.0178	0.0286
RMSE($\log(\alpha)$)	0.0014	0.0011	0.0033	0.0043	0.0395	0.0021	0.0151	0.0218

The heuristic QP method is applied to all scenarios. The underlying continuous quadratic program is solved using the quadprog package in R. The modified iKM algorithm, as implemented in the R package IPDfromKM,[8] is applied in the scenario where numbers of patients at risk is assumed known. The method of Rogula et al, as implemented in the package KMtoIPD[26] is applied in the case where only the total sample size and the marked censoring times are used.

Results

Table 2 presents the results from 1000 simulated datasets. As might be expected, the accuracy of the reconstruction increases with the additional information used across the QP methods. There is also a very slight improvement in accuracy for the MIQP method compared to the integer rounding heuristic solution. However, it is probably not sufficient to justify the additional computational complexity. The MIQP and QP methods based on numbers at risk do discernibly better than the corresponding modified iKM reconstructions that use the same information. It is also worth noting that only 4.5% of simulated datasets resulted in a modified iKM reconstruction that matched the total number at risk and total number of events supplied.

The QP reconstruction using only the information from tick marks gives a comparable match to the Kaplan–Meier curves to the Rogula *et al* approach, but somewhat better estimates of numbers at risk and the Weibull model parameters. It also also worth noting that just by adding the additional information on the total number of events/deaths to the QP method using only the marked censoring times, brings the quality of the reconstruction almost into line with those also using the intermediate estimates of time at risk.

Additional simulation results are shown in the Supplementary Materials where either the resolution of the reporting times or the number of decimal places the $\hat{S}(t)$ values are rounded, was varied. As might

be expected, the quality of all reconstructions improved as there were fewer tied events in the data and where there was less rounding error. The same pattern of relative performance between methods is seen in all scenarios. In addition, Table S9 of the Supplementary Materials presents a comparison of the QP method and modified iKM when only the total sample size and total number of events is given. In that case, the quality of the reconstruction is generally worse than using the location of marked times only, but is less sensitive to the presence of ties.

5.2 Estimation of treatment group differences

A second set of simulations aim to assess the ability of different methods to reconstruct a pair of Kaplan–Meier curves, corresponding to two arms of a clinical trial, in order to assess the ability to estimate a hazard ratio.

Data generating mechanisms

Patients in the control group are simulated from the same Weibull distribution as in the single sample case above. The sample size in each arm is taken as $N = 125$. For the treatment group we assume a Weibull distribution with the same shape parameter (ensuring proportionality) but with a lower hazard (corresponding to a log hazard ratio of 0.5). As before, the exact event times are coarsened by rounding up to the next 0.05 and the $S(t)$ values are the true values rounded to three decimal places. The same scenarios in terms of availability of auxiliary information are used as above.

Estimands and performance measures

The methods are assessed in terms of the estimated log hazard ratios (HR) obtained by fitting a Cox proportional hazard model and by fitting a fully-parametric Weibull proportional hazards model to the pooled pseudo-IPD obtained through each of the reconstruction methods.

To emulate the estimation of hazard ratios from two Kaplan–Meier curves, data was simulated from two arms of a clinical trial. Both a Cox proportional hazards model and a Weibull proportional hazards model was fitted to the pooled pseudo-IPD obtained through each of the reconstruction methods. The test statistic of the Grambsch-Therneau test of proportionality based on the scaled Schoenfeld residuals[27] was also computed in each case. Finally, the difference in restricted mean survival times[28] was computed, taking $\tau = 5$ as the upper time point.

For each of the estimands the performance is assessed by considering the bias and root mean squared error (RMSE), where in each case error is assessed in relation to the corresponding estimate using the true IPD.

The same set of comparator methods are used as in the first set of simulations except that only the heuristic QP method is considered, rather than the MIQP method.

Results

Table 3 gives the resulting bias and RMSE from 1000 simulated datasets. The QP method using the information from tick points performs exceptionally well, with bias and mean squared errors of 0.002 or less for all quantities except the Grambsch-Therneau test statistic. As in the one-sample case, the QP method without tick points gives lower root MSE for all quantities compared to the modified iKM method using the same information. The most notable discrepancies occur in the Grambsch-Therneau test statistics. Since the data were simulated from data with proportional hazards, the test statistic has an approximate χ^2_1 distribution using the true IPD. There is close agreement when the QP method is used with tick marks (bias 0.002, RMSE 0.026), substantially more variability for the QP method without tick marks (bias -0.002, RMSE 0.158). There is some tendency for both the QP method using only information from the tick marks and the corresponding Rogula et al method to give an inflated Grambsch-Therneau test statistic, and both the modified iKM and Rogula et al methods have relatively high RMSEs in relation to the statistic for the true IPD, indicating a higher risk of discordance between the conclusions regarding assessments of proportional hazard. As in the single sample case, a QP reconstruction using just the marked censoring times and the total number of events (deaths) gives markedly better accuracy than without the total number of events and is also better than the QP reconstruction using total number of events and intermittently observed numbers at risk.

5.3 Performance of competing risks reconstructions

The final set of simulations aim to confirm the accuracy of the proposed method for reconstructing IPD from cumulative incidence estimates from competing risks data. As above, the focus is on estimating treatment effects, but here the treatment effect could be with respect to either of two competing events and can be either based on the cause-specific hazard or based on the sub-distribution hazard using a Fine-Gray model.

Data generating mechanisms

Competing risks data are simulated from a process with two competing risks and two treatment groups.

Let $Z = 0, 1$ represent the treatment group, then the competing risks data are generated by assuming the CIF of the first risk, satisfies

$$F_1(t; Z = 0) = \kappa[1 - \exp\{-(\lambda_{01}t)^{\alpha_1}\}]$$

Table 3: Bias and RMSE in parameter estimates for two group comparisons based on pseudo-IPD reconstructed via different methods. nd = total number of deaths given.

	Bias					
	With ticks	Without ticks		Without at risk data		
	QP	QP	mod iKM	QP ticks + nd	QP ticks	Rogula
Cox log(HR)	0.000	-0.002	0.004	0.000	0.001	0.000
Weibull log(HR)	0.000	-0.001	0.001	0.001	-0.002	-0.001
Grambsch-Therneau X^2	0.002	-0.002	0.058	0.005	0.107	0.164
ΔRMST	0.000	0.001	0.002	0.000	0.001	0.000
	RMSE					
	With ticks	Without ticks		Without at risk data		
	QP	QP	mod iKM	QP ticks + nd	QP ticks	Rogula
Cox log(HR)	0.001	0.006	0.033	0.002	0.014	0.019
Weibull log(HR)	0.001	0.004	0.023	0.002	0.015	0.020
Grambsch-Therneau X^2	0.026	0.158	0.821	0.045	0.371	0.564
ΔRMST	0.002	0.004	0.008	0.002	0.003	0.004

and that F_1 adheres to the Fine-Gray assumption of proportional sub-distribution hazards with a log-hazard ratio of β_1 such that

$$F_1(t; Z = 1) = 1 - (1 - F_1(t; Z = 0))^{\exp \beta_1} = 1 - [1 - \kappa\{1 - \exp(-(\lambda_{01}t)^{\alpha_1})\}]^{\exp \beta_1}.$$

To allow $F_2(t; Z)$ to be a valid CIF, we note that the model implies $\lim_{t\to 0} F_1(t; Z) = 1-(1-\kappa)^{\exp(\beta_1 Z)}$, so then take

$$F_2(t; Z) = (1 - \kappa)^{\exp(\beta_1 Z)}[1 - \exp(-\{\lambda_{02} \exp(\beta_2 Z)t)^{\alpha_2}\}].$$

Hence the Fine-Gray model for cause 1 is correctly specified, while for the Fine-Gray model for cause 2 and for either of the Cox models with respect to the cause-specific hazards, the models are miss-specified.

Specifically, we take $(\kappa, \lambda_{01}, \alpha_1, \lambda_{02}, \alpha_2, \beta_1, \beta_2) = (0.6, 0.4, 1.2, 0.2, 1.5, -0.3, 0.3)$. The censoring distribution is assumed to be Unif$(1, 6)$. The resolution of reporting of times is again taken as the nearest 0.05 and each treatment group has $N = 125$ patients. For each simulated dataset, the Aalen-Johansen estimates of the cumulative incidence functions are computed for both of the competing events, where the CIF points are rounded to three decimal places to emulate digitization error.

Three scenarios are generated by varying the amount of auxiliary information supplied to the algorithm: use of marked censoring times, numbers at risk at times $1, 2, \ldots, 5$ and total number of events of each type; use of only numbers at risk at times $1, 2, \ldots, 5$ and total number of events of each type; use of only the marked censoring times.

Estimands and performance measures

For each of the scenarios, the Fine-Gray sub-distribution hazard ratio, and the cause-specific hazard ratio is estimated for both of the competing events. In each case, performance is assessed based on bias and RMSE defined with respect to the corresponding estimates obtained using the true simulated IPD.

Results

Table 4 displays the results of 1000 simulated datasets, where for each scenario the heuristic QP method is used. In all cases, the reconstructed data produces estimates close to those found using the original data. As one would expect, there is a larger discrepancy between the estimates from the true IPD when less information is used in the reconstruction.

6 Discussion

The results in the paper have confirmed previous work[6, 9] that has shown that accurate pseudo-IPD can be reconstructed from published Kaplan–Meier curves, provided either inteim information of numbers of

Table 4: Bias and RMSE in parameter estimates for competing risks analyses.

	Bias		
	QP with ticks	QP w/o ticks	QP only ticks
Fine-Gray HR_1	-0.001	0.001	-0.001
Fine-Gray HR_2	-0.001	0.000	-0.010
CSH HR_1	-0.001	0.000	0.006
CSH HR_2	-0.001	-0.001	-0.008
	RMSE		
	QP with ticks	QP w/o ticks	QP only ticks
Fine-Gray HR_1	0.001	0.003	0.026
Fine-Gray HR_2	0.001	0.003	0.037
CSH HR_1	0.002	0.008	0.025
CSH HR_2	0.002	0.011	0.031

patients at risk or the marked censoring times are provided. It also shows that some gain in accuracy can be achieved by adopting a quadratic programming approach which treats the numbers of patients at risk as strict equality constraints. Moreover, the quadratic programming approach easily accommodates both the information on individual censoring times and the numbers at risk, which was shown to give a more appreciable improvement in accuracy, particularly if the pseudo-IPD is to be used for a parametric analysis.

A limitation of the simulation study in Section 5.1 is that the underlying model for data generation is restricted to a particular Weibull distribution. While, the non-parametric nature of the Kaplan–Meier method should mean that the patterns seen in the performance measures Δ_S and Δ_R would likely translate to other underlying distributions. However, it is possible other models might behave differently in terms of the discrepancy between the IPD and reconstructed estimates of the model parameters.

The proposed quadratic programming approach also allows for reconstruction of competing risks survival data from published cumulative incidence curves. This could be particularly useful in performing meta-analyses in cases where, for instance, some studies report the Fine-Gray sub-distribution hazard ratio and other studies report the hazard ratio associated with the cause-specific hazard, or report neither. Existing methods for meta-analysis for competing risks events rely on strong assumptions such as constant hazards,[29] or only use information on the event counts and total follow-up in the studies.[30]

An **R** package, *CIFresolve*, has been prepared to implement the QP-based methods proposed in the

paper. For the integer rounding approach, the functions use the *quadprog* package to solve the continuous quadratic program. The MIQP approach can also be used via the *Rcplex* package which allows IBM CPLEX to be called from **R**, provided it is installed. The package is available via the author's Github page; `https://github.com/andrewtitman/CIFresolve`.

While the optimization problem in the proposed method is strictly a MIQP, simulation results indicate the additional computational burden of using methods for MIQP compared to simple integer rounding of the continuous solution is usually not justified. Hence using the quicker and widely available approach based on standard QP is recommended in most cases. If the total patient time at risk, $t_{\mathcal{P}}$ has been given (or can be inferred, for instance via the MLE of an exponential model) then this implies an equality constraint; $t_{\mathcal{P}} = \sum_i t_i(d_i + c_i)$. In practice, since the exact censoring times may not be known and the decrement points may have been digitized, the information would be better incorporated by stipulating;

$$(1 - \delta_{\mathcal{P}})t_{\mathcal{P}} \leq \sum_i t_i(d_i + c_i) \leq (1 + \delta_{\mathcal{P}})t_{\mathcal{P}}$$

for a suitably chosen $\delta_{\mathcal{P}}$, for instance, $\delta_{\mathcal{P}} = 0.001$. However, it is not clear how a heuristic rounding rule (such as in (2.5)) could be stipulated to preserve these inequalities when going from a continuous to integer solution of the QP. Hence, the MIQP approach to optimization is needed to incorporate these constraints.

The methods can also be adapted to provide pseudo-IPD from published Nelson-Aalen estimates of the cumulative hazard, or cumulative cause-specific hazard. In that case, the *increment* of the Nelson-Aalen estimate at time t_i can be used as o_i or o_{ij} in (2.3) and (3.4), with the methods otherwise proceeding the same. Similarly, the methods can be adapted to account for other sources of additional information. For instance, some Kaplan–Meier or cumulative incidence function plots include intermittently-reported cumulative event counts, in addition to numbers at risk.

A further extension is to allow reconstruction of non-parametric survival curves estimated on left-truncated and right-censored data using the Lynden-Bell generalization of the Kaplan–Meier estimator,[31] provided information on the, not necessarily decreasing, numbers at risk is available intermittently. In that case, it is not possible to infer individual-level data, but the sufficient statistics for parametric models under the standard assumption that the left-truncation times are quasi-independent of survival (aggregated numbers at risk and numbers of events over time) should be recoverable.

Recently in oncology health technology assessments, there is interest in fitting multi-state illness-death models using information from published progression-free survival and overall survival curves.[32,33] An important step in this process is ensuring good reconstructions of the pseudo-IPD from the PFS and OS curves. Potentially, the proposed QP approach could be used to ensure consistency between the sources of information. In particular, assuming the PFS and OS curves come from the same set

of patients and that each patient's right-censoring time is the same for PFS and OS, the number of patients at risk in the PFS data must necessarily be less than equal to that for OS at all follow-up times. Similarly, the number of patients censored for OS must be greater than or equal to the number of PFS censorings at all times. These constraints can be incorporated into relevant linear constraints in the QP provided care is taken to align the decrement points and marked censoring times on the digitized PFS and OS curves.

In some cases, the published Kaplan–Meier curve may provide no person-at risk information or marked censoring times, but will include pointwise 95% confidence intervals. If the method by which the confidence intervals have been constructed is known or can be inferred, it may be possible to gain additional information on the number of patients at risk, which could be incorporated into the objective function. For instance, if Greenwood's formula has been used for the standard error and naive (symmetric) asymptotic 95% confidence intervals have been constructed, then the limits will be of the form

$$\hat{S}(t)\left\{1 \pm 1.96\sqrt{\sum_{i:t_i \le t} \frac{d_i}{r_i(r_i - d_i)}}\right\}.$$

In theory, digitization could be used to infer $\frac{d_i}{r_i(r_i-d_i)}$ from the change in confidence intervals at decrements of $\hat{S}(t)$. However, the image quality is unlikely to be sufficient to capture this accurately since it is of order n^{-2}, at a given decrement of $\hat{S}(t)$. Nevertheless, it may be possible to incorporate information on the implied values of $\sum_{i:t_i \le t} \frac{d_i}{r_i(r_i-d_i)}$ at suitably spaced time points. Optimization in this case is more challenging since either the objective function or one or more of the constraints would need to be non-linear.

Appendix A: Form of the quadratic program

Let $\mathbf{H}$ be an $N_s \times 2N_s$ matrix where

$$\{\mathbf{H}\}_{ij} = \begin{cases} o_i & \text{if } j < i \text{ or } N_s < j < i + N_s \\ 1 & \text{if } j = i \\ 0 & \text{otherwise.} \end{cases}$$

Then, minimizing D in (2.3) with respect to $\mathbf{x} = (d_1, d_2, \ldots, d_{N_s}, c_1, c_2, \ldots, c_{N_s})'$ is equivalent to minimizing $\frac{1}{2}\mathbf{x}'\mathbf{Q}\mathbf{x} + \mathbf{q}'\mathbf{x}$ where $\mathbf{Q} = \mathbf{H}'\mathbf{H}$ and $q = -\mathbf{o_N}'\mathbf{H}$ where $\mathbf{o_N}$ is a vector of length N_s with ith entry $N \times o_i$.

The constraint matrix $\mathbf{A}$ and vector $\mathbf{b}$ depend on the specific auxiliary information provided, but $\mathbf{A}$

is an $N_c \times 2N_s$ matrix and the jth row of $\mathbf{A}$ defines either an equality or inequality constraint bounded by the corresponding entry of $\mathbf{b}$, which is a vector of length N_c.

A similar formulation can be constructed for D given in (3.4).

References

1. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct and reporting. *BMJ.* 2010;340:c221.

2. Parmar MKB, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literautre for survival endpoints. *Stat Med.* 1998; 17(24):2815-2834.

3. Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials.* 2007; 8:1-16.

4. Liu Z, Rich B, Hanley JA. Recovering the raw data behind a non-parametric survival curve. *Syst Rev.* 2014; 3:151.

5. Hoyle MW, Henley W. Improved curve fits to summary survival data: application to economic evaluation of health technologies. *BMC Med Res Methodol.* 2011; 11:139.

6. Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan–Meier survival curves. *BMC Med Res Methodol.* 2012; 12:1.

7. Liu N, Zhou Y, Lee JJ. IPDfromKM: reconstruct individual patient data from published Kaplan–Meier survival curves. *BMC Med Res Methodol.* 2021; 21:111.

8. Liu N, Lee JJ. *IPDfromKM: Map Digitized Survival Curves Back to Individual Patient Data.* R package version 0.1.10. 2020

9. Rogula B, Lozano-Oretga G, Johnston KM. A method for reconstructing individual patient data from Kaplan–Meier survival curves that incorporate marked censoring times. *MDM Policy Pract.* 2022;7(1).

10. Putter H, Fiocco M, Geskus RB. Tutorial in Biostatistics: Competing risks and multi-state models. *Stat Med.* 2007; 26(11):2389-2430.

11. Goldfarb D, Idnani A. A numerically stable dual method for solving strictly convex quadratic programs. *Math Program.* 1983; 27(1):1-33.

12. Berwin A, Turlach R, Weingessel A. *quadprog: Functions to solve quadratic programming problems.* R package version 1.5-7. 2019.

13. Duran MA, Grossmann IE. An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Math Program.* 1986; 36:307-339.

14. Gupta OK, Ravindran A. Branch and bound experiments in convex nonlinear integer programming. *Manage Sci.* 1985; 31(12):1533-1546.

15. IBM ILOG CPLEX. 2009. *Version 12.1: Users manual for CPLEX.* International Business Machines Corporation.

16. Corrada Bravo, H. and Theussl, S. *Rcplex: R interface to CPLEX.* R package version 0.3-3. 2016

17. Roettger M, Gelius-Dietrich G, Fritzemeier CJ. *cplexAPI: R interface to C API of IBM ILOG CPLEX.* R package version 1.3.6. 2019.

18. Kannan R, Monma CL. "On the computational complexity of integer programming problems." In *Optimization and Operations Research: Proceedings of a Workshop Held at the University of Bonn, October 2–8, 1977*, Springer Berlin Heidelberg; 1978; 161-172.

19. Prentice RL, Kalbfleisch JD, Peterson Jr AV., Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics.* 1978; 34:541-554.

20. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc.* 1999; 94:496-509.

21. El-Galaly, TC, Bilgrau AE, de Nully Brown P, Mylam, et al. A population-based study of prognosis in advanced stage follicular lymphoma managed by watch and wait. *Br J Haematol.* 2015; 169(3):435-444.

22. Rohatgi A. *WebPlotDigitizer, version 5.* 2022; available at `http://automeris.io/wpd/` [Accessed 18 July 2024]

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

24. Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med.* 2002; 21(15):2175–2197.

25. Jackson CH. flexsurv: A Platform for Parametric Survival Modeling in R. *J Stat Softw.* 2016; 70(8):1-33.

26. Rogula B. *KMtoIPD: Reconstruct Individual Patient Data from Kaplan–Meier Survival Curves.* R package version 1.0.0. 2025

27. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika.* 1994; 81(3):515-526.

28. Royston P, Parmar MKB. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol.* 2013; 13:152.

29. Bonofiglio F, Beyersmann J, Schumacher M, Koller M, Schwarzer G. Meta-analysis for aggregated survival data with competing risks: a parametric approach using cumulative incidence functions. *Res Synth Methods.* 2016; 7(3):282-93.

30. Ades AE, Mavranezouli I, Dias S, Welton NJ, Whittington C, Kendall T. Network meta-analysis with competing risk outcomes. *Value Health.* 2010; 13(8):976-983.

31. Lynden-Bell D. A method of allowing for known observational selection in small samples applied to 3CR quasars. *Mon Not R Astron Soc.* 1971; 155(1):95-118.

32. Jansen JP, Incerti D, Trikalinos TA. Multi-state network meta-analysis of progression and survival data. *Stat Med.* 2023; 42(19):3371-3391.

33. Pahuta MA, Werier J, Wai EK, Patchell RA, Coyle D. A technique for approximating transition rates from published survival analyses. *Cost Eff Resour Alloc.* 2019; 17:12.