OXFORD UNIVERSITY PRESS | JAMIA: Journal of the American Medical Informatics Associatio

Informative presence and observation in routine health data: A review of methodology for clinical risk prediction

Journal:	*Journal of the American Medical Informatics Association*
Manuscript ID	amiajnl-2020-009489.R2
Article Type:	Review
Keywords:	clinical prediction model, informative observation, informative presence, electronic health records

Informative presence and observation in routine health data:

A review of methodology for clinical risk prediction

Corresponding Author: Rose Sisk

Address: Room G.304

Jean McFarlane Building

University of Manchester

M13 9PY

Email: rose.sisk@postgrad.manchester.ac.uk

Telephone: +447903109945

Rose Sisk[1], Lijing Lin[1], Matthew Sperrin[1], Jessica K. Barrett[2,3], Brian Tom[2], Karla Diaz-Ordaz[4],

Niels Peek[1,5,6], Glen P. Martin[1]

1. Division of Informatics, Imaging and Data Sciences, School of Health Sciences, The University of Manchester, Manchester, UK.
2. MRC Biostatistics Unit, University of Cambridge, Cambridge, UK
3. Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
4. London School of Hygiene and Tropical Medicine, London, UK
5. NIHR Biomedical Research Centre, The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
6. The Alan Turing Institute, London, UK.

Keywords: clinical prediction model, electronic health records, informative observation, informative presence

Word Count: 4800

Abstract

OBJECTIVES

Informative presence (IP) is the phenomenon whereby the presence/absence of patient data is potentially informative with respect to their health condition, with informative observation (IO) being the longitudinal equivalent. These phenomena predominantly exist within routinely collected healthcare data, where data collection is driven by the clinical requirements of patients and clinicians. The extent to which IP and IO are considered when using such data to develop clinical prediction models (CPMs) is unknown, as is the existing methodology aiming at handling these issues. This review aims to synthesise such existing methodology, thereby helping identify an agenda for future methodological work.

MATERIALS & METHODS

A systematic literature search was conducted by two independent reviewers using pre-specified keywords.

RESULTS

Thirty-six papers were included. We categorised the methods presented within as: derived predictors (including some representation of the measurement process as a predictor in the model); modelling under IP; and latent structures. Including missing indicators/summary measures as predictors is the most commonly presented approach amongst the included studies (24/36 papers).

DISCUSSION

This is the first review to collate the literature in this area under a prediction framework. A considerable body relevant of literature exists, and we present ways in which the described methods

could be developed further. Guidance is required for specifying the conditions under which each method should be used to enable applied prediction modellers to use these methods.

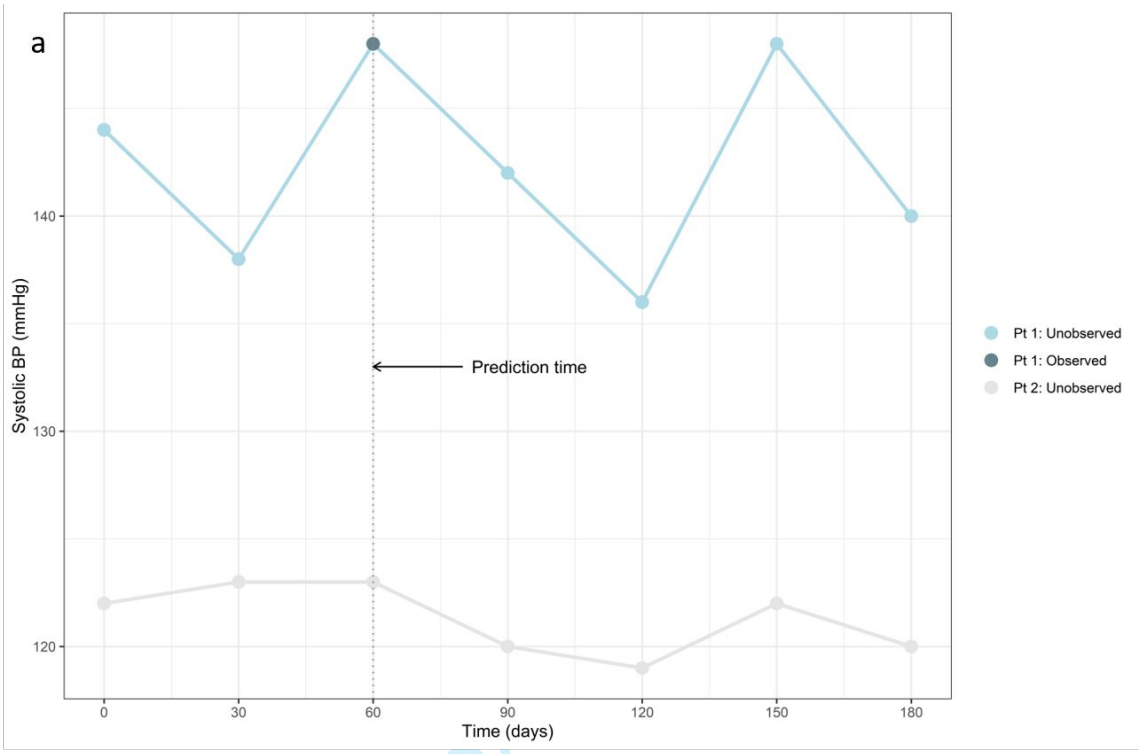
CONCLUSION

A growing recognition of IP and IO exists within the literature, and methodology is increasingly becoming available to leverage these phenomena for prediction purposes. IP and IO should be approached differently in a prediction context than when the primary goal is explanation. The work included in this review has demonstrated theoretical and empirical benefits of incorporating IP/IO, and therefore we recommend that applied health researchers consider incorporating these methods in their work.

BACKGROUND & SIGNIFICANCE

Clinical prediction models (CPMs) estimate the risk that a patient currently has (diagnostic), or will develop (prognostic), an outcome of interest based on known clinical and patient measures. Such risk models can guide clinical decision-making, amongst other uses.

Widespread adoption of electronic health records (EHRs) facilitates the development of CPMs,[1] since detailed clinical and patient information is collected through routine healthcare contacts. Such rich longitudinal information provides long-term patient follow-up without the need to recruit patients and conduct regular follow-up visits. The analysis of routinely collected data is not, however, without challenge. Observation times are not pre-specified as they would be in a typical research study (e.g. in a prospective cohort study with scheduled follow-up visits). Instead, data are collected opportunistically, where patient/clinician decisions directly dictate whether we observe clinical biomarkers and patient information.[2] For example, GP visits occur more frequently during periods of ill health,[3] and only information relevant to the particular consultation will be recorded. Equally, during inpatient care, clinicians will adapt their monitoring frequency to the changing needs and condition of the individual patient (see Figure 1).



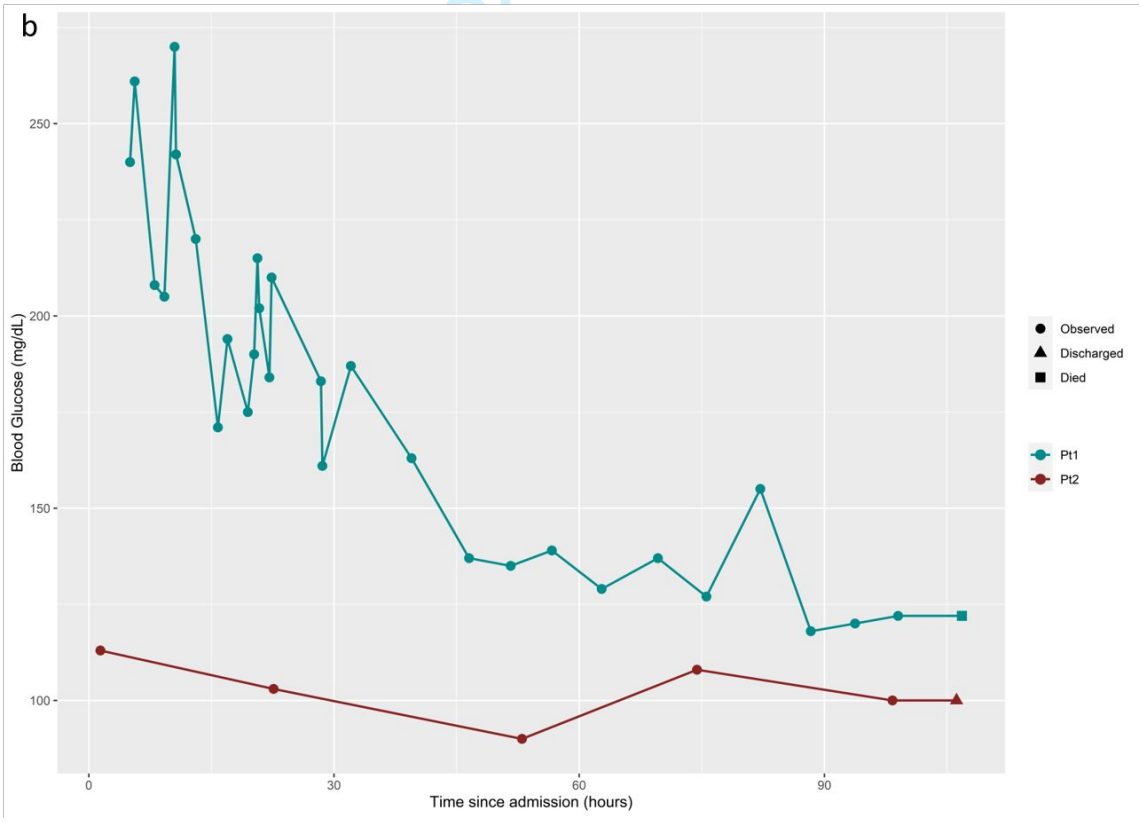


Figure 1a: An illustration of Informative Presence and how this could impact the information available at prediction time. We see the longitudinal pattern of blood pressure for two patients, with both their observed and unobserved values shown. Patient 1 has one single observed value of systolic BP, and this happens when their blood pressure was at its highest. Patient 2 has no observed values, but their blood pressure remains in the normal range - either the patient or clinician saw no clinical need to take a blood pressure measurement at any time.

Figure 1b: An illustration of Informative Observation, taken from the MIMIC dataset[4]. *Patient 1 has many more in-hospital measurements of blood glucose than Patient 2 throughout their ICU admission, likely due to the fact that their Blood Glucose is much higher and much more variable than Patient 2. A more severe condition often means more intense monitoring.*

We refer to the process by which visits, and hence measurements, occur as the *observation process* (also known elsewhere as the visiting or monitoring process). We define two key properties that an observation process may have, when presence of data is informative:

1. Informative presence (IP) (Figure 1a) – the presence or absence of a patient's data at any given time point carries information about their health status.
2. Informative observation (IO) – the timing, frequency or intensity (rate) of a patient's longitudinal pattern of observation carries information about their evolving health state. See Figure 1b for an example.

Informative presence is challenging from a statistical perspective as it implies a missing not at random (MNAR) process. IP is, however, conceptually different from missingness, as in the former, there was never any intention of collecting the data at a particular visit. Informative presence has previously been defined elsewhere[5,6], with Phelan et al[5] discussing how interactions contained within electronic health records are informative with respect to patient health.

Informative observation is the continuous time generalisation of informative presence: a longitudinal Visiting (at time t) Not at Random (VNAR) process, defined as "given data recorded up to time t, visiting at time t is not independent of outcome at time t".[7] By generalising the definition of informative presence above, one can draw value from how frequently a patient is observed over time. This is especially true when no schedule exists dictating when or how often visits should occur; we therefore focus on what an individual's longitudinal observation process could tell us about their condition.

A recent review of CPMs developed using routinely collected data revealed an apparent lack of understanding of, or proper handling of, IP/IO[1]. Moreover, much of the existing methodological literature in this area has focussed on IP/IO only in the context of effect estimation (i.e. in causal or associational studies),[8–14] and has generally viewed it as a "nuisance" – i.e. a phenomenon that potentially biases effect estimators and therefore needs to be corrected for in the analysis.

However, when developing a CPM, the primary focus is on achieving good predictive performance; predictor effect estimation is less important.

Instead, one could view IP and IO as opportunities to draw information from the EHR that is not explicitly recorded. In this paper we focus on informative measurement patterns in the predictors, and we do not discuss presence or absence of outcome data. Agniel et al.[15] demonstrated how the timing of a lab test better predicts mortality than the actual result of the test. Others have illustrated how incorporating the presence or absence of a particular test for an individual into a CPM can improve its accuracy.[16–18]

OBJECTIVES

This article aims to review the literature on methodology allowing CPMs to utilise IP or IO, both in overcoming some of the aforementioned challenges, and in harnessing information within informative measurement patterns. In doing so, we also highlight outstanding areas of methodological work that should be prioritised. Finally, we summarise existing software packages capable of implementing the methodology.

MATERIALS & METHODS

The strategy employed in this review loosely follows a scoping review framework.[19] Our protocol has been registered on the Open Science Framework.[20]

Search strategy

We searched MEDLINE, Embase and Web of Science for relevant articles using pre-specified search terms. Further details of the full search strategy (including search terms and an additional snowballing stage) can be found in the Supplementary materials and the published protocol[20].

Study selection

We had the following inclusion criteria: any paper presenting a method that allows CPMs to incorporate IP or IO. We excluded: papers that applied existing methods that had already been published elsewhere, and included those earlier publications instead, non-medical areas of application, IP/IO in outcome measures, and methods that handle sample selection bias, imputation or censoring only. See the Supplementary material for further justification of these exclusions.

We do not include textbooks within the review; while this could mean we miss some relevant literature, searching within textbooks is not widely feasible. Additionally, we believe that most methodological development in this area will be published in original research articles rather than textbooks.

Two independent reviewers (RS & LL) conducted a two-stage screening process. Titles and abstracts were screened first, and full texts of remaining articles were reviewed at the second stage. Reviewers met regularly to track agreement. Systematic differences were translated into new inclusion/exclusion criteria, in consultation with a third reviewer (GPM).

Primarily, we extracted information regarding the modelling method employed and any reported advantages and disadvantages. We also extracted information on the form of the observation processes, predictors, and outcome, including any clinical use cases presented.

RESULTS

Our database searches identified 6127 studies, of which 111 were retained for full text screening. Eleven of these were deemed eligible for inclusion. We identified a further 25 papers through forward and backward citation searching, giving a final set of 36 included papers (Figure 2).

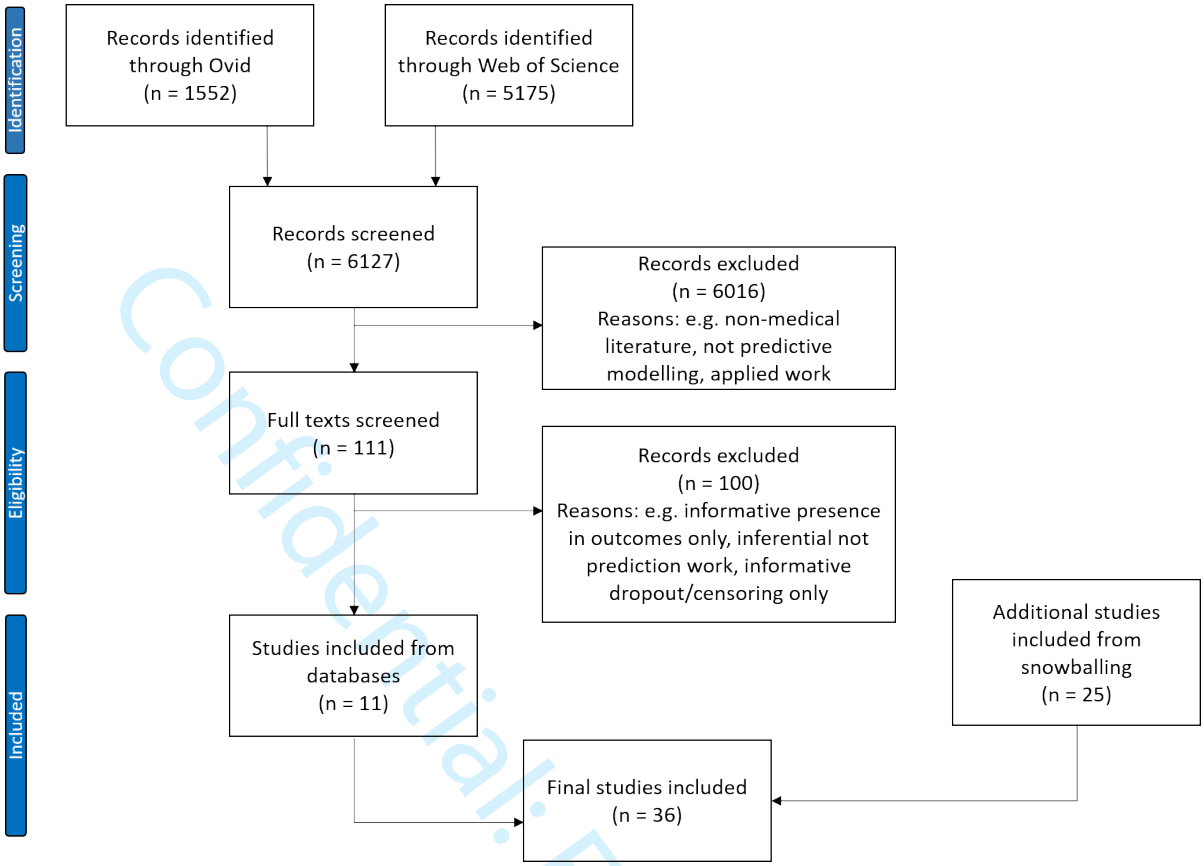


Figure 2: PRISMA flow diagram showing the various screening stages and reasons for exclusion at each stage

Throughout this section, we will illustrate each method with the following notation. Consider a binary outcome $Y(t)$ (or Y if only observed once) for patients $i = 1,\ldots,n$, at time t, where $Y = 1$ denotes that the event occurred, with marginal probability $P[Y = 1]$. Define a potentially time-varying continuous covariate process $X(t)$, with potential realisations x_{ij} for $i = 1,\ldots, n$ and $j = 1,\ldots m_i$, or simply x_i if X is not time-varying. The timing of the jth realisation of $X(t)$ is $t_{ij} \in \mathbb{R}^+$. Denote $R = 1$ if $X(t)$ is ever observed, and $R = 0$ if not. Define r_{ij} = 1 if the covariate process is observed at time t_{ij}. We assume that Z is a completely observed time-invariant covariate. $g(.)$ represents a link function, e.g. the logit function.

Broadly, the methods in this paper cover the three scenarios described in Table 1. To illustrate the prediction scenarios and methods, we consider a simplified version of the Sequential Organ Failure Assessment (SOFA) score,[21] used to predict mortality in critical care, assuming that the only predictors in the model are bilirubin and blood pressure. Of these two predictors, we assume that

blood pressure is completely observed for all patients, and bilirubin is informatively observed, as it has been shown to be within critical care. Depending on the specific scenario, it may be a one-time point observation, or a longitudinal process.[17]

There exists a breadth of methodological literature covering Scenario S2 (without accounting for IP/IO), which has recently been synthesised by Bull et al.[22] We therefore focus on modelling strategies that have specifically been proposed or extended to accommodate IP or IO.

Scenario	Scenario name	Description	Example (SOFA)
S1	Cross-sectional prediction	Interest lies in obtaining a single prognostic estimate (prediction) using a single value for each predictor.	Use values of bilirubin and Blood pressure (BP) obtained upon ICU admission to predict in-hospital survival (binary).
S2	Cross-sectional prediction with longitudinal predictor measurements	Interest lies in obtaining a single prognostic estimate but using the longitudinal history of predictor values.	Use all repeated lab tests obtained throughout inpatient admission for bilirubin and BP to predict in-hospital survival.
S3	Longitudinal prediction with longitudinal predictors and outcomes	Interested in prognostic estimates at multiple time points, potentially using the longitudinal history of predictor values.	Use all repeated measures of BP and bilirubin obtained throughout inpatient and ICU admission to predict survival at multiple future time points.

Table 1: A description of different prediction scenarios, covering cross-sectional vs longitudinal predictors and outcomes.

Identified Approaches to Handle Informative Presence and Observation

We identified three broad categories of method based on the included papers: (i) methods that incorporate IP/IO through derived predictors; (ii) methods for modelling under informative presence; and (iii) methods that incorporate IP/IO using latent structures. Within these three categories, we identified eight modelling strategies. A summary of the methods can be found in Table 2. Table 3 summarises the advantages, disadvantages, software, and assumptions for each method – here, the reported advantages and disadvantages were inferred by the research team since they are not consistently mentioned in the included literature. A summary table at paper-level can be found in Appendix 3.

Modelling approach	Broad category	Refs	Scenario(s)	IP or IO	Description	Example
Missing indicators & Separate class	Derived predictors	[16,23–30]	S1	IP	Creating a binary indicator, representing presence/absence of a predictor at a given time point or in a given window	Create a binary indicator taking 0 when bilirubin is observed, and 1 if missing. Enter this as an additional predictor alongside observed bilirubin and BP.
Summary measures	Derived predictors	[15,24,31–44]	S2	IO	Summarising the observation process into a single variable, e.g. counting visits, rates of visits over a window, weighted counts	Count the number of times bilirubin has been measured over the first 24 hours of each ICU admission. Enter this count as an additional predictor in the model.
Pattern-specific models	Modelling under informed presence	[45,46]	S1	IP	Derive separate models for each missingness pattern	Develop models for: bilirubin and BP observed, and only BP observed
Likelihood-based methods	Modelling under informed presence	[47,48]	S1	IP	Incorporating missingness mechanism into maximum-likelihood estimation of parameter estimates	Bilirubin is missing not at random. Estimate model parameters using method-of-weights and EM algorithm.
Similarity measures	Derived predictors	[49]	S2	IO	Calculate similarity between target patient and all others, based on predictor values and measurement timings. Develop models separately for "similar" groups of patients.	Develop separate models amongst cohorts of patients with similar bilirubin, BP and timings of those measures.
Latent variable	Latent structures	[50,51]	S1, S3	IP	Outcome can be partially latent, and the observation process infers the latent state.	The occurrence of a bilirubin measurement is used to infer patient state in a hierarchical model.
HMMs	Latent structures	[52,53]	S3	IO	Outcome is a partially latent process, and the observation process infers the state at any time.	The intensity of bilirubin measurements over the course of a patient's admission infers their severity at any time point.
Joint modelling/ shared random effects	Latent structures	[54–56]	S2, S3	IP and IO	Model the outcome, predictor and observation processes separately, but join them through random effects shared across the models.	Model the number of times bilirubin is measured throughout the admission as a point process, the repeated measures of bilirubin using a linear mixed model, and the binary outcome using a logistic regression. Link these via at least one shared random effect across the models.

Table 2: Descriptive summary table of methods, detailing when each method may be appropriate and how it would work with the running example of a simplified SOFA score.

Category 1: Derived predictors

The methods described in this section address IP or IO by deriving some representation of the observation process and including this as a separate predictor in the model to exploit the informativeness for predictive value. These approaches tend to be straightforward and have been proposed to handle both IP and IO. However, attention must be paid to the intended use of the final model, particularly where the model will be applied in clinical settings different to the one in which it was developed. Where measurement protocols change across different settings, these models may lack generalisability when transported to a new setting.[57–59] This should not be a concern where the development and application settings remain the same.

Missing indicators/separate class

The missing indicator approach[16,23–30] handles IP in a straightforward manner, by deriving a binary variable that indicates whether a predictor has been observed at a specific time (IP) or over a defined window of time. The indicators enter the prediction model as a separate predictor alongside other patient and clinical information. For example, if a prediction model requires an entry for bilirubin but this test has not been conducted, then a missing indicator would be included as a predictor with value 1 (or 0 when observed). For categorical variables, a separate "missing" category could instead be created.

Since most prediction models require a value for every predictor, the missing indicator approach is usually combined with imputation at both model development and prediction time (not necessary for categorical predictors with a separate class).The missing indicator approach results in a model of the form:

$$g(P[Y = 1|X, Z]) = \beta_0 + \beta_1 X + \beta_2 Z + \gamma R \qquad (1)$$

for continuous predictors within cross-sectional prediction (S1).

Similarly, for a categorical predictor x_i with k categories, then the missing indicator approach would set $x_i \in \{ Cat_1, \ldots, Cat_k, \text{Missing}\}$ and our model would be

$$g(P[Y = 1|X, Z]) = \beta_0 + \beta_1 X + \beta_2 Z \qquad (2)$$

The above two equations could be combined to consider prediction models with both continuous and categorical predictors. Alternatively, missing indicators and separate classes have been well developed in tree-based prediction algorithms[28–30].

Including a missing indicator or separate class is straightforward and has demonstrated improved predictive performance over models omitting them[17]. However, their inclusion could double the number of candidate predictors for a model. The approach also fails to capture complex representations of the measurement process.

Summary measures

An extension to missing indicators, capable of incorporating both IP and IO, is to derive a summary of the measurement process and include this as a predictor.[15,24,31–44] Examples include a count of the number of measurements (of e.g. throughout a critical care admission),[37] weighted counts,[42] combined missing indicators,[31] missingness rates over time,[32] time intervals between measures,[33–35], embedding vectors that represent missing values,[36] or information relating to hospital processes.[38,39]

In some cases, combined missing indicators and time intervals also alter the relationship between a predictor and outcome. Che's[24] method stipulates that the longer a measure has been missing, the less influence it should have on an individual's prediction, therefore the last observed measurement is decayed towards a mean value.

Piecewise-Constant Intensity Models (PCIMs) have also been proposed to handle informatively observed predictors.[40,41] PCIMs use decision trees to assign an intensity rate to the observation process, conditional on its history (timings, values and events).

Define a summary measure of the observation process Q, e.g. a count of the number of times $X(t)$ (whether continuous or categorical) has been observed: $Q = m_i$. For cross-sectional prediction with a time-varying covariate, we then have:

$$g(P[Y = 1|X, Z) = \beta_0 + \beta_1 X + \beta_2 Z + \gamma Q \qquad (3)$$

where X is a summary of $X(t)$ deemed to have predictive value, e.g. the mean, most recent or most extreme value. If $X(t)$ has never been observed, this should be imputed. Like missing indicators, summary measures are easily derived and implemented in any prediction model using standard software (since they are included as standard predictors). Combining missing indicators into one summary, or implementing a dimension-reduction technique such as Lasso, also overcomes the issue of including multiple missing indicators. However, selecting the most appropriate summary measure for a model requires careful consideration, and will depend on the clinical application. No current guidance exists on how best to choose the most appropriate summary measure. The association between a chosen summary measure and the outcome might lack generalisability where measurement processes vary across locations.[23,39] Simple summary measures such as counts may also fail to capture the complex relationship between the observation process and outcome.

Category 2: Modelling under informative presence

While the methods in the other categories can be used to handle both informative presence and informative observation, this category comprises methods that have specifically been proposed to handle informative presence.

Pattern-specific models

The pattern-specific approach[45,46] derives separate models for each missingness pattern, generalising the missing indicator approach. The model corresponding to the observed pattern in a new individual is then used for prediction. For example, in a model with a single partially-observed time-invariant continuous predictor, X we would derive the following submodels:

$$g(P[Y = 1 \mid R = 1, X, Z]) = \beta_{0,1} + \beta_{1,1}X + \beta_{2,1}Z \qquad (4)$$

$$g(P[Y = 1 \mid R = 0, Z]) = \beta_{0,2} + \beta_{2,2}Z \qquad (5)$$

Where Z is completely observed. Note that formulas 4 and 5 can also be combined by including interaction terms with the missing indicator, illustrating how this approach extends the missing indicator method.

Similar submodels could be derived for categorical and continuous predictors. Saar-Tschansky & Provost[45] propose using all available data to train each submodel, whereas Fletcher Mercaldo & Blume[46] recommend that only individuals in each observed pattern be used in the derivation of that pattern's submodel (also illustrated by Janssen et al.[60]). The latter approach does not require knowledge of the missingness mechanism.

The pattern-specific approach is flexible, as it can be applied to any form of prediction algorithm. However, a practical limitation is that the number of candidate submodels becomes intractable as the number of predictors increases.

Likelihood-based methods

A different approach assumes that missingness in the predictors is non-ignorable, and incorporates this into parameter estimates via likelihood-based methods[47,48]. The model formulation would take, e.g. the same form as equation (2), with parameter estimates obtained according to estimation procedures detailed in the following examples. Escarela et al.[47] assume a bivariate copula-based probability function for the missing covariates and the missingness mechanism. Kirkham[48] instead applies the "method of weights", which assumes a parametric model for the missingness mechanism and incorporates this into the maximum likelihood estimation of parameter estimates.

Escarela et al.[47] describe how their MNAR model can also be used to impute missing values. However, this does not remove the need to make untestable assumptions on the missing data mechanism.

Category 3: Latent structures

Similarity measures

Patient similarity measures apply a sequencing algorithm to establish the alignment of two sequences of patient data, e.g. longitudinal EHR data. Sha[49] presents a novel similarity measure, which recognises that the type of tests ordered and the time between tests can be indicative of patient condition. Their metric is therefore based on a distance measure incorporating the type, timings and results of tests and they assume that more intense monitoring indicates a more severe condition.

The sequencing algorithm produces a similarity matrix, defining the similarity between each pair of patients. We do not present the model formulation for this method since there are various approaches to using this matrix in prediction (described by Sharafoddini et al.[61]). One such method defines cohorts of "similar" patients within which to develop separate models. This approach can be viewed as an extension of the pattern submodel approach with longitudinally and irregularly measured predictors, where the patterns are defined by similar longitudinal sequences.

The benefit of this method is that, as with others, it can be applied to any form of prediction framework. Drawbacks include the computational burden of re-deriving multiple models, and requiring access to the training data at prediction time to train a model using similar patients.

Latent variable

A simple way of representing a latent clinical condition is to use a single (partially) latent binary variable, representing (e.g.) one of two states. This approach was used by Coley et al.[50] and Hubbard et al.[51], where IP and IO are incorporated by allowing the measurement process to infer a latent patient condition under a hierarchical structure.

Define the partially latent binary outcome $Y^L \sim Bern(\eta)$ representing one of two patient states, where only one state is entirely observed. In Coley[50]'s example, "true" cancer state (aggressive vs indolent) is the outcome, but is only observed for a subset of patients who underwent surgery. We then assume that the value of the outcome can influence the presence of x_i within the hierarchical model.

$$R \mid Y^L, Z \sim Bern(P[R = 1 \mid Y^L, Z, \beta]) \quad (6)$$

We have not provided the outcome model formulation since predictions are obtained by sampling from the posterior of the full hierarchical model.

Both studies note improved predictive performance where the measurement process influences predictions compared to a model that ignores IP/IO. These models can, however, be computationally intensive to fit.

Hidden Markov Models

Hidden Markov Models extend the latent variable approach by allowing a time-varying latent process. Zheng et al.[52] and Alaa et al.[53] use HMMs to capture IO, but the way they incorporate the observation process differs. HMM-based prediction models incorporate IO by allowing the measurement frequency or rate to infer the clinical state at any given time.

Alaa et al.[53] propose a latent semi-Markov process to capture a patient's evolving clinical state. The "state" variable $Y^L(t) \in \{1, \ldots, 4\}$, ranges from clinical stability to clinical deterioration, where stability (state 1) and deterioration (state 4) are observed states, but intermittent states are latent. Here the model aims to predict eventual clinical deterioration, i.e. $P[Y(\infty) = 4]$. The observation process (i.e. timings) of $X(t)$ is used to infer this clinical state, where it is assumed that increased monitoring indicates a less stable condition. A marked point process model (in this case a Hawkes process) is adopted to model the rate of patient monitoring, with the marks corresponding to the

observed value. Informative observation is captured through state-specific intensity functions for the monitoring frequency as follows:

$$\lambda(t\,|Y^L(t)=1) = \lambda_1 + \alpha_1 \sum_{\tau < t_m < t} e^{-\beta_1(t-t_m)} \quad (7)$$

...

$$\lambda(t\,|Y^L(t)=4) = \lambda_4 + \alpha_4 \sum_{\tau < t_m < t} e^{-\beta_4(t-t_m)} \quad (8)$$

λ_1,..., λ_4, α_1,..., α_4,β_1,..., β_4 are state-specific parameters to be estimated. t_m is the time of the last measure of $X(t)$. τ is the time of the most recent change in $Y^L(t)$, which is only observed if the state is absorbing. Details of the learning and prediction algorithm are presented in more detail in their paper.

A key advantage is that the Hawkes process allows for a time-varying intensity in the observation process. Model fitting and interpretation are, however, complex since there are multiple components to be estimated simultaneously.

Joint modelling

Joint modelling has been developed extensively within the prediction context, particularly for dynamic prediction, i.e. incorporating time-updated variables (S2, Table 1).[62–65] Joint modelling can be extended to handling IP and IO, by linking the outcome to the observation process via a shared random effect[54–56], which can be seen as an alternative approach to modelling latent variables. Separate models are defined for the outcome occurrence and the observation process, each of them containing an individual-level random effect representing individual "frailty". By sharing these random effects across the two models, the outcome and observation processes are linked. Liang et al.[54] and Choi et al.[56] both allow for irregularly observed visits, and therefore specify a hazard or intensity function that defines how often visits occur. The random effect, or

frailty term, controls how an individual's visit rate differs from average. Since this effect also appears in the model for the outcome, the visit rate indirectly affects the prediction for the outcome.

The method outlined in [55] only allows for scheduled, regular observations. Therefore, rather than specifying a model for the intensity/hazard of visiting, the "observation process" model is a repeated measures logistic regression model, where the outcome indicates whether an individual provided data at a specific time point.

Joint models take many different forms and provide the most general framework. We present an example of a trivariate joint model, with submodels for: the repeatedly and informatively measured covariate, the binary outcome and the observation process of the covariate x_{ij}. Assuming that measurement times are regular, i.e. $t_{ij} = t_j \; \forall \; i, j$.

$$X = \alpha_0 + \alpha_1 Z + \alpha_2 t + U \qquad (9)$$

$$g(P[Y = 1|Z, U, V]) = \beta_1 Z + \beta_2 U + \beta_3 V \qquad (10)$$

$$h(P[R_j = 1|U, V, Z]) = \delta_0 U + V + \delta_1 R_{j-1} + \delta_2 Z \qquad (11)$$

Here U and V are independent subject-specific random effects, and $g(.)$ and $h(.)$ are link functions. β_2 and δ_1 control the relationships between the longitudinal predictor and the outcome, and the longitudinal predictor and the observation process respectively. β_3 controls the association between the outcome and the missingness process. Missingness at time t depends on missingness at the previous measurement time.

The listed examples illustrate the flexibility of joint modelling, as the models for both the observation outcome processes can take different functional forms. Complex dependencies between the processes can be specified. However, fitting these models can be computationally intensive, and the interpretation of random effects in a prediction model can be challenging, especially for end users.[54]

Modelling approach	Advantages	Disadvantages	Software	Assumptions
Missing indicators & Separate class	Straightforward Flexible Low computational cost Easy to communicate	Potentially doubles no. of predictors Too simplistic for complex relationships between missingness and outcome Assumes discrete time intervals	Easily applied in common statistical software	Assumes absence is a proxy for some unmeasured patient feature Linear relationship with outcome
Summary measures	Straightforward Flexible Low computational cost Easy to communicate	Generalisability of models across centres may be questioned May fail to capture complex relationships between observation process and outcome	Easily applied in common statistical software	Assumes observation process is a proxy for some unmeasured patient feature Largely assumes linear relationship with outcome
Pattern-specific models	Straightforward Flexible	Number of models becomes large as no. of predictors increases	Easily applied in common statistical software	No assumptions placed on how missingness relates to observed or unobserved variables Assumes same functional form for all pattern-specific models
Likelihood-based methods	Also allows for imputation	Computationally intensive	None provided	Assumes absence is related to the unobserved value
Similarity measures	Flexible	Computationally intensive	None provided	None provided
Latent variable	Improved performance over methods not incorporating informative presence	Computationally intensive	R code provided by Coley and Hubbard	Association between outcome and observation process is captured through latent variable and other predictors
HMMs	Using a Hawkes process for intensity allows for time-varying intensity	Complex and computationally intensive	None provided	Assumes longitudinal predictors are normally distributed
Joint modelling/shared random effects	Flexible to different forms of outcome and observation process	Complex Computationally intensive Often requires independence assumption between processes given random effects	Frailtypack in R, WinBUGS, merlin in STATA for flexible user-defined models.	Assumes processes (outcome, observation) are independent conditional on random effects Existing methods assume constant intensity of observation

Table 3: Summary of (subjective assessments of) advantages, disadvantages, software and assumptions for each method described in this review.

DISCUSSION

This study has identified three broad categories of approaches to incorporate IP and/or IO into clinical prediction models: derived predictors; modelling under informed presence; and latent structures. This is a growing area of research, and much of the included literature illustrates that informative presence and informative observation can be incorporated into clinical prediction models in a meaningful way. Where missing data and non-random visit processes have been seen as a nuisance in effect estimation, a more positive outlook is possible when the goal is prediction. Although methodology allowing CPMs to accommodate IP and IO are emerging, further challenges remain, which will be discussed later.

Pullenayegum & Lim[7] and Neuhaus[66] have previously reviewed methods for handling informative observation in studies where the primary aim is to recover unbiased effect estimates. Both articles assume that the outcome is informatively observed, which differs from the focus of our work where we assume informatively measured predictors. Phelan et al[5] present a set of design considerations for EHR-based studies that could help to attenuate issues caused by IP and IO by carefully considering and defining the population of interest, e.g. in which part of the care system patient interactions occur, and how health status could affect patient interactions. None of these articles explicitly discuss prediction, where we anticipate the most appropriate methods will differ from those for effect estimation.

Empirical studies[67][37] have compared methods capable of handling repeatedly measured predictors in CPMs, and many of these methods can be extended to accommodate IO, such as summarising the process into a single measure (e.g. the mean or maximum - measurement patterns as predictors), or more complex latent process methods. Both studies found that joint modelling provided little benefit in predictive performance when compared to simple summary measures, but care should be taken in selecting an appropriate summary measure suited to the clinical context. Bull et al.[22] also recommend three key considerations when choosing the most suitable method

for harnessing a longitudinally measured predictor: the type and amount of information available at prediction time, how the CPM can benefit from the longitudinal information and the validity of assumptions for the particular application. We expect that these considerations will also be relevant to selecting the most appropriate means of incorporating IO.

To our knowledge, this is the first attempt at synthesising the methodology available to handle IP and IO specifically for prediction purposes. We have achieved this through a systematic search of the literature. A potential limitation is that only the health and biomedical literature was considered; as such, our search potentially did not capture methods that have been developed for use in other fields. Defining relevant terminology around IP and IO is challenging, since the nomenclature differs across the literature. This is illustrated by the fact that a minority (11/36) of included papers were discovered directly through database searches. However, this is a common challenge with methodological reviews.[68,69] It is possible that methods were missed as a result, but we aimed to mitigate against this by conducting a backward and forward citation search on papers identified through the search strategy and on a set identified as relevant a priori.

Many of the methods discussed herein remain underdeveloped and future studies should investigate the degree to which these methodological choices matter for prediction contexts. We have identified multiple avenues for further research. Missing indicators, capable of handling both IP and IO, is the most common approach (in terms of number of studies included) to incorporating the observation process. Although this method is straightforward and adaptable to any type of prediction model, key challenges remain, including but not limited to the requirement to impute missing values when developing and applying the model. Under most prediction frameworks, a value must be entered for any predictor in the model when a prediction is made. The impact of using different imputation techniques at model development and prediction time should be established.

Pattern-specific models present a promising extension to the missing indicator approach, and do not require imputation at either model development or application. Further development should

explore ways to borrow strength across models, or pool together sets of patterns, to overcome the issue of developing models with few data points for rarely observed missingness patterns.

Most methods capable of handling informative observation fall under the "summary measures" category (16 papers). The simplicity of this approach is attractive, but also a concern. Simple summaries of the entire process do not capture important changes in the observation process over time, such as a sudden increase in monitoring frequency which indicates worsening state. Latent structure approaches (e.g. modelling measurement times via a nonhomogeneous point process) may be better suited to capturing longitudinal variability but are computationally intensive. Developing a more sophisticated representation of the observation process to use as a predictor is a promising avenue of further research, offering a potential trade-off between the simplicity of summary measures and the sophistication of joint modelling. These more complex measures should be compared with both joint modelling techniques and simple summary measures to assess their added benefit in terms of predictive performance and computational efficiency. We plan to perform such comparisons in a separate full empirical study.

There already exists a vast body of literature on joint modelling for prediction, particularly covering scenario S2 (incorporating longitudinal predictors). Such methods have also recently been extended to functional data,[70] allowing them to accommodate complex structures in longitudinal predictors. Joint models have also been proposed to handle IO under an inferential framework,[8,9,71,72] so it follows that there is scope to extend joint models further to exploit IO for predictive benefit, as this review revealed that the method remains underdeveloped for this particular purpose.

There are broader challenges associated with exploiting IP and IO for prediction. First, since the association between the observation process and outcome is unlikely to be causal, this relationship may not generalise well to different settings. For example, clinicians' monitoring behaviours are likely to vary across units or clinical guidelines could recommend changes in the way patients are observed. This is particularly true following the introduction of a CPM into clinical practice; once this

happens the predictor variables in the model are far more likely to be observed. The predictive utility of any model incorporating the observation process should therefore be regularly validated and potentially updated.

A second challenge described by Alaa et al.[53] concerns models that use the observation process to inform predictions, but also update predictions as new information becomes available. An issue arises when clinicians change their monitoring behaviour based on predictions produced by the model; any changes in the way they monitor patients will be fed back into future predictions via the observation process. This should be accounted for to avoid the feedback loop, potentially by developing causal models to account for the possible time-varying confounding[73], or by explicitly modelling the effects of previous predicted values.

Despite these challenges, we view IP and IO as opportunities to improve the performance of predictive models, as opposed to a nuisance. The literature is divided on this point; much of the work in this review proposes methods that "overcome" the challenges of informative presence/observation, whereas others illustrate the added benefit of incorporating informative measurement patterns. Missing data has typically been seen as a threat to the estimation of parameters, but since this is not the key focus of prediction research, it may be useful to move away from terms such as "missingness", and instead focus on what the presence of an observation can tell us.

CONCLUSION

We have demonstrated that there is a growing recognition of both informative presence and informative observation within prediction research. Although parallels exist with missing data, informative presence should not be considered the same way, especially within the context of prediction and routinely collected data where there is no pre-specified observation process. By synthesising the available methods and software that could be applied to incorporate IO/IP into

CPMs, this paper can assist applied researchers in adopting suitable methods. Future research should investigate the challenges presented herein, which will require the development of formal guidelines and making existing methodology more accessible.

FUNDING STATEMENT

This work was supported by the Medical Research Council grants MC_UU_00002/5, MC_UU_00002/2, MR/N013751/11, and the Alan Turing Institute under the "Predictive Healthcare" project (Health and Medical Sciences Programme).

COMPETING INTERESTS STATEMENT

The authors have no competing interests to declare.

CONTRIBUTORSHIP STATEMENT

RS designed the study, conducted screening, and wrote the manuscript.

LL conducted screening and provided critical revisions to the final manuscript.

GPM, MS, NP provided substantial contributions to the conception, design and conduct of the work, and provided critical revisions to the final manuscript.

JKB, BT and KDO contributed to discussions on study design and conduct, and provided critical revisions to the final manuscript.

ACKNOWLEDGEMENTS

We thank two anonymous reviewers for their thoughtful comments on our manuscript, which have undoubtedly strengthened the final version.

References

1 Goldstein BA, Navar AM, Pencina MJ, *et al.* Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Informatics Assoc* 2017;**24**:198–208. doi:10.1093/jamia/ocw042

2 Haneuse S, Daniels M. A General Framework for Considering Selection Bias in EHR-Based Studies: What Data Are Observed and Why? *EGEMS (Washington, DC)* 2016;**4**:1203. doi:10.13063/2327-9214.1203

3 Weiskopf NG, Rusanov A, Weng C. Sick patients have more data: the non-random completeness of electronic health records. *AMIA . Annu Symp proceedings AMIA Symp* 2013;**2013**:1472–7.http://www.ncbi.nlm.nih.gov/pubmed/24551421 (accessed 27 Sep 2018).

4 Johnson AEW, Pollard TJ, Shen L, *et al.* MIMIC-III, a freely accessible critical care database. *Sci data* 2016;**3**:160035. doi:10.1038/sdata.2016.35

5 Phelan M, Bhavsar NA, Goldstein BA. Illustrating Informed Presence Bias in Electronic Health Records Data: How Patient Interactions with a Health System Can Impact Inference.

6 Goldstein BA, Bhavsar NA, Phelan M, *et al.* Controlling for Informed Presence Bias Due to the Number of Health Encounters in an Electronic Health Record. *Am J Epidemiol* 2016;**184**:847–55. doi:10.1093/aje/kww112

7 Pullenayegum EM, Lim LS. Longitudinal data subject to irregular observation: A review of methods with a focus on visit processes, assumptions, and study design. *Stat Methods Med Res* 2014;**25**. doi:10.1177/0962280214536537

8 Gasparini A, Abrams KR, Barrett JK, *et al.* Mixed effects models for healthcare longitudinal data with an informative visiting process: a Monte Carlo simulation study. 2018;:1–18.

9 Neuhaus JM, McCulloch CE, Boylan RD. Analysis of longitudinal data from outcome-dependent visit processes: Failure of proposed methods in realistic settings and potential improvements. *Stat Med* 2018;**37**:4457–71. doi:10.1002/sim.7932

10 Goldstein BA, Phelan M, Pagidipati NJ, *et al.* How and when informative visit processes can bias inference when using electronic health records data for clinical research | Journal of the American Medical Informatics Association | Oxford Academic. *J Am Med Informatics Assoc* 2019;**26**:1609–17.https://academic.oup.com/jamia/article-abstract/26/12/1609/5573796 (accessed 20 Jul 2020).

11 McCulloch CE, Neuhaus JM, Olin RL. Biased and unbiased estimation in longitudinal studies with informative visit processes. *Biometrics* 2016;**72**:1315–24. doi:10.1111/biom.12501

12 Liu L, Huang X, O'Quigley J. Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics* 2008;**64**:950–8. doi:10.1111/j.1541-0420.2007.00954.x

13 Tan KS, French B, Troxel AB. Regression modeling of longitudinal data with outcome-dependent observation times: extensions and comparative evaluation. *Stat Med* 2014;**33**:4770–89. doi:10.1002/sim.6262

14 Sun J, Park D-H, Sun L, *et al.* Semiparametric Regression Analysis of Longitudinal Data with Informative Observation. *Source J Am Stat Assoc* 2005;**100**:882–9. doi:10.1198/016214505000000060

15 Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *BMJ* 2018;**361**:k1479. doi:10.1136/BMJ.K1479

16 Lin J-H, Haug PJ. Exploiting missing clinical data in Bayesian network modeling for predicting medical problems. *J Biomed Inform* 2008;**41**:1–14. doi:10.1016/J.JBI.2007.06.001

17 Sharafoddini A, Dubin JA, Maslove DM, *et al.* A New Insight Into Missing Data in Intensive Care Unit Patient Profiles: Observational Study. *JMIR Med Inf 2019;7(1)e11605 https//medinform.jmir.org/2019/1/e11605/* 2019;**7**:e11605. doi:10.2196/MEDINFORM.11605

18 Sperrin M, Petherick E, Badrick E. Informative Observation in Health Data: Association of Past Level and Trend with Time to Next Measurement. *Stud Health Technol Inform* 2017;**235**:261–5. doi:10.3233/978-1-61499-753-5-261

19 Martin GP, Jenkins DA, Bull L, *et al.* Towards a Framework for the Design, Implementation and Reporting of Methodology Scoping Reviews. *J Clin Epidemiol* Published Online First: 26 July 2020. doi:10.1016/j.jclinepi.2020.07.014

20 Sisk R, Martin G, Sperrin M, *et al.* Scoping review of informative observation in clinical prediction models: protocol. Open Sci. Framew. 2019.https://osf.io/rtqsg/

21 Vincent JL, Moreno R, Takala J, *et al.* The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Med* 1996;**22**:707–10. doi:10.1007/BF01709751

22 Bull LM, Lunt M, Martin GP, *et al.* Harnessing repeated measurements of predictor variables for clinical risk prediction: a review of existing methods. *Diagnostic Progn Res* 2020;**4**:9. doi:10.1186/s41512-020-00078-z

23 Sharafoddini A, Dubin JA, Maslove DM, *et al.* A New Insight Into Missing Data in Intensive Care Unit Patient Profiles: Observational Study. doi:10.2196/11605

24 Che Z, Purushotham S, Cho K, *et al.* Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Sci Rep* Published Online First: 2018. doi:10.1038/s41598-018-24271-9

25 Helander E, Pavel M, Jimison H, *et al.* Time-series modeling of long-term weight self-monitoring data. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. Institute of Electrical and Electronics Engineers Inc. 2015. 1616–20. doi:10.1109/EMBC.2015.7318684

26 Lipton ZC, Kale DC, Wetzel R, *et al.* Modeling Missing Data in Clinical Time Series with RNNs. In: *Proceedings of Machine Learning for Healthcare*. 2016.

27 Jarrett D, Yoon J, van der Schaar M. Dynamic Prediction in Clinical Survival Analysis using Temporal Convolutional Networks. *IEEE J Biomed Heal Informatics* 2019;:1–1. doi:10.1109/jbhi.2019.2929264

28 Barclay LM, Hutton JL, Smith JQ. Chain Event Graphs for Informed Missingness. *Bayesian Anal* 2014;**9**:53–76. doi:10.1214/13-BA843

29 Twala BETH, Jones MC, Hand DJ. Good methods for coping with missing data in decision trees. doi:10.1016/j.patrec.2008.01.010

30 Ding Y, Simonoff JS. An Investigation of Missing Data Methods for Classification Trees Applied to Binary Response Data Yufeng Ding. 2010.

31 Rodenburg FJ, Sawada Y, Hayashi N. Improving RNN Performance by Modelling Informative

Missingness with Combined Indicators. *Appl Sci* 2019;**9**:1623. doi:10.3390/app9081623

32 Li Q, Xu Y. VS-GRU: A Variable Sensitive Gated Recurrent Neural Network for Multivariate Time Series with Massive Missing Values. *Appl Sci* 2019;**9**:3041. doi:10.3390/app9153041

33 Sengupta A, Ap P, Shukla SN, *et al.* Prediction and imputation in irregularly sampled clinical time series data using hierarchical linear dynamical models. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. Institute of Electrical and Electronics Engineers Inc. 2017. 3660–3. doi:10.1109/EMBC.2017.8037651

34 Du N, Dai H, Trivedi R, *et al.* Recurrent Marked Temporal Point Processes: Embedding Event History to Vector. In: *KDD*. 2016. doi:10.1145/2939672.2939875

35 Wu S, Liu S, Sohn S, *et al.* Modeling asynchronous event sequences with RNNs. *J Biomed Inform* 2018;**83**:167–77. doi:10.1016/j.jbi.2018.05.016

36 Ghorbani A, Zou JY. Embedding for Informative Missingness: Deep Learning with Incomplete Data. In: *2018 56th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2018*. Institute of Electrical and Electronics Engineers Inc. 2019. 437–45. doi:10.1109/ALLERTON.2018.8636008

37 Goldstein BA, Pomann GM, Winkelmayer WC, *et al.* A comparison of risk prediction methods using repeated observations: an application to electronic health records for hemodialysis. *Stat Med* 2017;**36**:2750–63. doi:10.1002/sim.7308

38 Badgeley MA, Zech JR, Oakden-Rayner L, *et al.* Deep learning predicts hip fracture using confounding patient and healthcare variables. *npj Digit Med* 2019;**2**:31. doi:10.1038/s41746-019-0105-1

39 Zhang Z, Goyal H, Lange T, *et al.* Healthcare processes of laboratory tests for the prediction of mortality in the intensive care unit: A retrospective study based on electronic healthcare records in the USA. *BMJ Open* 2019;**9**. doi:10.1136/bmjopen-2018-028101

40 Fauber J, Shelton CR. Modeling 'Presentness' of Electronic Health Record Data to Improve Patient State Estimation. 2018.

41 Islam KT, Shelton CR, Casse JI, *et al.* Marked Point Process for Severity of Illness Assessment. In: *Proceedings of Machine Learning for Healthcare*. 2017.

42 Zhao J, Henriksson A, Kvist M, *et al.* Handling Temporality of Clinical Events for Drug Safety Surveillance. *AMIA . Annu Symp proceedings AMIA Symp* 2015;**2015**:1371–80.

43 Zabihi M, Kiranyaz S, Gabbouj M. Sepsis Prediction in Intensive Care Unit Using Ensemble of XGboost Models. In: *Computing in Cardiology (CinC)*. 2019. doi:10.23919/CinC49843.2019.9005564

44 Bagattini F, Karlsson I, Rebane J, *et al.* A classification framework for exploiting sparse multi-variate temporal features with application to adverse drug event detection in medical records. *BMC Med Inform Decis Mak* 2019;**19**:7. doi:10.1186/s12911-018-0717-4

45 Saar-Tsechansky M, Provost F. Handling Missing Values when Applying Classification Models. 2007.

46 Fletcher Mercaldo S, Blume JD. Missing data and prediction: the pattern submodel. *Biostatistics* Published Online First: 6 September 2018. doi:10.1093/biostatistics/kxy040

47 Escarela G, Ruiz-de-Chavez J, Castillo-Morales A. Addressing missing covariates for the

regression analysis of competing risks: Prognostic modelling for triaging patients diagnosed with prostate cancer. *Stat Methods Med Res* 2016;**25**:1579–95. doi:10.1177/0962280213492406

48 Kirkham JJ. A comparison of hospital performance with non-ignorable missing covariates: An application to trauma care data. *Stat Med* 2008;**27**:5725–44. doi:10.1002/sim.3379

49 Sha Y, Venugopalan J, Wang MD. A Novel Temporal Similarity Measure for Patients Based on Irregularly Measured Data in Electronic Health Records. In: *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. 2016. doi:10.1145/2975167.2975202

50 Coley RY, Fisher AJ, Mamawala M, *et al.* A Bayesian hierarchical model for prediction of latent health states from multiple data sources with application to active surveillance of prostate cancer. *Biometrics* 2017;**73**:625–34. doi:10.1111/biom.12577

51 Hubbard RA, Huang J, Harton J, *et al.* A Bayesian latent class approach for EHR-based phenotyping. *Stat Med* 2019;**38**:74–87. doi:10.1002/sim.7953

52 Zheng K, Gao J, Ngiam KY, *et al.* Resolving the Bias in Electronic Medical Records. In: *KDD*. 2017. doi:10.1145/3097983.3098149

53 Alaa AM, Hu S, Schaar M. Learning from Clinical Judgments: Semi-Markov-Modulated Marked Hawkes Processes for Risk Prognosis. In: *International Conference on Machine Learning (ICML)*. International Conference on Machine Learning (ICML) 2017. 60–9.http://proceedings.mlr.press/v70/alaa17a.html (accessed 7 Feb 2019).

54 Liang Y, Li Y, Zhang B. Bayesian nonparametric inference for panel count data with an informative observation process. *Biometrical J* 2018;**60**:583–96. doi:10.1002/bimj.201700176

55 Zhang N, Chen H, Zou Y. A joint model of binary and longitudinal data with non-ignorable missingness, with application to marital stress and late-life major depression in women. *J Appl Stat* 2014;**41**:1028–39. doi:10.1080/02664763.2013.859235

56 Choi Y-H, Jacqmin-Gadda H, Król A, *et al.* Joint nested frailty models for clustered recurrent and terminal events: An application to colonoscopy screening visits and colorectal cancer risks in Lynch Syndrome families. *Stat Methods Med Res* 2019;:096228021986307. doi:10.1177/0962280219863076

57 Groenwold RHH. Informative missingness in electronic health record systems: the curse of knowing. *Diagnostic Progn Res* 2020;**4**:8. doi:10.1186/s41512-020-00077-0

58 van Smeden M, Groenwold RHH, Moons KG. A cautionary note on the use of the missing indicator method for handling missing data in prediction research. *J Clin Epidemiol* 2020;**0**. doi:10.1016/j.jclinepi.2020.06.007

59 Sperrin M, Martin GP, Sisk R, *et al.* Missing data should be handled differently for prediction than for description or causal explanation. *J Clin Epidemiol* 2020;**0**. doi:10.1016/j.jclinepi.2020.03.028

60 Janssen KJM, Vergouwe Y, Rogier A, *et al.* Dealing with Missing Predictor Values When Applying Clinical Prediction Models. *Clin Chem* 2009;**55**:994–1001. doi:10.1373/clinchem.2008.115345

61 Sharafoddini A, Dubin JA, Lee J. Patient Similarity in Prediction Models Based on Health Data: A Scoping Review. *JMIR Med Informatics* 2017;**5**:e7. doi:10.2196/medinform.6730

62 Rizopoulos D. Dynamic Predictions and Prospective Accuracy in Joint Models for Longitudinal

and Time-to-Event Data. 2011;**67**:819–29. doi:10.1111/j.l541-0420.2010.01546.x

63 Hickey GL, Philipson P, Jorgensen A, *et al.* Joint modelling of time-to-event and multivariate longitudinal outcomes: Recent developments and issues. *BMC Med Res Methodol* 2016;**16**:1–15. doi:10.1186/s12874-016-0212-5

64 Król A, Ferrer L, Pignon J-P, *et al.* Joint model for left-censored longitudinal data, recurrent events and terminal event: Predictive abilities of tumor burden for cancer evolution with application to the FFCD 2000-05 trial. *Biometrics* 2016;**72**:907–16. doi:10.1111/biom.12490

65 Alsefri M, Sudell M, García-Fiñana M, *et al.* Bayesian joint modelling of longitudinal and time to event data: A methodological review. *BMC Med Res Methodol* 2020;**20**:94. doi:10.1186/s12874-020-00976-2

66 Neuhaus JM, McCulloch CE, Boylan RD. Analysis of longitudinal data from outcome-dependent visit processes: Failure of proposed methods in realistic settings and potential improvements. *Stat Med* 2018;**37**:4457–71. doi:https://dx.doi.org/10.1002/sim.7932

67 Sweeting MJ, Barrett JK, Thompson SG, *et al.* The use of repeated blood pressure measures for cardiovascular risk prediction: a comparison of statistical models in the ARIC study. *Stat Med* 2017;**36**:4514–28. doi:10.1002/sim.7144

68 Martin GP, Jenkins D, Bull L, *et al.* Towards a Framework for the Design, Implementation and Reporting of Methodology Scoping Reviews. 2020. http://arxiv.org/abs/2001.08988 (accessed 4 Mar 2020).

69 Lawson DO, Thabane L, Mbuagbaw L. A call for consensus guidelines on classification and reporting of methodological studies. *J Clin Epidemiol* 2020;**121**:109–16. doi:10.1016/j.jclinepi.2020.01.017

70 Li K, Luo S. Dynamic predictions in Bayesian functional joint models for longitudinal and time-to-event data: An application to Alzheimer's disease. *Stat Methods Med Res* 2019;**28**:327–42. doi:10.1177/0962280217722177

71 Miao R, Chen X, Sun L quan. Analyzing longitudinal data with informative observation and terminal event times. *Acta Math Appl Sin* 2016;**32**:1035–52. doi:10.1007/s10255-016-0624-3

72 Qu L, Sun L, Song X. A Joint Modeling Approach for Longitudinal Data with Informative Observation Times and a Terminal Event. *Stat Biosci* 2018;**10**:609–33. doi:10.1007/s12561-018-9221-8

73 Sperrin M, Martin GP, Pate A, *et al.* Using marginal structural models to adjust for treatment drop-in when developing clinical prediction models. *Stat Med* 2018;**37**:4142–54. doi:10.1002/sim.7913

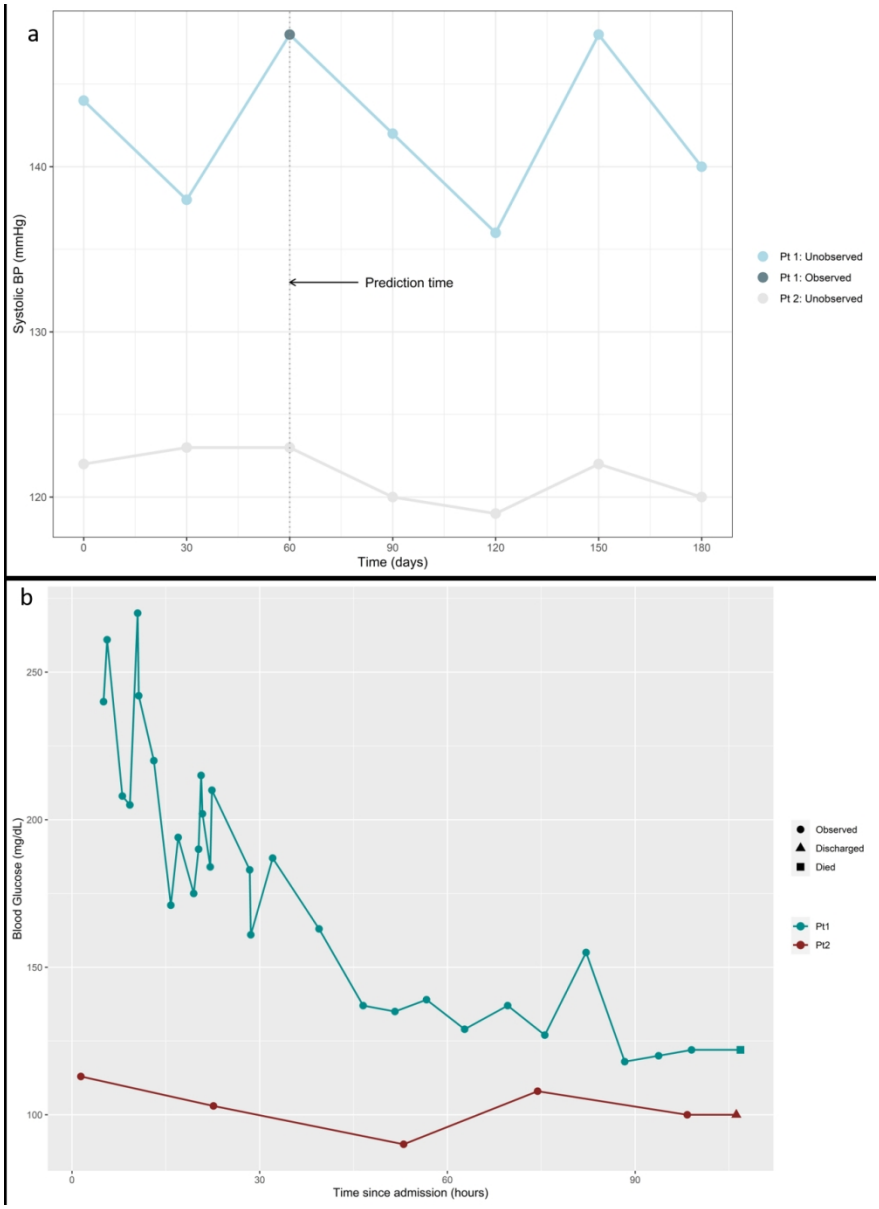


Figure 1a: An illustration of Informative Presence and how this could impact the information available at prediction time. We see the longitudinal pattern of blood pressure for two patients, with both their observed and unobserved values shown. Patient 1 has one single observed value of systolic BP, and this happens when their blood pressure was at its highest. Patient 2 has no observed values, but their blood pressure remains in the normal range - either the patient or clinician saw no clinical need to take a blood pressure measurement at any time.

Figure 1b: An illustration of Informative Observation, taken from the MIMIC dataset[4]. Patient 1 has many more in-hospital measurements of blood glucose than Patient 2 throughout their ICU admission, likely due to the fact that their Blood Glucose is much higher and much more variable than Patient 2. A more severe condition often means more intense monitoring.

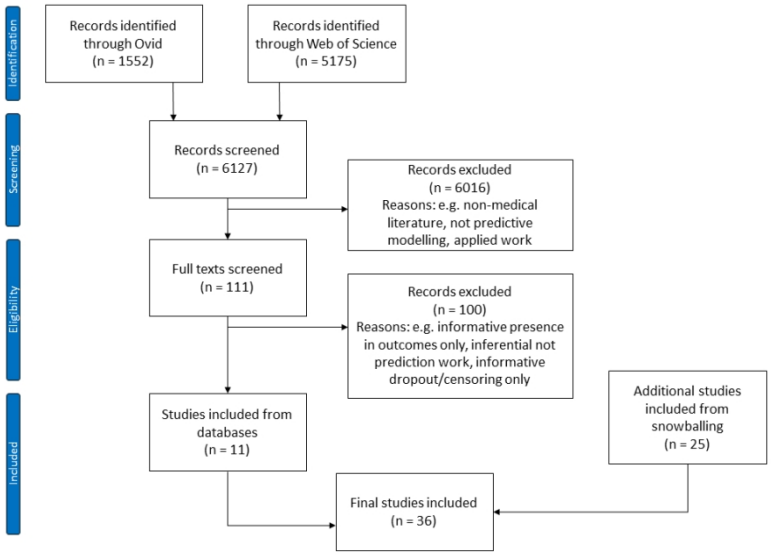


Figure 2: PRISMA flow diagram showing the various screening stages and reasons for exclusion at each stage

Supplementary material

Appendix 1: Database Searches

Searches were tailored to each database to maximise specificity of the strategy. For each database, a strategy was defined to search for terms relating to informative presence and observation. These were then combined with the Geersing filter to restrict the search to prediction-related work.

Web of Science IO & IP Strategy

Web of Science Search Terms – related to Informative Presence/Observation
TS = ((Informative* OR Informed OR Nonrandom OR Nonignorabl* OR Non-random OR Non-ignorabl*) NEAR/5 (Observation* OR Presence OR Absence OR Missing* OR Follow-up OR "follow up" OR completeness OR sampl* OR nonresponse OR non-response OR drop-out OR dropout))
TS = ("Observation process")
TS = (“Visit* process”)
TS = (“Visit* pattern”)
TS = ("Inconsistently collected data")
TS = (MNAR)
TI = ("Missing not at random")

Ovid (MEDLINE & Embase) Search Terms

Terms related to observation processes
Inform* presence
Inform* observ*
Observ* process
Inform* missing*
Inform* follow up
Inform* sampl*
Irregular* sampl*
Non random sampl*
Non random completeness
Inform* completeness
Inconsistently collected data
Visit* process
Visit* pattern
MNAR
Missing not at random[†]
Non ignorable missingness

Geersing Search Filter

The set of terms related to IP/IO was combined with the Geersing filter,[21] which has shown good sensitivity in detecting literature related to prediction model research. This filter was adapted to the

required syntax for each database, and combined with the IO/IP terms specified in the previous two sections. The filter has showed good sensitivity in picking up CPM research.

Some flexibility was allowed in the definition of the search strategy, due to uncertainty in the types of studies that would be returned. Terms were revisited and collapsed/expanded during the early phases of the first screening stage.

Ingui CPM Search Strategy + Geersing Update
(Validat* OR Predict*.ti. OR Rule*) OR (Predict* AND (Outcome* OR Risk* OR Model*)) OR ((History OR Variable* OR Criteria OR Scor* OR Characteristic* OR Finding* OR Factor*) AND (Predict* OR Model* OR Decision* OR Identif* OR Prognos*)) OR (Decision* AND (Model* OR Clinical* OR Logistic Models)) OR (Prognostic AND (History OR Variable* OR Criteria OR Scor* OR Characteristic* OR Finding* OR Factor* OR Model*)) OR Stratification OR ROC Curve OR Discrimination OR Discriminate OR c-statistic OR c statistic OR Area under the curve OR AUC OR Calibration OR Indices OR Algorithm OR Multivariable

Appendix 2: Justification of Exclusion criteria

We acknowledge that selection bias and censoring are concepts related to IP and IO. However we were primarily interested in the context where the sample accurately represents the population of interest, but the informativeness of the observation process of predictors is of primary interest. Non-medical literature was not considered, and while methods to incorporate IP/IO may exist within other fields, our primary interest is their handling within health research, and more specifically in EHRs. Imputation methods have been omitted since, under informative observation, imputing data is non-trivial, as there are no predefined time points at which data should be imputed[7] with the typical irregular patient/clinician-driven visit processes that characterises informative observation. Moreover, imputing data risks losing important information, by making all patients appear to have been monitored at equal intervals.

Appendix 3: Snowballing Set

A forward and backward citation search was performed on the following list of papers. These have all been deemed relevant to informative observation, but do not all meet the inclusion criteria for this review.

- Weiskopf, Nicole G, Alex Rusanov, and Chunhua Weng. 2013. “Sick Patients Have More Data: The Non-Random Completeness of Electronic Health Records.” *AMIA ... Annual Symposium proceedings. AMIA Symposium* 2013: 1472–77. http://www.ncbi.nlm.nih.gov/pubmed/24551421 (September 27, 2018).
- Rusanov A, Weiskopf NG, Wang S, Weng C. Hidden in plain sight: bias towards sick patients when sampling patients with sufficient electronic health record data for research. *BMC Med Inform Decis Mak*. 2014;14(1):51. doi:10.1186/1472-6947-14-51
- Phelan, Matthew, Nrupen A Bhavsar, and Benjamin A Goldstein. “Illustrating Informed Presence Bias in Electronic Health Records Data: How Patient Interactions with a Health System Can Impact Inference.”
- Sperrin, Matthew, Emily Petherick, and Ellena Badrick. 2017. “Informative Observation in Health Data: Association of Past Level and Trend with Time to Next Measurement.” *Studies in health technology and informatics* 235: 261–65.

http://www.ncbi.nlm.nih.gov/pubmed/28423794 (June 25, 2018).

- Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *BMJ*. 2018;361:k1479. doi:10.1136/BMJ.K1479
- Pivovarov R, Albers DJ, Sepulveda JL, Elhadad N. Identifying and mitigating biases in EHR laboratory tests. *J Biomed Inform*. 2014;51:24-34. doi:10.1016/J.JBI.2014.03.016
- Haneuse S, Daniels M. A General Framework for Considering Selection Bias in EHR-Based Studies: What Data Are Observed and Why? *EGEMS (Washington, DC)*. 2016;4(1):1203. doi:10.13063/2327-9214.1203
- Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Informatics Assoc*. 2017;24(1):198-208. doi:10.1093/jamia/ocw042
- Fletcher Mercaldo S, Blume JD. Missing data and prediction: the pattern submodel. *Biostatistics*. September 2018. doi:10.1093/biostatistics/kxy040
- Lin, Jau-Huei, and Peter J. Haug. 2008. "Exploiting Missing Clinical Data in Bayesian Network Modeling for Predicting Medical Problems." *Journal of Biomedical Informatics* 41(1): 1–14. https://www.sciencedirect.com/science/article/pii/S1532046407000524?via%3Dihub (October 8, 2018).

Appendix 4: paper-level summary of included articles

Author(s)	Year	Title	Broad group	Category	IP or IO	Description of method for incorporating IO/IP
Liang et al.	2018	Bayesian nonparametric inference for panel count data with an informative observation process.	Latent structures	Joint modelling	Both	- Cumulative count outcome e.g. recurrent events (tumour recurrence) - predicting future disease recurrences. - Bivariate joint model for panel count data when observation process and event processes are dependent. - Nonhomogeneous Poisson processes for event process and observation process - Stationary Gaussian processes for baseline functions of the two processes - Processes linked via correlated frailty terms, following a bivariate lognormal distribution.
Che et al.	2018	Recurrent Neural Networks for Multivariate Time Series with Missing Values.	Derived predictors	Missing indicator, summary measures	IO	-Takes multivariate time series (longitudinal predictors) data to predict diagnoses and mortality, as binary outcomes. - Uses a form of Recurrent Neural Network called the Gated Recurrent Unit. - Uses both presence/absence of predictors ("masking") and time intervals between measures as inputs in RNN. - Also allows the influence of predictors to decay over time when they have been missing for a while. - Allows for different decay rates for each predictor, to be learned from the data.

Coley et al.	2017	A Bayesian hierarchical model for prediction of latent health states from multiple data sources with application to active surveillance of prostate cancer.	Latent structures	Latent variable	IP	- Bayesian hierarchical model that predicts an individual's underlying health state via joint modelling of repeated PSA measures and biopsies. - Predictions are informed by a subset of patients for whom the true state is actually observed (those who underwent prostatectomy). Therefore prediction target "cancer state" is partially latent. - PSA (continuous predictor) is modelled using a multilevel model, with random effects (intercept and age effect) varying across latent states. - Biopsy occurrence modelled as logistic regression (binary indicator of biopsy vs no biopsy) within regular time intervals.
Sengupta et al.	2017	Prediction and imputation in irregularly sampled clinical time series data using hierarchical linear dynamical models.	Derived predictors	Summary measures	IO	- Authors develop Kalman filters that explicitly model the time difference between two measures, capturing the dependency between clinical variables and the measurement times. - The state at a given time is allowed to depend on the previous state and the time instant at which the previous observation was made. - Outcomes are all continuous physiological variables.
Zhang et al.	2013	A joint model of binary and longitudinal data with non-ignorable missingness, with application to marital stress and late-life major depression in women	Latent structures	Joint modelling	Both	- Predicting binary primary endpoint: probability of having Major Depressive Disorder (MDD), given individual trajectory of marital stress and an informative missing data mechanism. - Three components of the Shared Parameter Model: 1) Linear Mixed Model for longitudinal measures of marital stress, 2) GLM for binary primary endpoint (MDD), and 3) Shared parameter logistic regression model for the missingness mechanism - Subject-specific random effect shared across all models. - Include missingness at the previous visit as a predictor in missingness at current visit to account for dependence on prior missingness.

Escarela et al.	2016	Addressing missing covariates for the regression analysis of competing risks: Prognostic modelling for triaging patients diagnosed with prostate cancer	Modelling under informed presence	Likelihood-based methods	IP	- Likelihood-based method for estimating parameters under MAR and MNAR missingness in two categorical covariates. - Competing risks outcome, so mixture model is used. - They use a copula formulation for the covariate model and missing data mechanism.
Helander et al.	2015	Time-series modeling of long-term weight self-monitoring data.	Derived predictors	Missing indicator/summary measures	IO	- The goal is to predict future Weight, given a set of past weight data (time-series data). - The authors note that absence of weight data on a previous day predicts absence of data on the next day. - They build an ARIMA model to predict future weight, and incorporate absence flags for the M previous days in the model. M was varied between 0 and 15 days, and the best value chosen on the basis of AIC. For one subject a value of M = 3 was selected, for the other, a value of M = 9.
Barclay et al.	2014	Chain Event Graphs for Informed Missingness	Derived predictors	Separate class	IP	- A form of tree-based method, which incorporates missingness as a separate "event" in the Chain Event Graph, allowing for it to be informative of outcome. - By exploring predictions made under MAR and MNAR assumptions, the method allows us to assess plausibility of the MAR assumption.
Kirkham	2008	A comparison of hospital performance with non-ignorable missing covariates: An application to trauma care data	Modelling under informed presence	Likelihood-based methods	IP	- Outcome is 30-day survival following trauma as a dichotomous variable. - The method used to handle missing covariates is the "method of weights" in generalized linear models. - They adapt the work of Joseph Ibrahim, who proposed a ML based approach using the EM algorithm assuming a nonignorable missing mechanism. - The author anticipates that under many settings, missingness is related to the condition of the patient and therefore

						nonignorable (NMAR), and failure to observe depends on the values that would have been observed.
Alaa et al.	2017	Learning from clinical judgments: semi-Markov-modulated marked Hawkes processes for risk prognosis	Latent structures	Hidden (semi-) Markov Models	IO	- Method designed to account for an incorporate the fact that the frequency of patient monitoring in inpatient care is dependent on the patient's latent clinical state. - They propose representing the monitoring scheme as a marked Hawkes process. Intensity of the point process is defined by intensity parameters which depend on patient state (latent), and the observed physiological data are modelled using a switching multi-task Gaussian process. - Patient latent clinical states are represented as an absorbing semi-Markov jump process (absorbing to reflect the fact that episodes are informatively censored). - The target of prediction, and a patient's risk score at any time, is taken to be the probability of eventual absorption into a "clinical deterioration" state.

Zheng et al.	2017	Resolving the bias in electronic medical records	Latent structures	Hidden (semi-) Markov Models	IO	- This method recognises that EMR data are essentially an irregular time series, with irregular visiting times and different diagnoses/tests recorded at each visit. The goal is to transform it into a regular time series which is easier to analyse. - A multivariate time-series with regular intervals is created, and the hidden condition at each regular time point is learned, but uses the informative observation process to infer the hidden states. The goal is to then use methods developed for regular time series on the transformed series. - Authors define their specific type of bias as the fact that 1) patients visit hospital more often when sick and 2) doctors order lab tests that are likely to be abnormal. - A "hidden condition" is defined at each time point, which is inferred by how and whether patients with particular conditions are observed frequently. - They define the "observation rate" as "the probability of one medical feature being observed at a time point, based on its actual condition (e.g. present/absent, negative/normal)."
Islam et al.	2017	Marked Point Processes for Severity of Illness Assessment	Derived predictors	Summary measures	IO	- Prediction of mortality in the ICU from noisy, incomplete, heterogeneous, unevenly sampled patient data. - This paper fits a Piecewise Constant Conditional Intensity Model - a non-Markovian marked point process to model irregular observation streams in continuous time. - The PCIM point process can be expressed as decision tree, with internal nodes (e.g. "time between t-1 and t-5?") the binary test functions, and leaves as the "states", which define the intensity rate (easiest to visualise this - see diagrams in paper). - They learn separate PCIM models for patients who died, and those who did not. The log odds of the two models is then used as a severity score feature for individuals, and entered into a SVM classifier as a feature.

Lipton et al.	2016	Modeling Missing Data in Clinical Time Series with RNNs	Derived predictors	Missing indicator/summary measures	IO	- Goal of prediction model is multilabel classification; choosing from a set of 128 possible diagnoses, where each patient can experience more than one. - They aim to model missingness directly as a binary indicator, as the authors note that which tests were ordered can be more predictive than the results of said tests. - They also compare missingness indicators in combination with Zero Imputation and Forward Filling (LOCF) imputation techniques. The justification for LOCF is that items are likely to be measured when clinicians believe there has been a change in the value, and remain the same otherwise. - They compute a range of features related to missingness of individual items for use in the logistic regression model only, since a linear model "can only learn hard substitution rules". - They find that all models improve when either indicators or the manually computed features are included in the model, but this improvement is more modest in the logistic regression case.
Ghorbani & Zou	2018	Embedding for Informative Missingness: Deep Learning With Incomplete Data	Derived predictors	Summary measures	IP	- The aim is to provide a general framework for training neural network predictors when the training data has missing features. - The authors propose a flexible embedding method that learns a representation for missingness directly from the data. - The method does not require any imputation, and can handle informative missingness.
Li & Xu	2019	VS-GRU: A Variable Sensitive Gated Recurrent Neural Network for Multivariate Time Series with Massive Missing Values	Derived predictors	Missing indicator/summary measures	IO	- The proposed method is called variable sensitive GRU (VS-GRU). It considers the missingness rates of each variable individually rather than as a whole. - For each variable at each time point, create a missingness indicator to differentiate between imputed and observed values. Also calculate the missing rate of each predictor. - The missing indicators for each individual variable as used as

						inputs/features in the VS-GRU model, as well as the missing factor (defined in the next column).
Rodenburg et al.	2019	Improving RNN Performance by Modelling Informative Missingness with Combined Indicators	Derived predictors	Missing indicator/summary measures	Both	- The method proposes summing missing indicators to avoid the issue of potentially doubling the number of predictors in a model where using a single missing indicator for each predictor.
Sharafoddini et al.	2019	A New Insight Into Missing Data in Intensive Care Unit Patient Profiles: Observational Study	Derived predictors	Missing indicator/summary measures	IP	- Uses simple missing indicators to predict patient mortality in the ICU. - Each patient's data was summarised over every day of their admission, with indicators representing which lab tests were ordered in a particular day. - The missing indicators were added to the predictor matrix to create an augmented dataset. Missing values were imputed using Hot Deck and predictive mean matching single imputation techniques. - Feature selection methods were employed to select the most informative missing data indicators. - They attempt fitting a model on missing indicators alone, and find fairly good predictive performance. However they note that these models would not be sufficient for use in clinical practice.

Lin & Haug	2008	Exploiting missing clinical data in Bayesian network modeling for predicting medical problems	Derived predictors	Missing indicator/separate class	IP	- The method explicitly represents missing items in a clinical decision system to improve predictive performance. - All methods are a form of Bayesian Network used to predict diagnoses; a naïve Bayes structure, a human-composed network structure, and two networks based on structural learning algorithms. - They compare different ways of incorporating (or ignoring) information in the missingness, and find that those methods explicitly modelling missing items perform best. - Missingness is represented as either a separate class or a separate indicator variable.
Badgeley et al.	2019	Deep learning predicts hip fracture using confounding patient and healthcare variables	Derived predictors	Summary measures	IP	- Hospital process variables (related to image acquisition) are added as predictors in a model. - Variables considered are: department, scanner model, scanner manufacturer, laterality, study date (and day of week), order priority, technician, radiologist, radiation dose, time from image order to acquisition, time from image acquisition to initial interpretation, time from image acquisition to final interpretation. - They fit: logistic regression models and convolutional neural networks - Most hospital process vars were found to be statistically significantly associated with fracture ($p < 0.05$) - Missing items were imputed
Zhang et al.	2019	Healthcare processes of laboratory tests for the prediction of mortality in the intensive care unit: a retrospective study based on electronic healthcare records in the USA	Derived predictors	Summary measures	IO	- Similar to the Badgeley et al paper; this time predicting mortality in the ICU using variables related to the collection of lab tests. - Process variables this time are: the clock hour, the number of measurements and the measurement time from ICU admission - GLMs (logistic regression) are fitted with hospital mortality as the outcome. - AUROC increased with addition of the process variables.

Sha et al.	2016	A Novel Temporal Similarity Measure for Patients Based on Irregularly Measured Data in Electronic Health Records	Latent structures	Similarity measures	IO	- Authors create a patient similarity measure which incorporates the ordering and time intervals between lab tests. - They hypothesise that the timestamps, order, and frequency of measurements in addition to the results could carry meaningful information about patient condition. - Their similarity is novel since it incorporates time-varying information, as well as information on e.g. time intervals between tests. - The measure takes only the 10 most commonly ordered lab tests from each dataset used (MIMIC-II and CHOA). - Their novel similarity measure is compared against two non-temporal similarity measures, and find improved predictive performance. - The similarity measure is used to define patient cohorts within which to develop separate models.
Hubbard et al.	2018	A Bayesian latent class approach for EHR-based phenotyping	Latent structures	Latent variable	IP	- Develops a method that can handle informatively missing predictors using a Bayesian latent class approach in a phenotyping context. - This is an unsupervised learning method, where the true gold standard phenotype is not available for any patients. - Method assumes that true disease state is unavailable, but may influence which data are available for an individual. - The approach appears to perform well, even under MNAR. - Prior knowledge about classification accuracy of biomarkers and codes can be incorporated through suitable choice of priors.

Goldstein et al.	2017	A comparison of risk prediction methods using repeated observations: an application to electronic health records for hemodialysis	Derived predictors	Summary measures	IO	- This paper primarily compares methods that generally allow for repeated biomarker measurements in a prediction model. - However, they explore the incorporation of informed presence, by adding in the number of times a measurement is taken as a predictor. - Authors comment on the predictability of the number of measurements taken as a simple summary statistic. However they note that this is only the case for the vital signs, as these are not measured on a scheduled basis as labs would be in this setting.
Fletcher Mercaldo & Blume	2018	Missing data and prediction: the pattern submodel	Modelling under informed presence	Pattern-specific models	IP	- Develops separate models for each missingness pattern: the pattern submodel (PS) - Therefore accommodates missing data at both model development and prediction time, and does not require any imputation at development or prediction time. - The key difference with regular pattern mixture models is that only data in the observed patterns are used to develop the models; this means a that no assumptions must be placed on the missing data mechanism - The paper focuses on assessing performance at prediction time, comparing the pattern submodel with commonly applied imputation techniques.
Fauber & Shelton	2018	Modeling “Presentness” of Electronic Health Record Data to Improve Patient State Estimation	Derived predictors	Point processes	IO	- Uses Piecewise-Constant Intensity Models to build a generative model of observation times and values. - The model is used to predict future values of vital signs based on the history of these events. - They note that data are rarely MAR in medical settings, and instead that the frequency or absence of events should be used to estimate patient state. - An existing PCIM model is extended to incorporate not only the rate of events, but also values.

Zabihi et al.	2019	Sepsis Prediction in Intensive Care Unit Using Ensemble of XGboost Models	Derived predictors	Summary measures	IO	- Authors note that the pattern of missing data may convey useful information, and should therefore be used to aid prediction. - Missing data are first imputed and summary measures are computed to be used as features in the model. - Authors define sequence abstraction: each sequence is defined as a set of consecutive measures where values are either missing or present, e.g. SBP measures for a 6 hour period (1 hour intervals): {NA, 122, 98, NA, NA, 123}. Based on their definition, we have four sequences of {NA}, {122,98}, {NA, NA} and {123}. Sequence abstraction calculates and uses as features: 1) Mean and variance of the lengths of sequences along each covariate. 2) Summation and variance of the lengths of sequences with only valid values (no missing) along each covariate, and 3) Mean and variance of the lengths of sequences along each observation, in the last 5 hours. - These features representing different aspects of the missingness patterns are entered into a classifier.
Du et al.	2016	Recurrent Marked Temporal Point Processes: Embedding Event History to Vector	Derived predictors	Marked point processes	IO	- Proposes 'Recurrent Marked Temporal Point Processes' (RMTPP) to simultaneously model event timings and markers. - They aim to predict the time and type of future events from the history of a sequence of many events. - The key idea of the approach is to view the intensity function of a temporal point process as a nonlinear function of the history of the process, and parameterize the function using a recurrent neural network. - Using our model, event history is embedded into a compact vector representation which can be used for predicting the next event time and marker type. - Based on the hidden unit of RNN, we are able to learn a unified representation of the dependency over the history.

Choi et al	2019	Joint nested frailty models for clustered recurrent and terminal events: An application to colonoscopy screening visits and colorectal cancer risks in Lynch Syndrome families	Latent structures	Joint modelling	IO	- Authors introduce and use a joint nested frailty model (JNFM) to predict risk of colorectal cancer (CRC). - They incorporate the visit process of screening visits as a recurrent events process and cancer occurrence (prediction target) as a terminal event. - Model provides dynamic predictions, allowing predictions to update as new information becomes available. - They allow for an individual-specific frailty which links the processes. - Their data also contains clustering at the family level, which is included in the model via another frailty term for each family. - The number of screening visits per individual is highly irregular, and the timing between visits varies both within and between individuals.
Jarrett et al	2019	Dynamic Prediction in Clinical Survival Analysis using Temporal Convolutional Networks	Derived predictors	Missing indicators	Both	- Authors are proposing Match-Net: a Missingness-Aware Temporal Convolutional Hitting-time Network. Designed to capture temporal dependencies and heterogeneous interactions in covariate trajectories and patterns of missingness. - The model makes no assumptions regarding the underlying longitudinal or time-to-event processes. - The model can provide dynamically updated survival predictions, as well as accommodating informative patterns of missingness. - The model explicitly accounts for informative missingness by learning correlations between patterns of missingness and disease progression. - The model accounts for potential informativeness of both irregular sampling (intervals between consecutive visits and measures may vary) and asynchronous sampling (not all features are measured at the same time or same frequency).

Saar-Tsechansky	2007	Handling Missing Values when Applying Classification Models	Modelling under informed presence	Pattern-specific models	IP	- Key method of interest here is the "reduced models" approach, where separate models are developed for different missingness patterns (as described later by Fletcher Mercaldo). -Proposes developing a separate model for each missingness pattern, but using all available data, NOT just those observed within the pattern (as with Fletcher-Mercaldo's more recent paper). - Authors also propose a workaround for the possibility of having to develop huge numbers of models when p is large - develop models for "important" patterns, and using "lazy learning" or imputation for less important patterns.
Ding & Simonoff	2010	An Investigation of Missing Data Methods for Classification Trees Applied to Binary Response Data	Derived predictors	Separate class	IP	- Authors find that "separate class" (adding in an additional category for missing values) is the best method to use when the training set contains missing values and missingness is related to the outcome. - All methods here are considered under a classification tree framework, but also extended to logistic regression. Predictors must be categorised in tree-based methods, so the separate class works well here. - They also study different methods in a logistic regression model: missing indicator method, separate models for data with/without missing data (by-group method), imputing missing values with mean/mode and complete case. Missing indicator and separate models observations with/without missing values are the same as the separate class method in tree methods.
Bagattini et al	2019	A classification framework for exploiting sparse multi-variate temporal features with application to adverse drug event	Derived predictors	Summary measures	IO	- Provides a framework for using multivariate time series data to detect adverse drug events, considering that the sparsity in the available data may be useful in determining the existence of an ADE. - Proposes and compares three different methods for handling sparsity, one of which explicitly exploits it.

		detection in medical records				
Wu et al	2018	Modeling Asynchronous Event Sequences with RNNs	Derived predictors	Missing indicators/summary measures/time intervals	IO	- Discusses different ways of measuring time, and of incorporating this into RNNs, .e.g time between events, time since a landmark event, burstiness of events. - Then establishes how this information should be used in RNNs; either concatenated into the predictor matrix, or used to mediate the importance of an input. i.e. the longer something has been unobserved, the less important it is.
Zhao et al	2015	Handling Temporality of Clinical Events for Drug Safety Surveillance	Derived predictors	Summary measures	IO	- This method handles informative observation (longitudinally measured predictors) by proposing different ways of counting the number of measures (or clinical events) that occur. - The setting is in detecting Adverse Drug Events (ADEs), which are not necessarily recorded in the patient record. - The first method (Bag of events - BE) simply counts the number of times a measure occurs within D days. - Bag of Binned Events (BBE) counts the number of occurrences of each x in each day within D days. So each day has a separate feature calculated. - Bag of Weighted Events (BWE) assigns different weights to event x that occurred at different days d, and takes into account the weights when counting the number of occurrences of x. Weights are assigned according to the time distance between the event and the target ADE (prediction target). Those further away from the target ADE receive proportionally less weight.
Twala et al	2008	Good methods for coping with missing data in decision trees	Derived predictors	Separate class	IP	- Proposes "missingness incorporated in attributes" and compares against competing methods. - Method is very similar to separate class, but has also been

						extended for use in continuous predictors, where missingness can be used as the basis of a split in a tree-based model.
Agniel et al	2018	Biases in electronic health record data due to processes within the healthcare system: retrospective observational study	Derived predictors	Summary measures	Both	Explores the predictive ability of time of day, day of the week, and time between measures on mortality in inpatient admissions. Shows that the timing is a more accurate predictor of mortality than the result itself of some blood tests.