
12-1-2022

A Multistate Competing Risks Framework for Preconception Prediction of Pregnancy Outcomes

Kaitlyn Cook

Harvard Pilgrim Health Care Institute, kcook93@smith.edu

Neil J. Perkins

National Institute of Child Health and Human Development (NICHD)

Enrique Schisterman

University of Pennsylvania Perelman School of Medicine

Sebastien Haneuse

Harvard T.H. Chan School of Public Health

Follow this and additional works at: https://scholarworks.smith.edu/sds_facpubs



Part of the [Data Science Commons](#), and the [Statistics and Probability Commons](#)

Recommended Citation

Cook, Kaitlyn; Perkins, Neil J.; Schisterman, Enrique; and Haneuse, Sebastien, "A Multistate Competing Risks Framework for Preconception Prediction of Pregnancy Outcomes" (2022). Statistical and Data Sciences: Faculty Publications, Smith College, Northampton, MA.

https://scholarworks.smith.edu/sds_facpubs/63

This Article has been accepted for inclusion in Statistical and Data Sciences: Faculty Publications by an authorized administrator of Smith ScholarWorks. For more information, please contact scholarworks@smith.edu

RESEARCH

Open Access



A multistate competing risks framework for preconception prediction of pregnancy outcomes

Kaitlyn Cook^{1*}, Neil J. Perkins², Enrique Schisterman³ and Sebastien Haneuse⁴

Abstract

Background: Preconception pregnancy risk profiles—characterizing the likelihood that a pregnancy attempt results in a full-term birth, preterm birth, clinical pregnancy loss, or failure to conceive—can provide critical information during the early stages of a pregnancy attempt, when obstetricians are best positioned to intervene to improve the chances of successful conception and full-term live birth. Yet the task of constructing and validating risk assessment tools for this earlier intervention window is complicated by several statistical features: the final outcome of the pregnancy attempt is multinomial in nature, and it summarizes the results of two intermediate stages, conception and gestation, whose outcomes are subject to competing risks, measured on different time scales, and governed by different biological processes. In light of this complexity, existing pregnancy risk assessment tools largely focus on predicting a single adverse pregnancy outcome, and make these predictions at some later, post-conception time point.

Methods: We reframe the individual pregnancy attempt as a multistate model comprised of two nested multinomial prediction tasks: one corresponding to conception and the other to the subsequent outcome of that pregnancy. We discuss the estimation of this model in the presence of multiple stages of outcome missingness and then introduce an inverse-probability-weighted Hypervolume Under the Manifold statistic to validate the resulting multivariate risk scores. Finally, we use data from the Effects of Aspirin in Gestation and Reproduction (EAGeR) trial to illustrate how this multistate competing risks framework might be utilized in practice to construct and validate a preconception pregnancy risk assessment tool.

Results: In the EAGeR study population, the resulting risk profiles are able to meaningfully discriminate between the four pregnancy attempt outcomes of interest and represent a significant improvement over classification by random chance.

Conclusions: As illustrated in our analysis of the EAGeR data, our proposed prediction framework expands the pregnancy risk assessment task in two key ways—by considering a broader array of pregnancy outcomes and by providing the predictions at an earlier, preconception intervention window—providing obstetricians and their patients with more information and opportunities to successfully guide pregnancy attempts.

Keywords: Competing risks, Discrete survival models, Missing data, Multinomial classification, Pregnancy, Risk prediction

*Correspondence: kaitlyn_cook@harvardpilgrim.org

¹Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, US

Full list of author information is available at the end of the article



© The Author(s). 2022 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

The role of risk prediction in preconception care

In obstetrics and gynecology, preconception care refers to a “set of interventions that aim to identify and modify biomedical, behavioral, or social risks to [a potentially child-bearing individual’s] health or pregnancy outcome,” with particular emphasis placed on those factors that must be intervened on prior to conception or early in the pregnancy [1]. Integral to these efforts is the initial preconception consultation, a pre-pregnancy check-up during which an obstetrician may collect baseline demographic and medical history data in order to better guide and inform the planned pregnancy attempt. These consultations thus serve an important dual purpose: they assist patients in preparing for an upcoming pregnancy, while also representing the first point at which clinicians might intervene to improve the likelihood of conception and an eventual full-term live birth by, for example recommending immediate initiation of assisted reproductive technology [2] or by identifying and reviewing modifiable risk factors of adverse pregnancy outcomes [3, 4].

To that end, accurate and personalized predictions of both the time to conception and the outcome of any subsequent pregnancy are particularly relevant to preconception care [1]. Yet current pregnancy outcome prediction tools are limited in several key dimensions: existing risk scores typically (i) focus on prediction among individuals with already viable pregnancies, using as predictors biomarker and other biomedical data that may be unavailable or difficult to measure at an initial preconception visit [5–7], and (ii) either collapse all pregnancy outcomes into a single adverse event indicator [8, 9] or focus on prediction of a single pregnancy outcome [10, 11]. In focusing on a binarized representation of the pregnancy attempt, these models ignore the complex interdependencies and competing risks that may exist between pregnancy outcomes. This improper accounting may, in turn, have consequences for estimation biases in, and the predictive accuracy of, these post-conception prediction models [12, 13].

The few obstetric risk prediction tools explicitly developed for preconception contexts (such as those in Sep et al. [14], van Kuijk et al. [15], and Mehta-Lee et al. [16]) similarly consider only a single isolated adverse maternal or neonatal event, and are further limited by concerns of both internal and external validity. In particular, these models are often constructed and trained on retrospectively ascertained datasets whose inclusion criteria require that individuals (i) had a clinically-recognized pregnancy that (ii) resulted in a live birth during the course of the study period; such criteria systematically exclude individuals struggling with sub-fertility or clinical pregnancy loss—individuals who may also be at higher risk for other adverse pregnancy outcomes—while preventing the generalization of these models to practical

preconception clinical settings in which neither conception nor an eventual live birth are guaranteed. And while greater attention has been paid to the task of modeling fertility and time to conception in the statistical literature [17–21], to the best of our knowledge no obstetric/gynecological model has attempted to integrate these fertility predictions into a broader preconception risk assessment framework.

In short, the existing pregnancy-related prediction tools are either all adapted for patient populations and clinical scenarios that are defined downstream of the initial preconception visit or are too narrow in scope to be used to guide preconception care.

The effects of aspirin in gestation and reproduction trial

To address this limitation, we aim to develop a clinical risk assessment tool that is implementable during the course of a pre-pregnancy check-up and that simultaneously considers both (i) the likelihood of a clinically-recognized pregnancy occurring and (ii) the likelihood of that pregnancy ending in either a full-term live birth or an adverse event. This prediction task effectively reframes the pregnancy attempt as a multistate competing risks process comprised of two stages—a conception stage resulting in an implementation event (or a lack thereof) and a gestation stage resulting in a final birth outcome—that represent fundamentally different biological processes and thus may be governed by fundamentally different sets of risk factors [22].

One of the few existing pregnancy studies with data on both of these outcome stages, as well as a rich set of potential predictors, is the Effects of Aspirin in Gestation and Reproduction (EAGeR) trial, a randomized controlled trial of the effects of low-dose preconception aspirin use on both the times to conception and the subsequent birth outcomes for women with at least one prior pregnancy loss [23]. The study enrolled 1228 women at the time of their initial preconception consultation, and then followed these women for one of four mutually exclusive pregnancy attempt outcomes:

1. *failure to conceive* within six menstrual cycles of the initial preconception visit;
2. *pregnancy resulting in a clinical pregnancy loss*, including both spontaneous abortion and stillbirth;
3. *pregnancy resulting in a preterm birth*, defined as a live birth occurring at or before 37 weeks’ gestation; or
4. *pregnancy resulting in a full-term birth*, defined as a live birth occurring after 37 week’s gestation.

By treating the initial preconception visit as the time origin from which women are then prospectively followed for an implantation event, the EAGeR dataset overcomes the left and right truncation issues typically noted in time to

pregnancy studies [24]; by recording pregnancy attempt outcomes on all women—regardless of whether those women ultimately fail to conceive within the study period or experience a clinical pregnancy loss—the trial circumvents the selection and generalizability concerns endemic to other preconception datasets.

It does, however, present different statistical challenges to the construction and validation of a preconception prediction model. In light of the relatively short follow-up during the conception stage of the EAGeR study, as well as the complex biological restrictions that naturally exist on the support of the subsequent birth outcomes, correct specification of the transition intensities for the various pregnancy attempt outcomes is difficult. Furthermore, the time-to-event data on each stage of the pregnancy attempt were recorded with different levels of granularity; the gestation-related outcomes (namely clinical pregnancy loss, preterm birth, and full-term birth) were reported in terms of gestational age, a continuous measure of time since implantation, while the conception outcome was reported in terms of menstrual cycles since the preconception visit. This partial coarsening of the time axis further complicates the use of traditional multistate modeling analyses for the preconception prediction of pregnancy outcomes, as these analyses typically assume that the outcome process unfolds and is measured entirely in either continuous or discrete time [25]. Finally, during the conduct of the EAGeR trial, women were lost to follow-up at both stages of the outcomes process—prior to the observation of a clinically-recognized pregnancy, as well as prior to the observation of that pregnancy’s outcome—such that the missing data process was also multistage in nature. As a consequence, we must contend with multiple possible sources of estimation bias when constructing the final risk prediction model, both as a consequence of missingness in the pregnancy attempt outcomes themselves and as a consequence of missingness in the selection event for the second-stage outcomes.

Outline

In what follows, we present a unified framework for the preconception prediction of pregnancy outcomes that

adapts the traditional multistate competing risks model in order to accommodate the coarsening of the timescale information and the multistage nature of the missingness processes, while also circumventing concerns regarding biological plausibility. In the “Methods” section, we first introduce this prediction framework in greater detail, and then discuss its estimation and inference in the presence of complete data (“Complete data setting” section) as well as extensions to the missing at random setting (“Estimation and validation in the presence of outcome missingness” section). We also present an analysis of data from the EAGeR trial (“Preconception risk prediction in the EAGeR study population” section), illustrating how preconception risk prediction might be accomplished and utilized by clinicians in practice. The results of this analysis are given in the “Results” section, and we conclude with a brief discussion in the “Discussion” section and conclusions in the “Conclusion” section.

Methods

Prediction in a multistate competing risks framework

Suppose that individual i wishes to conceive, and schedules a preconception obstetrics consultation in order to plan their current pregnancy attempt. We let the stochastic process $(Y_i(t), t \geq 0)$ characterize the status of this attempt at time t since preconception visit, with $Y_i(t)$ taking on values over some discrete state space \mathcal{S} that captures all potential pregnancy outcomes of interest. In the motivating EAGeR trial, for example, this state space is $\mathcal{S} = \{0, 1, 2, 3, 4\}$, where $\{0\}$ indicates the absence of a successful implantation event, $\{1\}$ an active and ongoing pregnancy, $\{2\}$ a clinical pregnancy loss, $\{3\}$ a preterm birth, and $\{4\}$ a full-term birth. More complex constructions of \mathcal{S} including additional terminal (e.g., elective abortion) and non-terminal (e.g., the development of gestational diabetes or pre-eclampsia) events are possible but not considered here. We assume that the transitions between these states are governed by the multistate model shown in Fig. 1, where

$$\lambda_{jk}(t) = \lim_{\delta \rightarrow 0} P \{ Y_i(t^- + \delta) = k | Y_i(t^-) = j, \mathcal{Y}_{it^-} \}$$

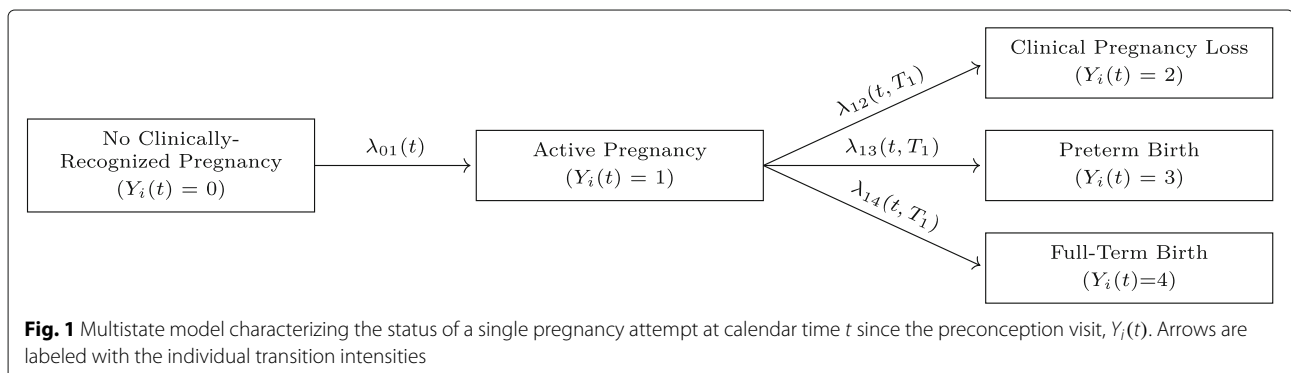


Fig. 1 Multistate model characterizing the status of a single pregnancy attempt at calendar time t since the preconception visit, $Y_i(t)$. Arrows are labeled with the individual transition intensities

characterizes the instantaneous probability of transitioning from pregnancy outcome state j to pregnancy outcome state k at time t since the preconception visit, conditional on the history of the pregnancy attempt up until that time, \mathcal{Y}_{it^-} . Note that the model in Fig. 1 naturally partitions the pregnancy attempt into a conception stage (the $0 \rightarrow 1$ transition) and a gestation stage (the $1 \rightarrow 2$, $1 \rightarrow 3$, and $1 \rightarrow 4$ transitions); we assume that the transition intensity for the conception stage depends only on the time t since the preconception visit, and that the transition intensities for the gestation stage depend additionally on the time to clinically-recognized pregnancy, $T_1 = \inf\{t : Y_i(t) = 1\}$. Given this framework, our aim is to make meaningful predictions about individual i 's pregnancy attempt, $Y_i(t)$ —and in particular to construct an individualized risk profile characterizing the likely outcomes of that attempt—based on the covariate information, \mathbf{w}_i , that is routinely collected as part of the preconception consultation.

Several authors have considered how to incorporate covariate information into the estimation of the transition intensities, $\lambda_{jk}(t)$, and the state occupation probabilities, $\pi_{ij}(t) = P\{Y_i(t) = j\}$, of semi-Markovian stochastic processes like the multistate model in Fig. 1 (see Andersen and Perme [25] for an in-depth review). However, the nature of preconception pregnancy prediction in general—and the structure of the EAGeR data in particular—present several challenges to the construction of individual preconception risk profiles in terms of either of these quantities.

1. Directly specifying and estimating $\lambda_{jk}(t)$ in terms of \mathbf{w}_i and some parametric baseline intensity function, as discussed in Kalbfleisch and Prentice [26] and Beyersmann et al. [22] among others, is complicated by the complex restrictions that exist on the support of the gestational outcomes: clinical pregnancy loss, preterm birth, and full-term birth. The possible calendar times t at which $\lambda_{12}(t)$, $\lambda_{13}(t)$, and $\lambda_{14}(t)$ are each non-zero are dictated by limits on fetal viability, biological and medical constraints on pregnancy duration, and the definitions of the outcomes themselves, such that biologically plausible—let alone correct—specification of the baseline intensity functions is unlikely.
2. Alternatively, direct estimation of the state occupation probabilities using either the pseudo-value approach of Andersen and Klein [27] or the weighted estimating equation approach of Scheike and Zhang [28] and Scheike et al. [29] is complicated by the lack of granularity in the EAGeR event time data, and in pregnancy outcome reporting more generally. In particular, time to pregnancy is typically measured and reported in terms of menstrual cycles to conception, which partitions calendar time into

discrete time units; the gestational outcomes (clinical pregnancy loss, preterm birth, and full-term birth) are instead reported in *continuous* time (gestational age) with an unknown and estimated time origin: the exact calendar time of conception. Using such data to estimate attributes of a single stochastic process $(Y_i(t), t \geq 0)$ with a cohesive and consistent internal time scale is challenging.

3. Similar data coarsening issues arise when accounting for censoring in the estimation task. In the EAGeR trial, for example, while exact loss to follow-up times are available for the conception stage of the stochastic process (i.e., for the $0 \rightarrow 1$ transition), the remaining pregnancy outcomes were determined using medical chart abstraction, so that only a binary indicator of missingness was available for the $1 \rightarrow 2$, $1 \rightarrow 3$, and $1 \rightarrow 4$ transitions. The mechanisms driving missingness in the conception stage may also differ substantially from the mechanisms driving missingness in the clinical result of that conception.
4. Finally, while the multistate model for a single pregnancy attempt naturally includes a non-absorbing “active pregnancy” state (Fig. 1), this state is only of indirect interest for our particular prediction task. The relative likelihoods of occupying states $\{0\}$, corresponding to no clinically-recognized pregnancy; $\{2\}$, corresponding to clinical pregnancy loss; $\{3\}$, corresponding to preterm live birth; and $\{4\}$, corresponding to full-term live birth, have clear clinical and prognostic value for the pregnancy attempt. State $\{1\}$, corresponding to active pregnancy, is relevant only in so much as it defines the population of individuals on whom the gestation-related outcomes will be observed.

In light of these challenges, we instead define the individual preconception risk profile as

$$p_i(\tau) := \left\{ \pi_{i0}(\tau), \pi_{i2|1}(\tau)\pi_{i-0}(\tau), \pi_{i3|1}(\tau)\pi_{i-0}(\tau), \pi_{i4|1}(\tau)\pi_{i-0}(\tau) \right\}, \quad (1)$$

where $\pi_{i-0}(\tau) := 1 - \pi_{i0}(\tau)$ and where

$$\pi_{ij|1}(\tau) := \lim_{t \rightarrow \infty} P\{Y_i(t) = j | T_1 \leq \tau\}, \quad j = 2, 3, 4$$

are limiting conditional state occupation probabilities defined over a conception window of τ menstrual cycles. $p_i(\tau)$ thus characterizes the likelihood that individual i fails to conceive within τ menstrual cycles of the preconception consultation, or that they successfully conceive within that window and that the subsequent pregnancy results in either a clinical pregnancy loss, a preterm birth, or a full-term birth: $\pi_{i0}(\tau)$, $\pi_{i2|1}(\tau)\pi_{i-0}(\tau)$, $\pi_{i3|1}(\tau)\pi_{i-0}(\tau)$, and $\pi_{i4|1}(\tau)\pi_{i-0}(\tau)$, respectively.

Constructing $p_i(\tau)$ in this fashion allows us to reformulate $(Y_i(t), t \geq 0)$ as the composition of a first-stage

outcome, $Y_{i1} \in \{0, 1\}$, which represents conception within τ menstrual cycles, and a (potentially undefined) second-stage outcome, $Y_{i2} \in \{2, 3, 4\}$, which represents the final result of that pregnancy (Fig. 2). We may then estimate $p_i(\tau)$ in terms of the baseline covariate information \mathbf{w}_i by separately modeling the outcomes of each of these stages; this two-stage approach is reminiscent of the nested competing risks construction of [22] and allows us to explicitly account for the different timescales, the different governing risk factors, and the different censoring patterns of the conception and gestation processes while avoiding concerns regarding the biological plausibility of the underlying transition intensity model.

Model estimation and validation

Complete data setting

When the reformulated pregnancy attempt outcome $\mathbf{Y}_i = (Y_{i1}, Y_{i2})$ is completely observed for all n individuals, we take

$$\text{logit} \{P(Y_{i1} = 1|\mathbf{x}_i)\} = \boldsymbol{\beta}^T \mathbf{x}_i \tag{2}$$

$$\text{log} \left\{ \frac{P(Y_{i2} = k|\mathbf{z}_i, Y_{i1} = 1)}{P(Y_{i2} = 4|\mathbf{z}_i, Y_{i1} = 1)} \right\} = \boldsymbol{\alpha}_k^T \mathbf{z}_i, \quad k = 2, 3 \tag{3}$$

where $\mathbf{x}_i \subset \mathbf{w}_i$ and $\mathbf{z}_i \subset \mathbf{w}_i$ are the baseline covariates informing the conception and gestation processes, respectively, and where the first element of \mathbf{x}_i and \mathbf{z}_i is assumed to be a 1 if an intercept is included. Note that the dependence of models (2) and (3) on the conception window τ is implicit in the definition of $Y_{i1} = I(T_{1i} \leq \tau)$: different choices of τ will naturally lead to different definitions of the outcome for model (2) and the selection event for model (3), which might in turn lead to differences in the true values of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. We further note that the second-stage model in (3) is agnostic to the timing of the clinically-recognized pregnancy beyond the implicit restriction that $Y_{i1} = 1 \iff T_{1i} \leq \tau$; more complex specifications of (3), in which we first predict a time to conception and then incorporate this prediction into (3),

are possible but not explored here. We then use standard maximum likelihood estimation to fit (2) to the full sample of n individuals and (3) to the subset of individuals with at least one clinically-recognized pregnancy within τ menstrual cycles of their preconception visit. From the resulting estimates of $\boldsymbol{\beta}$, $\boldsymbol{\alpha}_2$, and $\boldsymbol{\alpha}_3$, we have

$$\begin{aligned} \hat{\pi}_{i-0}(\tau) &= \text{expit}(\hat{\boldsymbol{\beta}}^T \mathbf{x}_i) \\ \hat{\pi}_{i2|1}(\tau) &= \hat{\pi}_{i4|1}(\tau) \exp(\hat{\boldsymbol{\alpha}}_2^T \mathbf{z}_i) \\ \hat{\pi}_{i3|1}(\tau) &= \hat{\pi}_{i4|1}(\tau) \exp(\hat{\boldsymbol{\alpha}}_3^T \mathbf{z}_i) \\ \hat{\pi}_{i4|1}(\tau) &= \left\{ 1 + \exp(\hat{\boldsymbol{\alpha}}_2^T \mathbf{z}_i) + \exp(\hat{\boldsymbol{\alpha}}_3^T \mathbf{z}_i) \right\}^{-1}, \end{aligned}$$

and we obtain the predicted risk profile for individual i , $\hat{p}_i(\tau)$, by substituting these estimates into (1).

To quantify the extent to which these risk profiles meaningfully discriminate between the four pregnancy outcomes, we use the Hypervolume Under the Manifold (HUM) statistic, a generalization of the Area Under the Curve (AUC) statistic to outcomes with $K > 2$ outcome classes. Much as the AUC measures the correct classification rate of discordant pairs of binary outcomes, the HUM measures the correct classification rate of sets of K individuals, with one individual from each of the K outcome classes. Operationalizing the HUM thus requires specifying a decision rule with which to conduct this K -alternative forced-choice decision task. To that end, let \mathbf{e}_k be the k th basis vector, here indicating assignment to outcome class k ($k = 1, \dots, K$). We consider the following general class of decision rules, which selects the classification (c_1, c_2, \dots, c_K) that minimizes the weighted sum of the Euclidean distances from each individual's risk profile to their assignment vector:

$$\begin{aligned} \underset{(c_1, c_2, \dots, c_K)}{\text{argmin}} \quad & \left(\omega_1 \|p^{(1)}(\tau) - \mathbf{e}_{c_1}\| + \omega_2 \|p^{(2)}(\tau) - \mathbf{e}_{c_2}\| \right. \\ & \left. + \dots + \omega_K \|p^{(K)}(\tau) - \mathbf{e}_{c_K}\| \right), \end{aligned} \tag{4}$$

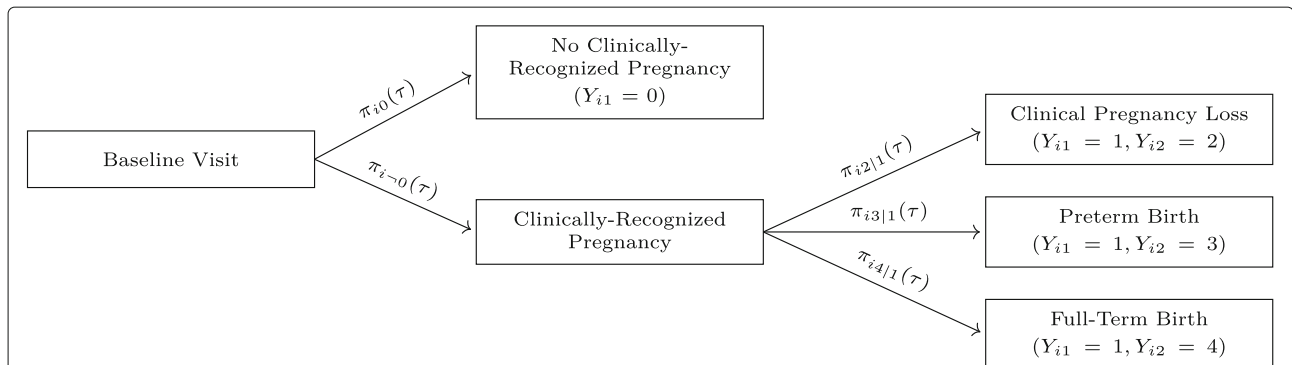


Fig. 2 Reformulation of $(Y_i(t), t \geq 0)$ as a binomial first-stage outcome, Y_{i1} , composed with a multinomial second-stage outcome, Y_{i2} , conditional on a conception window of τ menstrual cycles. Arrows are labeled with the relevant individual (conditional) state occupation probabilities

where $p^{(k)}(\tau)$ is the risk profile for the individual from class k and where $\|\cdot\|$ is the standard L_2 norm. Note that the choice of weight vector, $\omega = (\omega_1, \dots, \omega_K)$, allows one to differentially prioritize the discriminatory ability of $p_i(\tau)$ with respect to each of the different outcome classes; a weighting scheme in which $\omega_1/\|\omega\| > 0.25$ will, for example, result in an HUM that rewards risk profiles with high discriminatory ability with respect to the first outcome class. Appropriate selection of ω is thus highly context dependent. In the absence of any strong rationale to the contrary, we recommend implementing (4) with all outcomes weighted equally, i.e., with ω set to $\mathbf{1}$, so that the resulting HUM speaks to overall classification performance and may be more readily compared across analyses and applications.

Let $CR(\cdot)$ indicate whether all K individuals have been correctly classified on the basis of their risk profiles and the selected decision rule. Then

$$HUM = P\left[CR\left\{p^{(1)}(\tau), p^{(2)}(\tau), \dots, p^{(K)}(\tau)\right\} = 1\right]$$

and a nonparametric estimator of the HUM is

$$\widehat{HUM} = \frac{1}{n_1 n_2 \dots n_K} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \dots \sum_{i_K=1}^{n_K} CR\left\{\widehat{p}_{i_1}^{(1)}(\tau), \widehat{p}_{i_2}^{(2)}(\tau), \dots, \widehat{p}_{i_K}^{(K)}(\tau)\right\}, \quad (5)$$

where n_k is the number of individuals belonging to outcome class k and $\widehat{p}_{i_k}^{(k)}(\tau)$ is the estimated risk profile for the i_k th individual in that class. Note that if the risk profiles have no inherent predictive value, such that they correspond to classification by random chance, then the correct classification rate will simply be $HUM = 1/K!$. For pregnancy outcome prediction with $K = 4$ outcome classes, this non-informative $HUM = 1/4! \approx 0.0417$.

Estimation and validation in the presence of outcome missingness

Longitudinal pregnancy outcome studies often, however, feature loss to follow-up and study withdrawal, which may occur at any point during the pregnancy attempt. As a result, $Y_i = (Y_{i1}, Y_{i2})$ may be partially or completely unobserved for some subset of individuals in the study. Let R_{i1}, R_{i2} , and R_i be indicators of non-missingness in Y_{i1}, Y_{i2} , and Y_i , respectively, where $R_i = I(R_{i1} = 1 \cap R_{i2} = 1)$. Then study subjects may be classified as belonging to one of four observed data categories:

1. Censored prior to conception ($R_{i1} = 0, R_{i2} = 0$)
2. No pregnancy within τ menstrual cycles ($R_{i1} = 1, R_{i2} = 1$)

3. Pregnancy within τ menstrual cycles, unknown result ($R_{i1} = 1, R_{i2} = 0$)
4. Pregnancy within τ menstrual cycles, known result ($R_{i1} = 1, R_{i2} = 1$)

For an individual to have complete data with respect to model (2), their pregnancy status after τ cycles must be known ($R_{i1} = 1$); for an individual to have complete data with respect to model (3), both their pregnancy status and, provided that a clinically-recognized pregnancy occurred, the final result of that pregnancy must be observed ($R_{i1} = 1, Y_{i1} = 1, R_{i2} = 1$).

Modeling and predicting Y_i thus requires addressing potentially multiple stages of outcome missingness, where the factors governing that missingness may vary according to stage. To that end, we assume that Y_i is missing sequentially at random, so that

$$\begin{aligned} P(R_{i1} = 1 | \mathbf{w}_i, Y_{i1}) &= P(R_{i1} = 1 | \mathbf{x}'_i) \\ P(R_{i2} = 1 | \mathbf{w}_i, R_{i1} = 1, Y_i) &= P(R_{i2} = 1 | \mathbf{z}'_i, R_{i1} = 1, Y_{i1}) \\ P(R_i = 1 | \mathbf{w}_i, Y_i) &= P(R_{i2} = 1 | \mathbf{z}'_i, R_{i1} = 1, Y_{i1}) \\ &P(R_{i1} = 1 | \mathbf{x}'_i), \end{aligned}$$

where $\mathbf{x}'_i \subset \mathbf{w}_i$ and $\mathbf{z}'_i \subset \mathbf{w}_i$ are the baseline covariates informing the first and second stages of the missingness process, respectively. We then (i) reframe estimation of $\pi_{i \rightarrow 0}(\tau) = P(Y_{i1} = 1 | \mathbf{x}_i)$ in terms of a discrete-time survival model for time to clinically-recognized pregnancy, and (ii) implement the modularized missing data framework of Haneuse and Daniels [30] using either multiple imputation or inverse probability of censoring weights to address selection into and estimation of the model for $\pi_{i \rightarrow k}(\tau) = P(Y_{i2} = k | \mathbf{z}_i, Y_{i1} = 1)$ with $k = 2, 3, 4$.

Let, as before, T_{1i} be the time to clinically-recognized pregnancy, and now take C_i to be the time to censoring or withdrawal from the first stage of Fig. 1. As a result, we do not necessarily observe $Y_{i1} = I(T_{1i} \leq \tau)$ directly, but rather observe (U_i, δ_i) , where $U_i = (T_{1i} \wedge C_i) \wedge \tau$, $\delta_i = I[T_{1i} \leq (C_i \wedge \tau)]$, and “ \wedge ” denotes the minimum of its arguments. Note that $\delta_i \equiv Y_{i1}$ for all individuals with $R_{i1} = 1$; for all remaining individuals, we have only the partial information that $T_{1i} > C_i$. To leverage this when estimating $\pi_{i0}(\tau)$, we additionally assume that $T_{1i} \perp\!\!\!\perp C_i | \mathbf{x}_i$ and replace the logistic regression model in (2) with the discrete-time survival model

$$\text{logit}[P(T_{1i} = t | T_{1i} \geq t, \mathbf{x}_i)] = g_t + \boldsymbol{\gamma}^T \mathbf{x}_i, \quad t = 1, 2, \dots, \tau \quad (6)$$

where $P(T_{1i} = t | T_{1i} \geq t, \mathbf{x}_i)$ is the transition intensity function $\lambda_{01}(t)$ under a discrete-time representation of the conception process.

To fit model (6), let $\delta_{it} = I(\delta_i = 1 \cap U_i = t)$. Then after rearrangement, the log-likelihood of $\mathbf{g} = (g_1, \dots, g_\tau)$ and $\boldsymbol{\gamma}$ may be written as

$$\begin{aligned} & \log L(\mathbf{g}, \boldsymbol{\gamma} | \mathbf{U}, \boldsymbol{\delta}, \mathbf{x}) \\ &= \sum_{i=1}^n \sum_{t=1}^{u_i} \delta_{it} \text{logit} [P(T_{1i} = t | T_{1i} \geq t, \mathbf{x}_i)] \\ & \quad + \sum_{i=1}^n \sum_{t=1}^{u_i} \log [1 - P(T_{1i} = t | T_{1i} \geq t, \mathbf{x}_i)] \\ &= \sum_{i=1}^n \sum_{t=1}^{u_i} \left\{ \delta_{it} (\mathbf{g}_t + \boldsymbol{\gamma}^T \mathbf{x}_i) - \log [1 + \exp(\mathbf{g}_t + \boldsymbol{\gamma}^T \mathbf{x}_i)] \right\}, \end{aligned}$$

which is the log-likelihood for a logistic regression model fit to $N = \sum_{i=1}^n U_i$ independent observations with δ_{it} as the outcome [31]. Estimation of \mathbf{g} and $\boldsymbol{\gamma}$ then proceeds by:

1. Creating U_i duplicates of individual i 's record, with each duplicate corresponding to a distinct unit of time during which individual i was under observation.
2. Recording, for each unit of time, the binary outcome δ_{it} . By definition, $\delta_{it} = 0$ for observation times $t = 1, \dots, U_i - 1$. If individual i was censored, then $\delta_{iU_i} = 0$ as well; otherwise, $\delta_{iU_i} = 1$.
3. Using standard maximum likelihood estimation to fit the logistic regression model $\text{logit}\{P(\delta_{it} = 1 | \mathbf{x}_i)\} = \mathbf{g}_t + \boldsymbol{\gamma}^T \mathbf{x}_i$ under working independence.

We may then recover $\hat{\pi}_{i-0}(\tau) = \hat{P}(Y_{i1} = 1 | \mathbf{x}_i) = \hat{P}(T_{1i} \leq \tau | \mathbf{x}_i)$ by taking

$$\begin{aligned} \hat{P}(T_{1i} \leq \tau | \mathbf{x}_i) &= \sum_{t=1}^{\tau} \left[\hat{P}(T_{1i} = t | T_{1i} \geq t, \mathbf{x}_i) \right. \\ & \quad \left. \times \prod_{j=1}^{t-1} \{1 - \hat{P}(T_{1i} = t - j | T_{1i} \geq t - j, \mathbf{x}_i)\} \right] \\ &= \sum_{t=1}^{\tau} \left[\text{expit}(\hat{\mathbf{g}}_t + \hat{\boldsymbol{\gamma}}^T \mathbf{x}_i) \prod_{j=1}^{t-1} \left\{ 1 - \text{expit}(\hat{\mathbf{g}}_{t-j} + \hat{\boldsymbol{\gamma}}^T \mathbf{x}_i) \right\} \right]. \end{aligned}$$

Estimation of the remaining components of the personalized risk profile requires fitting model (3), which in turn requires addressing missingness in both the selection event, Y_{i1} , and the outcome, Y_{i2} . To do so, we modularize the data provenance for Y_i into two submechanisms—one corresponding to whether or not the implantation event (or lack thereof) is observed within τ menstrual cycles, and the other corresponding to whether or not the final clinical result of that pregnancy is recorded—and use either multiple imputation or inverse probability of censoring weights to address missingness at each of these stages.

In the sequential multiple imputation analysis of (3), we start by imputing the selection event $I(T_{1i} \leq \tau)$, which is necessary for delineating the subgroup on whom model (3) will subsequently be fit. This selection event is, in turn, a function of the time to clinically-recognized pregnancy, T_{1i} . Thus, to fit model (3) in the presence of two-stage outcome missingness, we first impute T_{1i} for all individuals with $R_{i1} = 0$ using the multinomial model

$$\log \left\{ \frac{P(T_{1i} = t | T_{1i} > u_i, \mathbf{x}_i^*)}{P(T_{1i} = u_i + 1 | T_{1i} > u_i, \mathbf{x}_i^*)} \right\} = \boldsymbol{\zeta}_{u_i}^T \mathbf{x}_i^*, \quad t = u_i + 1, \dots, \tau + 1,$$

where $\mathbf{x}_i^* \subset \mathbf{w}_i$ comprises both the predictors \mathbf{x}_i and any other auxiliary variables thought helpful for the imputation process (such as \mathbf{x}'_i ; see [32] for guidance on specifying the imputation model) and where the mass at cycle $t = \tau + 1$ corresponds to administrative censoring at time τ . We then transform the imputed \tilde{T}_{1i} to obtain the corresponding indicator of an implantation event, $\tilde{Y}_{i1} = I(\tilde{T}_{1i} \leq \tau)$. For all individuals meeting the selection criteria for model (3) (i.e., with either $Y_{i1} = 1$ or $\tilde{Y}_{i1} = 1$) but with missing second-stage outcome data (i.e., with $R_{i2} = 0$), we also conduct a second imputation for the final pregnancy result, \tilde{Y}_{i2} . We repeat this two-stage imputation procedure M times, and in the m th completed dataset fit model (3) to all individuals with either an observed or imputed implantation event within τ menstrual cycles. Combining the resulting point estimates, we find $\hat{\boldsymbol{\alpha}} = M^{-1} \sum_{m=1}^M \hat{\boldsymbol{\alpha}}^{(m)}$. Alternatively, in the two-stage inverse-probability-weighted analysis of model (3), we first estimate the probability of having complete data with respect to model (3),

$$\begin{aligned} \pi_i^C &= P(R_{i2} = 1, Y_{i1} = 1, R_{i1} = 1 | \mathbf{w}_i) \\ &= P(R_{i1} = 1 | \mathbf{z}'_i, R_{i1} = 1, Y_{i1} = 1) \\ & \quad P(Y_{i1} = 1 | \mathbf{x}_i, R_{i1} = 1) P(R_{i1} = 1 | \mathbf{x}'_i), \end{aligned}$$

using (for example) a series of logistic regression models, and arrive at $\hat{\boldsymbol{\alpha}}$ by fitting (3) to the complete cases, weighted by $1/\hat{\pi}_i^C$. Recent work has also considered the task of model estimation and inference when different missing data techniques are adopted for each of the proposed sub-mechanisms [33, 34]. Regardless of how final point estimates for (3) are obtained, given the resulting $\hat{\boldsymbol{\alpha}}$ we may then estimate $\hat{\pi}_{i2|1}(\tau)$, $\hat{\pi}_{i3|1}(\tau)$, and $\hat{\pi}_{i4|1}(\tau)$ as in the previous section.

Remark 1 For a given data provenance sub-mechanism, the choice between conducting multiple imputation and conducting inverse probability weighting should largely be informed by which model the analyst feels better able to specify correctly: the imputation model or the weighting

model. The former requires modeling the full conditional distribution of the missing data given the observed data—which (at a minimum) draws upon knowledge of the outcome process, and which can become challenging if additional covariates beyond the outcome are missing—while the latter requires modeling $P(R_1 = 1)$ and draws upon knowledge of the missingness process. Misspecification of these models will, in general, lead to bias in the point estimates $\hat{\alpha}$ from the corresponding imputed or weighted final analysis, though iterative multiple imputation procedures such as multiple imputation by chained equations have been shown to be robust to misspecification of the full conditional distribution so long as the component univariate conditional distributions are valid [35]. Other considerations for selecting a missing data approach include efficiency (when the imputation and weighting models are both correctly specified, multiple imputation is known to be more efficient than inverse probability weighting) and the stability of the inverse probability weights (in settings where either $\hat{P}(R_{i1} = 1|\mathbf{x}'_i)$ or $\hat{P}(R_{i2} = 1|\mathbf{z}'_i, R_{i1} = 1, Y_{i1})$ is near zero for certain levels of \mathbf{x}'_i or \mathbf{z}'_i , the corresponding inverse probability weights are large and the resulting estimator may be unstable).

We must also account for both partial and complete missingness in $Y_i = (Y_{i1}, Y_{i2})$ when quantifying the discriminatory performance of the predicted risk profiles, $\hat{p}_i(\tau)$. In particular, the nonparametric HUM estimator given in (5) requires knowledge of each individual’s true pregnancy outcome; a complete-case implementation—in which (5) is estimated using only the subset of individuals for whom this outcome is completely observed—may be subject to verification bias unless both Y_{i1} and Y_{i2} are missing completely at random. Zhang and Alonzo [36] consider verification bias adjustment in the three outcome classification setting in which the outcomes of interest are ordinal and the classification decision is based on a continuous measurement; under the additional assumption that missingness occurs at random, they develop an inverse-probability-weighted version of the nonparametric HUM estimator. Here, we extend this result to $K = 4$ unordered outcome classes, missing sequentially at random, where classification occurs on the basis of a multinomial risk profile.

Let R_{i1} , R_{i2} , and R_i be defined as before, and let n_k^* be the number of individuals in outcome class k with $R_i = 1$. Then a verification-bias-adjusted estimator for the HUM is

where $\hat{p}_{i_k}^{(k)}(\tau)$ is the estimated risk profile for the i_k th individual in outcome class k , and where $\hat{\pi}_{i_k}^R$ is the estimated probability of fully observing Y_i for that same individual:

$$\hat{\pi}_i^R = \hat{P}(R_{i2} = 1|\mathbf{z}'_i, R_{i1} = 1, Y_{i1}) \hat{P}(R_{i1} = 1|\mathbf{x}'_i).$$

Preconception risk prediction in the EAGeR study population

We illustrate the broad utility of this multistate competing risks framework for pregnancy outcomes by constructing and validating a preconception risk assessment tool on the study population of the Effects of Aspirin in Gestation and Reproduction (EAGeR) trial [23, 37]. As discussed in the “Background” section, the trial enrolled 1228 women who were actively attempting to conceive, and who were between the ages of 18 and 40, had a history of 1–2 previous pregnancy losses, had up to two previous live births, and had at most one elective termination or ectopic pregnancy; these women had no prior history of infertility or subfertility, and were neither currently undergoing nor planning to undergo medical fertility treatment during the course of the trial. Each woman attended a baseline preconception clinic visit, at which point biomedical, sociodemographic, and medical data were collected. The women were then randomized to take either low-dose aspirin or a placebo over the remainder of their pregnancy attempt. The women were followed for up to six menstrual cycles for the occurrence of a clinically-recognized pregnancy and—for those women who successfully conceived during the study period—were subsequently followed for the outcome of that pregnancy (Table 1).

Towards building a risk prediction model for the outcome of these pregnancy attempts given a conception window of $\tau = 6$ menstrual cycles, we restricted our attention to the 1093 women who had been attempting to conceive for fewer than twelve menstrual cycles at the start of the trial and who contributed at least one cycle of follow-up. We considered the following covariates as possible predictors, \mathbf{w}_i : preconception aspirin use (0, placebo; 1, low-dose aspirin), age at enrollment, body mass index (BMI), hypertensive status (0, non-hypertensive; 1, hypertensive), smoking history over the past year (0, none; 1, at least one cigarette), race (0, other; 1, non-Hispanic white), income level ($\leq \$19,999$; $\$20,000–\$39,999$; $\$40,000–\$74,999$; $\$75,000–\$99,999$; $\geq \$100,000$), educational attainment (0, some college or less; 1, college or postgraduate degree), parity (0, nulliparous; 1, parous),

$$\frac{\sum_{i_1=1}^{n_1^*} \sum_{i_2=1}^{n_2^*} \sum_{i_3=1}^{n_3^*} \sum_{i_4=1}^{n_4^*} (\hat{\pi}_{i_1}^R \hat{\pi}_{i_2}^R \hat{\pi}_{i_3}^R \hat{\pi}_{i_4}^R)^{-1} CR(\hat{p}_{i_1}^{(1)}(\tau), \hat{p}_{i_2}^{(2)}(\tau), \hat{p}_{i_3}^{(3)}(\tau), \hat{p}_{i_4}^{(4)}(\tau))}{\sum_{i_1=1}^{n_1^*} \sum_{i_2=1}^{n_2^*} \sum_{i_3=1}^{n_3^*} \sum_{i_4=1}^{n_4^*} (\hat{\pi}_{i_1}^R \hat{\pi}_{i_2}^R \hat{\pi}_{i_3}^R \hat{\pi}_{i_4}^R)^{-1}}, \tag{7}$$

Table 1 Observed pregnancy outcomes for the $n = 1093$ EAGeR participants who were not clinically infertile at the baseline preconception visit

Pregnancy Outcome ^a	R	Y	N (%)
Censored prior to conception	$(R_1 = 0, R_2 = 0)$	—	88 (8.1%)
No pregnancy within 6 cycles	$(R_1 = 1, R_2 = 1)$	$Y_1 = 0$	319 (29.2%)
Pregnancy within 6 cycles			
Unknown result	$(R_1 = 1, R_2 = 0)$	$(Y_1 = 1, —)$	12 (1.1%)
Clinical pregnancy loss	$(R_1 = 1, R_2 = 1)$	$(Y_1 = 1, Y_2 = 2)$	122 (11.2%)
Preterm birth	$(R_1 = 1, R_2 = 1)$	$(Y_1 = 1, Y_2 = 3)$	49 (4.5%)
Full-term birth	$(R_1 = 1, R_2 = 1)$	$(Y_1 = 1, Y_2 = 4)$	501 (45.8%)

^aA pregnancy was considered to be clinically-recognized if it was confirmed by an ultrasound scan at 6–7 weeks' gestation. Clinical pregnancy loss comprised both miscarriages (prior to 20 weeks' gestation) and stillbirths (after 20 weeks' gestation), preterm birth comprised live births prior to 37 weeks' gestation, and full-term birth comprised live births after 37 weeks' gestation

number of previous pregnancy losses (either 1 or 2 per study inclusion criteria), and number of menstrual cycles spent attempting to conceive at the time of enrollment.

Implementation of the proposed model and validation metric

Given that women in the EAGeR trial reported missingness at both stages of the pregnancy outcome process (Table 1), as well as at low levels in the available baseline covariate data (BMI, 1.4%; hypertensive status, 0.5%; smoking status, 0.7%; educational attainment, 0.7%), we estimated and validated the personalized risk profiles according to the procedure proposed in the “[Estimation and validation in the presence of outcome missingness](#)” section. In particular, we conducted multiple imputation by chained equations to create $M = 10$ imputed datasets: for the first-stage model in (6), we imputed missing baseline covariate data only, while for the second-stage model in (3), we also used a sequential multiple imputation procedure to impute both missing selection events and missing pregnancy outcome data. We then selected predictors separately for the first- and second-stage models using an AIC-based backward selection procedure applied to the combined dataset of Mn individuals in order to decide among all possible first-order covariate effects [38]. Both race and number of previous losses were a priori thought to be clinically relevant, and so were manually included in both stages of the outcome model. We fit the selected models on each imputed dataset separately—with the first-stage model fit using all fertile women in the EAGeR trial, and the second-stage model fit using all women with either an observed or imputed clinically-recognized pregnancy—and combined the point estimates across the datasets using the standard Rubin's method.

To quantify the predictive capability of the final pregnancy outcome model, we implemented the verification-bias-adjusted HUM estimator in (7). In light of the low levels of second-stage outcome missingness (only 12 women had recognized pregnancies but no recorded final pregnancy outcome), we fit a single logistic regression model to estimate the probability of verification weights, $\hat{\pi}_i^R = \hat{P}(R_i = 1 | \mathbf{w}_i)$, and once again used backward selection followed by Rubin's method to arrive at the final model for the verification weights.

Finally, to adjust for the optimistic prediction performance inherent to both fitting and validating models (6) and (3) on the full EAGeR dataset, we implemented the optimism correction described in Harrell et al. [39]. Let \widehat{HUM}_{init} be the initial optimistic verification-bias-adjusted HUM estimate. We first created $B = 1000$ bootstrapped datasets, and in each of these datasets repeated the multiple imputation, model selection, and model estimation procedures in order to refit models (6) and (3) and reestimate the verification weights, $\hat{\pi}_i^R$ [40]. For the b th set of bootstrapped models, we then evaluated their prediction performance (i.e., computed the verification-bias-adjusted HUM estimator) on both the corresponding bootstrapped dataset and the original EAGeR dataset: $\widehat{HUM}_{boot}^{(b)}$ and $\widehat{HUM}_{orig}^{(b)}$, respectively. The final optimism-corrected estimate was then

$$\widehat{HUM} = \widehat{HUM}_{init} - \frac{1}{B} \sum_{b=1}^B \left(\widehat{HUM}_{boot}^{(b)} - \widehat{HUM}_{orig}^{(b)} \right).$$

Remark 2 The above bootstrap procedure explicitly accounts for three sources of uncertainty and optimism in \widehat{HUM}_{init} , namely that resulting from the multiple imputation, the model selection, and the model estimation procedures. However, several of the covariate patterns in \mathbf{w}_i have low prevalence in the original and resampled EAGeR datasets, such that the final selected models were highly variable across the resampled datasets and prone to large and unstable coefficient estimates; these large fluctuations in coefficient magnitude in turn led to increased separation between the individual risk predictions and thus to improved in-sample (and worsened out-of-sample) discriminatory performance (a model selection analog of the winner's curse [41]). As a result, the resampled \widehat{HUM}_{boot} were overly optimistic relative to \widehat{HUM}_{init} , and their empirical bootstrap distribution was centered well to the right of the observed-sample statistic. In light of this idiosyncratic small-sample behavior, we opt to report bootstrapped optimism-corrected results that condition on the form of the originally-selected risk prediction model, i.e., for which only the multiple imputation and model estimation procedures were bootstrapped.

All analyses were conducted using R version 3.6.1, and the code is available online at <https://github.com/kaitlyncook/preconception-risk-prediction>. The analysis code makes use of the package `discSurv` to fit the discrete-time survival model in (6), `nnet` to fit the multinomial regression model in (3), `MASS` to conduct model selection, and `mice` to conduct sequential multiple imputation.

Results

The final first- and second-stage prediction models are given in Table 2, and the resulting preconception risk profiles, calculated for all women in the EAGeR trial with complete covariate data, are given in Fig. 3. Among these women, the predicted probability of a clinically-recognized pregnancy within the first six menstrual cycles post preconception visit ranged from 14.5% to 91.9%; the predicted probability of successful conception peaked during the first menstrual cycle post clinic visit (at 20.3%, on average) and decreased with each subsequent cycle (Fig. 4). Assuming that each woman did, in fact, successfully conceive, the most likely outcome of their pregnancies was a full-term birth (with predicted conditional probabilities ranging from 56.7% to 80.2%), followed by clinical pregnancy loss (10.1% to 37.0%) and preterm birth (3.5% to 25.7%). The optimistic verification-bias-adjusted HUM was estimated to be 0.108 (95% confidence interval (CI) with bootstrapped standard errors: 0.078, 0.144), while the optimism-corrected HUM was estimated to be

0.093 (95% CI with bootstrapped standard errors: 0.064, 0.128); both point estimates and confidence interval limits fell well above the non-informative HUM for a four-level outcome (≈ 0.042), representing significant improvement over classification by random chance.

To illustrate the potential clinical utility of this multi-stage prediction framework, we specifically highlight the predicted risk profiles for three hypothetical patients: one who represents the median participant in the EAGeR trial with respect to each of the included baseline covariates; another whose medical history, and in particular their hypertension and experience of previous pregnancy loss with no previous live birth, places them at higher risk of an adverse pregnancy outcome; and a third who is of advanced maternal age (Table 3). Comparing the corresponding risk profiles to one another illustrates a stark difference in predicted pregnancy attempt trajectories for the three hypothetical patients, which may then inform the subsequent preconception care that each patient receives (see the gray bands superimposed on Fig. 3). The first patient and their obstetrician may, for example, find their preconception risk profile to be acceptable, and proceed with the pregnancy attempt according to the standard of care. The second patient, however, has an elevated probability of failing to conceive (44.2%) and, should they conceive, of having either a clinical pregnancy loss or preterm birth (22.0%). Given this profile, the obstetrician may instead recommend more active monitoring of the pregnancy attempt and, should the patient fail to

Table 2 Selected models for the time to clinically-recognized pregnancy (first-stage model), the result of that pregnancy (second-stage model) and the probability of verification. All results are reported as: estimate (SD)

	First Stage	Second Stage		Verification
	$\text{logit}[P(T_i = t T_i \geq t)]$	$\text{log}(\pi_{i2 1} / \pi_{i4 1})$	$\text{log}(\pi_{i3 1} / \pi_{i4 1})$	$\text{logit}(\pi_i^R)$
Intercept	-0.542 (0.370)	-3.025 (0.754)	-0.827 (1.096)	2.050 (0.515)
$I(t = 2)$	-0.047 (0.117)	—	—	—
$I(t = 3)$	-0.148 (0.129)	—	—	—
$I(t = 4)$	-0.152 (0.141)	—	—	—
$I(t = 5)$	-0.456 (0.167)	—	—	—
$I(t = 6)$	-0.402 (0.178)	—	—	—
Menstrual cycles trying prior to baseline	-0.120 (0.019)	—	—	—
Aspirin use	0.185 (0.086)	—	—	—
Age (in years)	-0.026 (0.010)	0.041 (0.022)	-0.032 (0.035)	—
Number of previous pregnancy losses	0.058 (0.092)	0.265 (0.210)	-0.089 (0.330)	—
Non-Hispanic white	0.444 (0.159)	0.012 (0.422)	-0.698 (0.463)	0.773 (0.278)
College degree	0.251 (0.093)	—	—	0.474 (0.236)
BMI	-0.027 (0.007)	—	—	-0.024 (0.015)
Parous	0.357 (0.090)	—	—	0.366 (0.214)
Smoker (past year)	—	—	—	-0.973 (0.258)
Hypertension	—	0.412 (0.253)	0.741 (0.367)	—

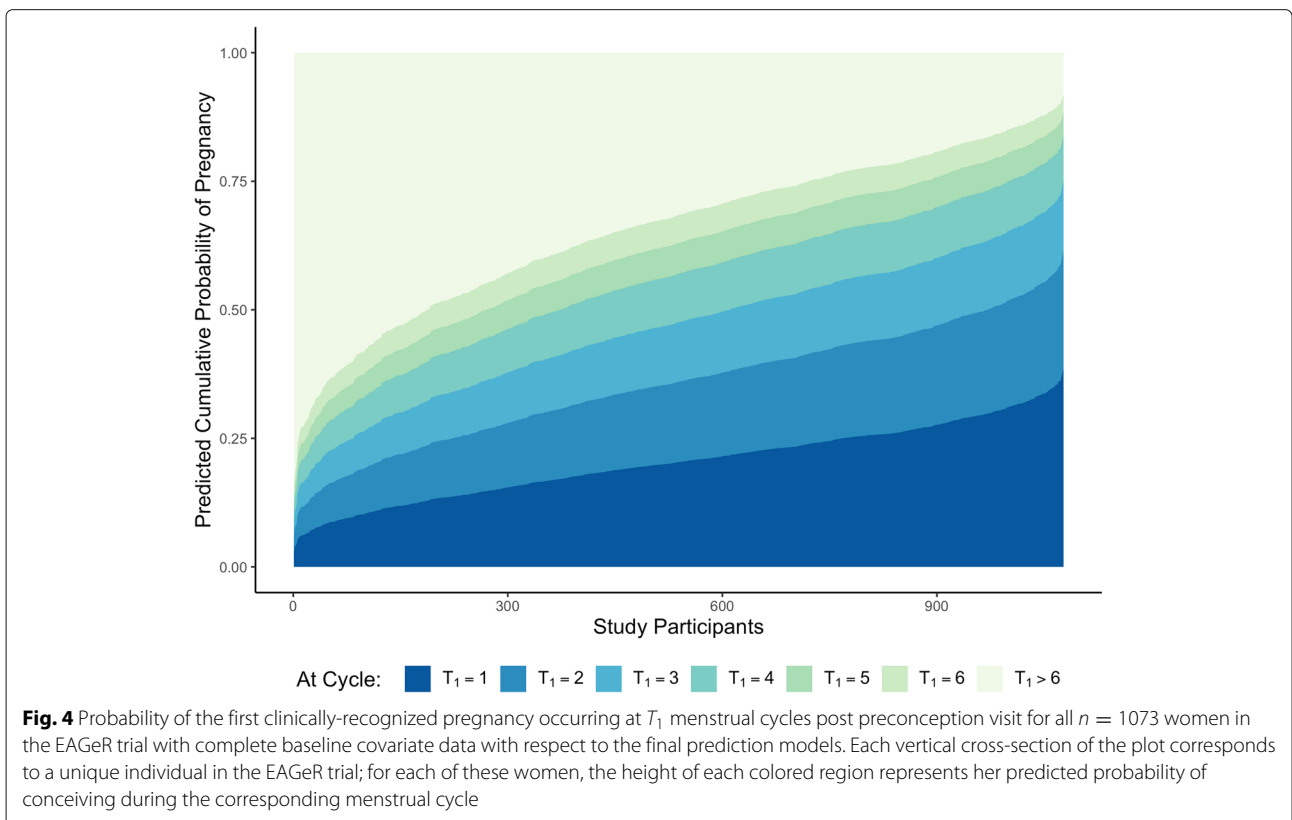
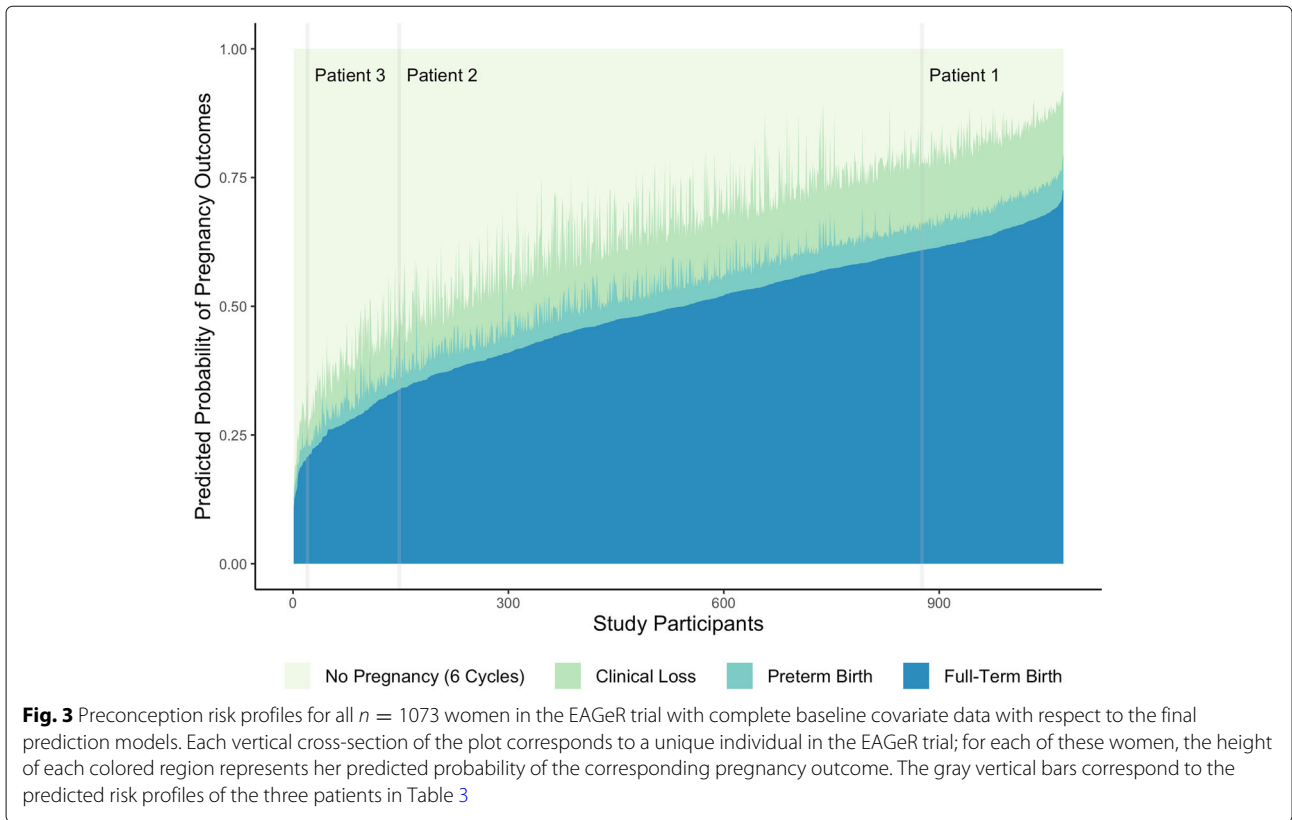


Table 3 Preconception medical history and sociodemographic profiles for three hypothetical patients, along with their corresponding predicted preconception risk profiles

	Patient 1	Patient 2	Patient 3
<i>Characteristics at the Preconception Visit</i>			
Age	28.3 years	23.7 years	39.3 years
Non-Hispanic white	Yes	No	No
College degree	None	None	None
BMI	24.4	21.1	27.2
Hypertensive status	No	Yes	Yes
Parous	Yes	No	Yes
Number of previous pregnancy losses	1.0	1.0	1.0
Preconception aspirin use	Yes	No	No
Menstrual cycles trying prior to visit	1.0	2.0	4.0
<i>Preconception Risk Profile</i>			
No pregnancy within $\tau = 6$ cycles	0.212	0.443	0.643
Pregnancy ending in clinical loss	0.125	0.085	0.099
Pregnancy ending in a preterm birth	0.050	0.134	0.050
Pregnancy ending in a full-term birth	0.613	0.338	0.208

conceive within six additional menstrual cycles, a fertility treatment such as intrauterine insemination. Finally, the third patient's advanced maternal age and high risk of either failing to conceive (64.3%) or experiencing a clinical pregnancy loss (9.9%) might lead the obstetrician to recommend immediate use of assisted reproductive technology.

Discussion

In this manuscript, we have advocated for the use of a multistate modeling framework in order to construct and validate individualized pregnancy outcome profiles for use in routine preconception care. We have discussed how the task of pregnancy outcome prediction in particular presents challenges to the typical construction and application of these multistate models—namely regarding correct specification of the transition intensities and potential coarsening of the available data on both the outcome and censoring processes—and presented a two-stage estimation process that circumvents these challenges in order to derive patient-specific absolute risk profiles for the joint probability of all competing birth outcomes. We also developed an inverse-probability-weighted HUM statistic to quantify the model's classification performance.

By adopting this multistate prediction framework, we were able to reframe—and broaden the purview of—pregnancy risk prediction to include endpoints related to

both the conception and gestation processes. Existing prediction tools have typically focused on a single adverse pregnancy outcome, whose risk is determined on the basis of biomedical markers collected over the course of the pregnancy; they are necessarily calculated downstream of the initial preconception consultation. This initial visit represents, however, an important window in patients' preconception and prenatal care. It marks the first time at which risk factors for adverse conception and perinatal outcomes may be identified and modified, with potentially long-ranging implications for the pregnancy attempt. To that end, the multistate competing risks framework allows one to derive patient-specific risk profiles—estimable at the time of the preconception visit—that quantify both (i) the likelihood of successful conception and (ii) the likelihood of that pregnancy ending in a clinical pregnancy loss, preterm birth, or full-term birth while permitting different sets of demographic and medical characteristics to govern and inform each of those processes.

We illustrated the clinical utility of these risk profiles by constructing and validating a multistage competing risks model using the EAGeR trial. While the patient population in the EAGeR trial—women with at least one previous pregnancy loss and at most two previous live births—prevents generalization of this prediction model to a broader clinical population, it still represents an important proof of concept. Despite using only the (more limited) information available at the time of the initial preconception visit, the resulting risk profiles were able to meaningfully distinguish between women on the basis of their future pregnancy outcomes. This prediction framework also easily generalizes to other clinical settings in which the endpoint of interest is effectively a multistage outcome. For example, predicting which patients are at high-risk of hospital-acquired infections and subsequent adverse events—which considers first whether a patient acquires an infection and, among those with a hospital-acquired infection, the duration of stay or severity of that infection—readily lends itself to a similar multistage framing.

Before these sorts of multistate competing risks prediction models are implemented in clinical practice, however, several limitations regarding their construction and validation must be addressed. Returning to the task of pregnancy outcome prediction, some fraction of the broader population is understood to be *sterile* and thus unable to conceive or to experience any of the downstream pregnancy outcomes; this phenomenon is naturally modeled through the inclusion of either a latent sterility variable or a so-called cure fraction (here representing those who are sterile) into the competing risks model. These extensions are not without their own methodological challenges. An individual's latent sterility status is measured imperfectly through their observed *fertility*, which itself is measured

only after a fixed period of time—an individual is classified as clinically infertile only after unsuccessfully attempting to conceive for 12 or more menstrual cycles—and so will be structurally missing for most individuals in the estimation and validation datasets. Once measured, sterility is also subject to differential misclassification: while all individuals who are unable to conceive will be labeled as clinically infertile, not all individuals labeled as clinically infertile are biologically unable to conceive. Generalizing the multistate pregnancy outcome prediction framework to include infertility will thus require addressing both of these statistical features. Additional work will also be needed to improve model selection procedures for multistate competing risks models, e.g., through the development of penalized regression methods, particularly when these models are fit using either inverse probability of censoring weights or multiple imputation.

Finally, the proposed HUM validation metric for multinomial outcomes is less well-studied and well-known than its binary counterpart, the AUC, and so lacks the same common understanding of what a “poor”, “good”, or “excellent” HUM might be for a given applied problem. Constructing confidence intervals for the estimated HUM is an important first step towards contextualizing the predictive performance of multinomial or nested multinomial (multistate) models, but further research on both the statistical properties and the practical implications of the HUM is warranted.

Conclusion

As illustrated by our analysis of the EAGeR trial data, our proposed multistate competing risks framework has the potential to shape preconception care by (i) considering joint risk prediction for a broader array of possible pregnancy attempt outcomes and (ii) providing those predictions at an earlier intervention period (during the preconception clinic visit) than existing obstetric prediction models. It thus presents an important first step in expanding the types of clinical risk assessment tools available for pregnancy outcome prediction, and suggests broader applications to the prediction of other complex, multistage health outcomes.

Abbreviations

EAGeR: Effects of Aspirin in Gestation and Reproduction; HUM: Hypervolume under the manifold; AUC: Area under the curve; BMI: Body mass index; CI: Confidence interval

Acknowledgements

The authors are grateful to all individuals who participated in the EAGeR trial, as well as to Kyu Ha Lee and Catherine Lee for providing source code for the HUM implementation. They would also like to thank two anonymous reviewers for their helpful comments and feedback on an earlier version of this manuscript.

Funding

This work was funded in part by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development,

National Institutes of Health, Bethesda, MD. The funder had no role in the study design, methods development, analyses, or writing of the manuscript.

Authors' contributions

K.C. and S.H. led the development of the statistical methods and prediction framework, as well as their implementation in software. N.J.P. and E.S. curated the analytic dataset, directed the analysis, and interpreted the findings. K.C. drafted the main manuscript text and figures, and all authors reviewed and approved the final version.

Availability of data and materials

The data analyzed in this study are available upon request at the NICHD/DIPHR Biospecimen Repository Access and Data Sharing (BRADS) site (<https://brads.nichd.nih.gov>).

Declarations

Ethics approval and consent to participate

For the EAGeR study, individual sites obtained Institutional Review Board approvals at each clinical center and the DCC: Salt Lake City, Utah Study Site, IRB #1002521 (5/24/2007); Scranton, Pennsylvania Study Site, IRB #HHSN275200403394 (05/2007); Buffalo, New York Study Site, IRB #SPM0900107A (12/20/2007); Denver, Colorado Study Site, IRB #08-0982 (04/2009). Participants provided written informed consent. The trial was registered on clinicaltrials.gov, #NCT00467363. An independent Data Safety and Monitoring Board (DSMB) ensured continued patient safety and ongoing monitoring of the trial, and all methods were carried out in accordance with the relevant guidelines and regulations. All data were de-identified.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Author details

¹Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, US. ²Epidemiology Branch, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD, US. ³Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, US. ⁴Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, US.

Received: 2 August 2021 Accepted: 10 March 2022

Published online: 30 May 2022

References

- Posner SF, Johnson K, Parker C, Atrash H, Biermann J. The national summit on preconception care: a summary of concepts and recommendations. *Matern Child Health J*. 2006;10(1):199–207.
- Farquhar CM, Liu E, Armstrong S, Arroll N, Lensen S, Brown J. Intrauterine insemination with ovarian stimulation versus expectant management for unexplained infertility (tui): a pragmatic, open-label, randomised, controlled, two-centre trial. *Lancet*. 2018;391(10119):441–50.
- Cnattingius S, Signorello LB, Annerén G, Clausson B, Ekbohm A, Ljunger E, Blot WJ, McLaughlin JK, Petersson G, Rane A, et al. Caffeine intake and the risk of first-trimester spontaneous abortion. *N Engl J Med*. 2000;343(25):1839–45.
- Metwally M, Ong KJ, Ledger WL, Li TC. Does high body mass index increase the risk of miscarriage after spontaneous and assisted conception? a meta-analysis of the evidence. *Fertil Steril*. 2008;90(3):714–26.
- Kleinrouweler CE, Cheong-See FM, Collins GS, Kwee A, Thangaratnam S, Khan KS, Mol BWJ, Pajkrt E, Moons KG, Schuit E. Prognostic models in obstetrics: available, but far from applicable. *Am J Obstet Gynecol*. 2016;214(1):79–90.
- Meertens LJ, van Montfort P, Scheepers HC, van Kuijk SM, Aardenburg R, Langenveld J, van Dooren IM, Zwaan IM, Spaanderma ME, Smits LJ. Prediction models for the risk of spontaneous preterm birth based on

- maternal characteristics: a systematic review and independent external validation. *Acta Obstet Gynecol Scand.* 2018;97(8):907–20.
7. Heestermans T, Payne B, Kayode GA, Amoakoh-Coleman M, Schuit E, Rijken MJ, Klipstein-Grobusch K, Bloemenkamp K, Grobbee DE, Browne JL. Prognostic models for adverse pregnancy outcomes in low-income and middle-income countries: a systematic review. *BMJ Global Health.* 2019;4(5):001759.
 8. Goyal NK, Hall ES, Greenberg JM, Kelly EA. Risk prediction for adverse pregnancy outcomes in a medicaid population. *J Womens Health.* 2015;24(8):681–8.
 9. Chu R, Chen W, Song G, Yao S, Xie L, Song L, Zhang Y, Chen L, Zhang X, Ma Y, et al. Predicting the risk of adverse events in pregnant women with congenital heart disease. *Journal of the American Heart Association.* 2020;9(14):016371.
 10. Schaaf JM, Ravelli AC, Mol BWJ, Abu-Hanna A. Development of a prognostic model for predicting spontaneous singleton preterm birth. *Eur J Obstet Gynecol Reprod Biol.* 2012;164(2):150–5.
 11. Koivu A, Sairanen M. Predicting risk of stillbirth and preterm pregnancies with machine learning. *Health Inf Sci Syst.* 2020;8(1):1–12.
 12. Slama R, Ballester F, Casas M, Cordier S, Eggesbø M, Iniguez C, Nieuwenhuijsen M, Philippat C, Rey S, Vandentorren S, et al. Epidemiologic tools to study the influence of environmental factors on fecundity and pregnancy-related outcomes. *Epidemiol Rev.* 2014;36(1):148–64.
 13. Townsend R, Manji A, Allotey J, Heazell A, Jorgensen L, Magee L, Mol B, Snell K, Riley R, Sandall J, et al. Can risk prediction models help us individualise stillbirth prevention? a systematic review and critical appraisal of published risk models. *BJOG Int J Obstet Gynaecol.* 2021;128(2):214–24.
 14. Sep SJ, Smits LJ, Prins MH, Spaanderman ME, Peeters LL. Simple prepregnant prediction rule for recurrent early-onset hypertensive disease in pregnancy. *Reprod Sci.* 2009;16(1):80–7.
 15. van Kuijk SM, Nijdam M-E, Janssen KJ, Sep SJ, Peeters LL, Delahajie DH, Spaanderman M, Bruinse HW, Franx A, Bots ML, et al. A model for preconceptional prediction of recurrent early-onset preeclampsia: derivation and internal validation. *Reprod Sci.* 2011;18(11):1154–9.
 16. Mehta-Lee SS, Palma A, Bernstein PS, Lounsbury D, Schlecht NF. A preconception nomogram to predict preterm delivery. *Matern Child Health J.* 2017;21(1):118–27.
 17. Scheike TH, Jensen TK. A discrete survival model with random effects: an application to time to pregnancy. *Biometrics.* 1997;53(1):318–29.
 18. Keiding N. Time to pregnancy. In: Armitage P, Colton T, editors. *Encyclopedia of Biostatistics*, vol. 4, 2nd ed. West Sussex: Wiley; 2005.
 19. Zhou H. Statistical models for human fecundability. *Stat Methods Med Res.* 2006;15(2):181–94.
 20. Sundaram R, McLain AC, Buck Louis GM. A survival analysis approach to modeling human fecundity. *Biostatistics.* 2012;13(1):4–17.
 21. van Eekelen R, Putter H, McLernon DJ, Eijkemans MJ, van Geloven N. A comparison of the beta-geometric model with landmarking for dynamic prediction of time to pregnancy. *Biom J.* 2020;62(1):175–90.
 22. Beyersmann J, Allignol A, Schumacher M. *Competing Risks and Multistate Models with R*. New York: Springer; 2011.
 23. Schisterman EF, Silver RM, Perkins NJ, Mumford SL, Whitcomb BW, Stanford JB, Leshner LL, Faraggi D, Wactawski-Wende J, Browne RW, et al. A randomised trial to evaluate the effects of low-dose aspirin in gestation and reproduction: design and baseline characteristics. *Paediatr Perinat Epidemiol.* 2013;27(6):598–609.
 24. Scheike TH, Petersen JH, Martinussen T. Retrospective ascertainment of recurrent events: an application to time to pregnancy. *J Am Stat Assoc.* 1999;94(447):713–25.
 25. Andersen PK, Perme MP. Inference for outcome probabilities in multi-state models. *Lifetime Data Anal.* 2008;14(4):405.
 26. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*: Wiley; 2011.
 27. Andersen PK, Klein JP. Regression analysis for multistate models based on a pseudo-value approach, with applications to bone marrow transplantation studies. *Scand J Stat.* 2007;34(1):3–16.
 28. Scheike TH, Zhang M-j. Direct modelling of regression effects for transition probabilities in multistate models. *Scand J Stat.* 2007;34(1):17–32.
 29. Scheike TH, Zhang M-J, Gerds TA. Predicting cumulative incidence probability by direct binomial regression. *Biometrika.* 2008;95(1):205–20.
 30. Haneuse S, Daniels M. A general framework for considering selection bias in EHR-based studies: what data are observed and why? *eGEMs.* 2016;4(1):1203.
 31. Allison PD. Discrete-time methods for the analysis of event histories. *Sociological Methodology.* 1982;13:61–98.
 32. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med.* 2011;30(4):377–99.
 33. Seaman SR, White IR, Copas AJ, Li L. Combining multiple imputation and inverse-probability weighting. *Biometrics.* 2012;68(1):129–37.
 34. Quartagno M, Carpenter J, Goldstein H. Multiple imputation with survey weights: a multilevel approach. *J Surv Stat Methodol.* 2020;8(5):965–89.
 35. Liu J, Gelman A, Hill J, Su Y-S, Kropko J. On the stationary distribution of iterative imputations. *Biometrika.* 2014;101(1):155–73.
 36. Zhang Y, Alonzo TA, Initiative ADN. Inverse probability weighting estimation of the volume under the roc surface in the presence of verification bias. *Biom J.* 2016;58(6):1338–56.
 37. Schisterman EF, Silver RM, Leshner LL, Faraggi D, Wactawski-Wende J, Townsend JM, Lynch AM, Perkins NJ, Mumford SL, Galai N. Preconception low-dose aspirin and pregnancy outcomes: results from the eager randomised trial. *Lancet.* 2014;384(9937):29–36.
 38. Wood AM, White IR, Royston P. How should variable selection be performed with multiply imputed data?. *Stat Med.* 2008;27(17):3227–46.
 39. Harrell Jr 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.
 40. Schomaker M, Heumann C. Bootstrap inference when using multiple imputation. *Stat Med.* 2018;37(14):2252–66.
 41. Hansen PR. A winner's curse for econometric models: on the joint distribution of in-sample fit and out-of-sample fit and its implications for model selection. *Res Pap.* 2010;1:39.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

