Covariates of individual patient vital sign observation timeliness in hospital wards

Rupert Ironside-Smith
School of Computer Science and Informatics
Cardiff University
Cardiff, United Kingdom
Ironside-SmithR@Cardiff.ac.uk

Stuart Allen
School of Computer Science and Informatics
Cardiff University
Cardiff, United Kingdom
AllenSM@Cardiff.ac.uk

Beryl Noë
School of Computer Science and Informatics
Cardiff University
Cardiff, United Kingdom
noeb@Cardiff.ac.uk

Liam Turner
School of Computer Science and Informatics
Cardiff University
Cardiff, United Kingdom
TurnerL9@Cardiff.ac.uk

***Abstract*—Vital sign observations are typically carried out by healthcare staff at regular intervals, known as ward rounds, to monitor the overall health status of individual patients. Patients who present symptoms of deterioration, or are deemed 'at risk' by clinical staff, will have their vital signs observed more frequently (e.g., hourly instead of on a 12 hour interval), decided by the nature of their condition. The frequency and documentation accuracy of vital signs observations has been well studied, but less consideration has been given to how clinical staff manage specific observation intervals amongst those that have been routinely scheduled. The primary aims of this study are to assess the current adherence to prescribed observation intervals and identify any primary factors that affect timeliness. This study uses empirical survival estimation methods to determine the empirical likelihood that, for a patient with a specified observation interval, the subsequent observation would be recorded within the planned interval. We discuss the management of various patient observation intervals across 20 study wards in south Wales and how these formulate routine ward rounds, or coexist with them. A semi-parametric proportional hazards model is then used to determine the extent that individual patient covariates, such as Early Warning Score (EWS), time of day, and sepsis, mediate a significant change to the baseline of vital sign observation timeliness. At the ward level, our findings suggest regular batching of vital sign observations irrespective of the planned schedules and a moderately positive linear relationship between observation interval length and the likelihood of timely recording. On an individual patient basis, elevated EWS was the strongest indicator that a patient's subsequent observation would be taken earlier than baseline, however, it is clear that baseline vital sign observation management is largely governed by existing ward round policies.**

***Index Terms*—Vital Sign Observations, Electronic Health Record, Secondary Data Study, Survival Analysis, Proportional Hazards**

I. Introduction

Healthcare staff use routinely collected vital signs observations (such as blood pressure, heart rate, respiratory rate, temperature, level of consciousness, and oxygen saturation as part of the NEWS-2 standard [36], [37]) to track the health of the individual patient in hospital wards. Patients should be observed at regular intervals (*routine* observations), such as every 4 to 12 hours, depending on the requirements of the wards and the hospital policy. The *observation interval* (*I*) (that is, the target gap between vital sign measurements) can also be shortened for an individual patient as a cautionary response (*individual* observations) when vital signs exceed thresholds [22]. In most cases these become hourly observations, but the *I* can be as short as 15 minutes depending on the underlying condition. The introduction of handheld electronic patient vital signs observation recording systems (also known as e-observations) have become essential for offering clinical staff access to real-time data for general frequency and documentation quality measures. However, the specifics of how clinical staff manage a range of observation intervals among routine operations have yet to be established and could provide new-wave insights for hospital stakeholders looking to find ways to improve staff resource allocation.

The combination of individual patient and routine observation intervals means the daily pattern of vital signs observations within a ward is non-uniform, reflecting ward and patient requirements. In practice, it has been shown that clinical staff typically consolidate most routine patient observations into 'ward rounds' [1] that take place 2 to 4 times a day [16], [24], [33]. When patients require individual plans, there is an opportunity for observations to be either 'pulled forward' or 'pushed back' into staff dealing with a larger group of beds. These patterns naturally reflect operational management at the ward level, how staff may implement policies, and other latent heuristics that support the delivery of care. In this work we assess the probability that a patient with a specified *I* will have their vital signs observed 'on time', and if not, what covariates are significant mediators.

Re-purposing of data from routine ward activities surrounding patient care, such as vital sign observations, can provide

TABLE I
CHARACTERISTICS OF THE 20 STUDY WARDS.

Ward	Year	Ward type	N_{obs}	Unique staff IDs	Unique patient IDs	Unique bed IDs
W1	2022	Surgical	25,629	177	453	32
W2	2022	Medical	28,642	139	422	30
W3	2022	Cardiology	25,476	167	406	32
W4	2022	Rehabilitation	12,073	101	178	18
W5	2022	Cardiology	28,466	54	233	34
W6	2022	Trauma & Orthopaedics	9,335	40	118	12
W7	2022	Trauma & Orthopaedics	25,137	157	543	30
W8	2022	Medical	25,998	82	1,766	36
W9	2022	Gastroenterology	35,359	150	1,353	24
W10	2022	Respiratory	37,875	155	309	30
W11	2022	Medical	30,659	221	1,268	28
W12	2022	Rehabilitation	39,366	120	651	30
W13	2022	Care of the elderly	31,461	132	224	25
W14	2019	Rehabilitation	44,768	188	904	32
W15	2019	Rehabilitation	47,671	132	1,397	30
W16	2019	Medical	45107	257	1,336	32
W17	2019	Rehabilitation	42,605	237	1576	32
W18	2022	Medical	54,449	172	1,405	31
W19	2022	Medical	22,249	92	735	17
W20	2022	Surgical	45,716	194	1,608	31

a basis for summarising how activities on wards are undertaken without the practical challenges and costs of bespoke equipment or third-party observers. A retrospective review of ward activity may be requested by hospital managers, policymakers, or other stakeholders for which the results of this study will provide a baseline of typical ward vital sign observation management, allowing for deviations thereof to be identified. This statistical representation may also provide utility in supporting changes to the structure of patient observation interval escalation or staff resource allocation.

This paper presents three main contributions: (**1**) Insight into real vital sign observations management across a range of ward sizes, specialisms, and operating characteristics, including a consideration of conformance to timeliness targets, (**2**) an evaluation of the statistical effect that individual patient characteristics have on how their subsequent observation will be managed, and (**3**), a selection of policy/ software adjustment suggestions that could improve the data quality and patient handling.

II. RELATED WORKS

In this section, we briefly summarise the current implementation and obstacles identified with respect to the practice of vital sign observations and discuss the significance of improving the manual collection of vital signs despite the introduction of continuous monitoring methods. Following this, we discuss how survival analysis methods have been used in other areas of healthcare, and how the current literature includes the effects of patient Early Warning Scores (EWS)[1] or vital sign documentation conformance.

A. Vital sign observations timeliness

The recent shift towards e-observations of vital signs provides practical and affordable clinical improvements [12] and also facilitates new analysis of inter-ward patient management [8], [13] with metrics such as timeliness and conformance [43]. However, many obstacles have been identified to completing patient observations to a target frequency and documentation standard [11], [12], [21], [22], [26], [41], partially attributed to staff interaction with e-observation systems [30], [43] and the impact of staffing levels or shift lengths on conformance and timeliness [3], [7], [9], [13], [38], [39].

Continuous automated monitoring of vital signs could be the solution to these shortcomings [42], however, the associated costs from installation, implementation, operation, and maintenance costs continue to be a barrier to the widespread adoption. Furthermore, hospitals must also consider whether patients feel confident in the level of care they receive from virtual monitoring [17], and clinical staff will always need to check more subjective patient factors, like ruminative self-focus, emotional monitoring, or social rhythms in person as well. In the status-quo, the hourly volume of vital sign observations changes throughout the day [16], [28]. It is accepted that these patterns are reflective of typical 'ward round' practice, where staff consolidate routine observations into short periods.

Stratifying vital signs observations by their timeliness (that is, the difference between the planned observation schedule and the actual time the observation is taken) has highlighted that shorter observation intervals and 'high' EWS patients (e.g., > 5 total points, or > 3 points in any single parameter for the NEWS-2 system used in the UK [36], [37]) have been shown to have the most vital sign observation omissions

[1]In a busy ward environment it is slow for healthcare staff to concisely describe the severity of a patient case using multiple individual parameters (e.g., blood pressure, heart rate, respiratory rate, temperature, level of consciousness, and oxygen saturation). So, staff combine multiple vital sign measurements and clinical judgement into a convenient single-digit number using a standardised scoring system, called an *Early Warning Score* [10], [29].

[24], [34], [38]. However, the specific causes behind late observations have not been widely studied, and these works have little mention to observations that are pulled forward unnecessarily.

B. Survival analysis

Survival analysis is a widely used statistical method to evaluate how long it will take for an event to happen, which in this study is the subsequent observation for a given patient. To the best of our knowledge, vital sign observation timeliness has not yet been modelled as time-to-event data, however, the method is prevalent in other aspects of healthcare and has been used to evaluate some other features from observations data sets. For example, Kaplan-Meier [23] survivor functions have been used to predict acute mortality in stroke patients based on their EWS score, whilst considering how individual patient observation scheduling changes in relation to EWS to reduce risk [27], and evaluate how EWS can predict in-hospital mortality when stratifying patients by risk (and therefore observation schedule) [25]. Neither work takes into account how conformance to the expected observations schedule may be a factor of individual patient risk. Likewise, previous work has utilised the Cox proportional hazards model [6] to evaluate different EWS methods for predicting patient outcome (Rapid Emergency Medicine Score, Worthing Physiological Scoring system, Charlson score, admission Barthel Index, and altered mental status at presentation) [14], [19], but do not consider that the predictive capacity of these measures could be mediated by individual ward levels of observation frequency or documentation conformance.

III. Data overview

This study works with a large, anonymised vital signs observations dataset ($N_{obs} = 770,720$) spanning 20 hospital wards from 7 sites and 8 specialisms (Medical, Surgical, Rehabilitation, Care Of The Elderly, Orthopaedic, Cardiology, and Acute Stroke) run by Aneurin Bevan University Health Board (ABUHB) in South Wales, UK. The selected wards have a well established implementation of e-observations. The time period was chosen such as that e-observations were consistently recorded for a full year (4 wards for 2019 and 16 for 2022, for which the continued effects of Covid-19 are considered to not have significant impact on vital sign observations management [24]). As part of this study, a 'vital signs observation' is defined as an 9-dimensional vector that includes: Observation ID, time of observation, ward ID, bed ID, patient NEWS[2], patient observation interval, sepsis labelling[3], 'is concerned' (a checkbox, optionally selected at the bedside, indicating either 'yes' or the default 'no') labelling, and staff role. The vital signs observation dataset was compiled within CareFlow eObservations software installed on handheld mobile devices on wards which, among other features, allow vitals sign data to be entered at the bedside with automatic NEWS and I calculations [40].

[2]The *NEWS-2* system is used in this study [36], [37].

[3]Sepsis occurs when the body's response to infection damages its own tissues and organs, where the chemicals released into the bloodstream to fight an infection trigger damaging inflammatory responses throughout the body. It is a medical emergency that requires immediate intervention, but early onset symptoms can be spotted (e.g., fever, elevated heart rate, and rapid breathing) and treatment administered from timely vital signs monitoring [32].

TABLE II
TIMELINESS BANDS TO DESCRIBE THE RATIO OF THE TTNO TO THE PLANNED OBSERVATION INTERVAL (E.G., IF TTNO=72 MINUTES AND I=60 MINUTES, TTNO:I=1.2, AND THE SUBSEQUENT OBSERVATION WILL BE CATEGORISED AS 'LATE A').

Category	Lower band	Upper band
Early		0.2
On time	0.2	1
Late A	1	1.33
Late B	1.33	1.67
Miss	1.67	

IV. Methods

Vital sign observations datasets can be appropriately transformed to a time-to-event dataset by calculating the Time-To-Next-Observation (TTNO): the actual time differential be two subsequent observations on the same patient [16]. Some works have used TTNO synonymously with the planned schedule [33], to describe the planned schedule indirectly (e.g., "expected TTNO" [24]), or inconsistently for both [4]. We would like to, at least for this study, establish the TTNO of observation v as when the subsequent observation for the same patient *is* taken, rather than when it *should* be taken (which is the observation interval, denoted as I). In cases where there is not a subsequent observation, or the TTNO exceeds 24 hours (e.g., patients between visits [4]), the outcome is considered 'right-censored' in line with convention [23]. Right-censoring data when using electronic health record datasets can lead to structural incompleteness [2], but for vital sign observations this can be mitigated by setting the TTNO as 24 hours (the end of the relevant follow-up period).

A definition for stratifying vital observations by the ratio of TTNO to the observation interval was formalised in November 2018 [13] with subsequent literature (albeit with some slight variation) maintaining the definition [18], [24], [33]. For this study, since we are simply evaluating conformance not the design of the observation interval itself, we follow convention of these standard timeliness definitions, shown in Table II.

1) Empirical TTNO analysis: We estimate the hazard distribution (i.e., the likelihood of the TTNO occurring at time t) as the derivative of the non-parametric Nelson-Aalen cumulative hazard function estimator[4] [31]. The data is stratified by ward and planned observation interval, where we include intervals that have been been used consistently by clinical staff (threshold set at $N_{obs} > 1000$). Each hazard curve describes the likelihood that an observation with a specified I will have a TTNO of time t.

[4]The cumulative sum of estimates is more stable than point-wise estimates, but the rate of change of this curve, i.e., the hazard function ($h(t)$), is more interpretable.

First, we consider whether observations across stratifications are managed similarly. To answer this, we use the *survival* distribution ($S(t)$, which represents the probability that the subsequent observation will *not* have occurred by time t) [23] in place of the hazard distribution for this null hypothesis to align with appropriate methods for this question. Although the *gold standard* for this task is the Log-rank test [35], there is a key statistical consideration; it is most powerful under proportional hazards and loses power under non-proportional hazards. Theoretically, the survival curves we derive from this data *should* proportional to the planned observation interval, but this may not the case in practice, especially if comparing curves between wards. Instead, we measure the difference in *Restricted Mean Survival Time* (RMST) [15] for 24 hours following the observation[5]. The RMST aggregates all survival information in the follow-up time to provide a heuristic, clinically meaningful interpretation for the interval management, however, as a single-number summary it may mask important temporal features. Pairwise differences in RMST (dRMST) (the average gain or loss in survival time) for the following 24 hours between each stratification will broadly indicate which intervals in which wards are similarly managed. This information can then be summarised as an edge-weighted network diagram and compared against visual representations of the hazard functions.

The cumulative density of each $S(t)$ will also be used to determine if the observation interval length correlates to the likelihood that the observation will be taken within its specified interval (i.e., *on time*, when considering the timeliness limits defined in Table II).

Although in this study we will not establish whether the empirical hazard distributions can be appropriately parametrised, survival times are often assumed to follow a specific distribution, and we consider what a hazard distribution for a 'so far as is reasonably practicable' (H_{sfairp}) treatment of a given observation interval may look like. What H_{sfairp} *should* look like is not currently known, however, we suggest it could be Weibull distribution with shape (k) and scale (λ) parameters aligned with I (Equation 1);

$$k = I \text{ (hours)}, \quad \lambda = k - \frac{k}{k + \exp\left(\frac{1-k}{2}\right)} \tag{1}$$

This leaves room for some vital sign observations be completed early or late for intentional reasons, or for reasons outside of staff control (such as overseeing a student observation, or a patient away from their bed). H_{sfairp} is only intended to be a useful baseline comparison to current ward management, not a target for hospital ward stakeholders. We simulate H_{sfairp} as a scale of I and with respect to different I values in Figures 1 and 2, respectively.

2) Semi-parametric TTNO analysis: As the reasons why routine and non-routine observations may be taken in advance of, or after, the specified observation interval (known as 'pulled forward' and 'pushed back', respectively) are not known, we employ a standard Cox proportional hazards model to determine the effect of the 5 covariants listed in Table III. These were selected for two reasons; their ease of identification by clinical staff at the time of observation, and their routine usage in the data for which similar studies may benefit. The categorical limits for NEWS and NEWS increase covariates have been selected to conform with existing guidance [4], [36], [37]. The hazard function for the TTNO is given by Equation 2.

$$h(t;\ TTNO) = h_0(t)^{(\beta_{ews}X_{ews} + \ \dots\ + \beta_{sps}X_{sps})} \tag{2}$$

The null model is fitted to predict the observation TTNO

[5] Measured as the area under the survival curve.

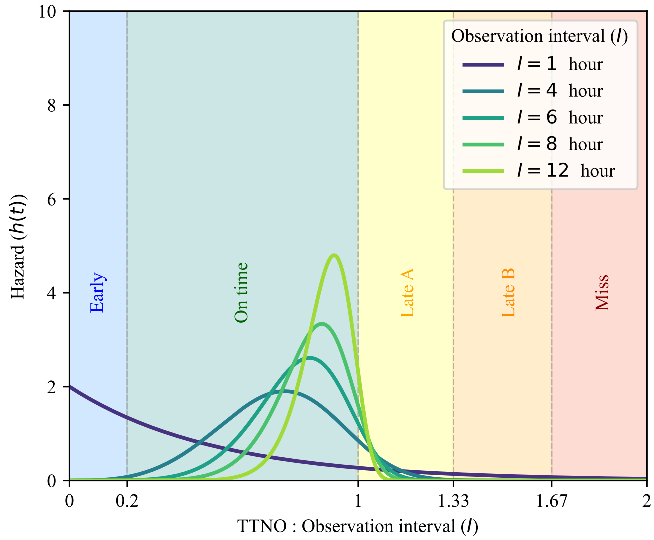

Fig. 1. An illustration of the nominal best-case hazard distribution curves ($h(t)$) for a vital sign observation with a planned interval, $h_{sfairp}(t)$.

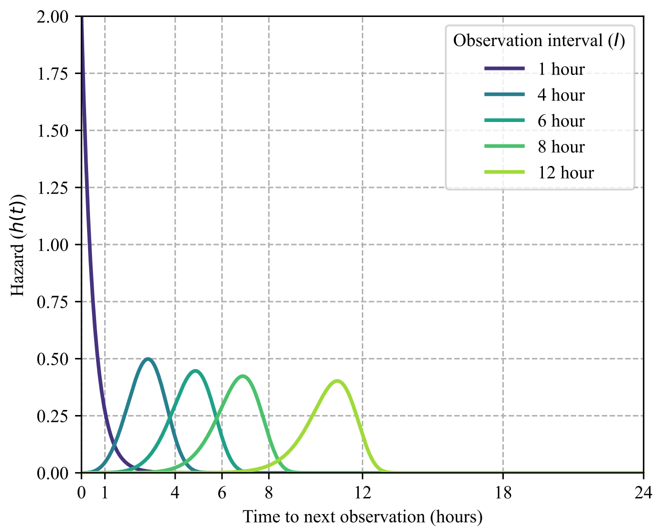

Fig. 2. An illustration of best-case hazard distribution functions, $h_{sfairp}(t)$, for a range of observation intervals (1 hour, 4 hour, 6 hour, 8 hour, and 12 hour).

TABLE III
COVARIATE DESCRIPTIONS FOR THE HAZARD FUNCTION DESCRIBED IN EQUATION 2.

Name	Notation	Values
Patient NEWS	X_{ews}	$ews \in \{0,1,2\}$, for stable (0-3); at risk (4-6); critical (7+)
NEWS increase	$X_{ews\ inc}$	$ews\ inc \in \{0,1\}$, for no low increase in NEWS; increase in NEWS of 2 or more
Observation time	X_{ToD}	$ToD \in \{0,1,2,4\}$, for overnight (0–5); morning (6–11); afternoon (12-17); evening (18–24)
Sepsis label	X_{sps}	$sps \in \{0,1\}$
‘is concerned‘ label	X_{ic}	$ic \in \{0,1\}$
Oxygen state	X_{oxy}	$oxy \in \{0,1\}$; not on oxygen (0); on oxygen (1)

over the follow-up period. We assume that the TTNO between distinct individuals in the sample are independent. We are conscious that our dataset could be perceived to contain correlated subjects, as subjects appear more than once in the dataset (i.e., patients have multiple observations taken during their stay), which would break the independent-and-identically-distributed assumption, but sampling process should account for this and avoid bias towards admissions that had more vital sign observations or covariate instances. With this, we assume a sample of vital sign observations that are taken independently and as a factor of the patient condition and/or other covariates, not the identity of the patient. We also check the proportional-hazards assumption: that there is a multiplicative relationship between the predictors and the hazard and that there is a constant hazard ratio over time.

V. RESULTS

Our findings unfold in two parts; Part *A*, where we evaluate empirical vital sign observation hazard distributions for common schedules in all wards. Figure 3 serves as a visual scale in this section to the various vital sign observation management heuristics discussed and to give context to the statistical methods used. Part *B* then discusses the assumptions of the model (using Table IV and Figure 7) and the strengths of the individual patient covariate coefficients (Table V).

A. Empirical Kaplan-Meier curve evaluation

Figure 3 introduces the analytical framework by illustrating hazard functions for commonly used vital sign observation schedules. Right inset axes illustrate the proportion of observations taken each hour, where it is agreed that the activity peaks in these plots likely represent routine ‘ward round’ procedures [16], [24], [33]. Left inset axis show the proportion of each observation interval used. The number of daily activity peaks correlate to operational differences between wards, namely, that wards scheduling 2 daily rounds appear to employ fewer varieties of observation intervals. We follow convention in discussions and group the study wards by daily observation interval distributions [24], [33], where wards W1-W8 are ‘group 1’ and W9-W20 are ‘group 2’.

Comparing Figure 3 to the reference Figure 2 highlights differences between theoretical observations management and what occurs in practice. There are distinct peaks that appear in all wards, but not necessarily aligned to individual observation intervals, rather more closely aligned to the daily activity peaks. For example, in W1, although there is a small peak that aligns to 1-2 hours for the short intervals in use, there is a much larger peak that aligns to 12 hours - likely the subsequent ward round when viewing the inset axis. The influence of ward rounds is reiterated across the rest of the study wards where peaks in the hazard distribution are created in alignment to the gaps between ward rounds. For example, in W2, like W1 there are only 2 peaks in the daily observation volumes, but there are also 2 peaks in the hazard distributions because of how the daily activity volume peaks are staggered (9 and 15 hours apart).

In most wards there is high consolidation of different observation intervals, especially between medium length and longer intervals. The similarity of observation interval management can be summarised by a network representation of pairwise dRMST between all common schedules in all wards, which we illustrate in Figure 4 (where there is an edge threshold for maximum dRMST of 1). For when 6 and 8 hour intervals are present (mostly in group 2 wards), they appear to be closely related, supporting the notion that clinical staff batch patient observations on similar intervals. As expected, most 1 hour intervals for ward group 2 ward and 12 hour intervals for ward group 1 are managed in groups, but unexpectedly, 1 hour intervals in group 1 and 12 hour intervals in group 2 wards appear to be managed similarly.

Figure 5 shows longer observations have a much higher likelihood of being *on time*, but also the largest variation in conformance between wards. Overall, there is a moderate positive linear relationship between observation interval length and the probability the observation will be taken within the interval (coef=0.735, $p < 0.005$, standard error=0.000084). We test if the distribution of probabilities for 12 hour intervals for wards with 2 daily ward rounds and for those with over 2 daily ward rounds are similar using a Kolmogorov-Smirnov test (variance ratio is 15.165:1). We find the separation of distributions is statistically significant ($p < 0.005$) and may occur because wards operating with over 2 daily rounds also use a wider variety of intervals (e.g., 4 hour, 6 hour, and 8 hour), and 12 hour intervals are simply consolidated with these during ward rounds (also see Figure 3).

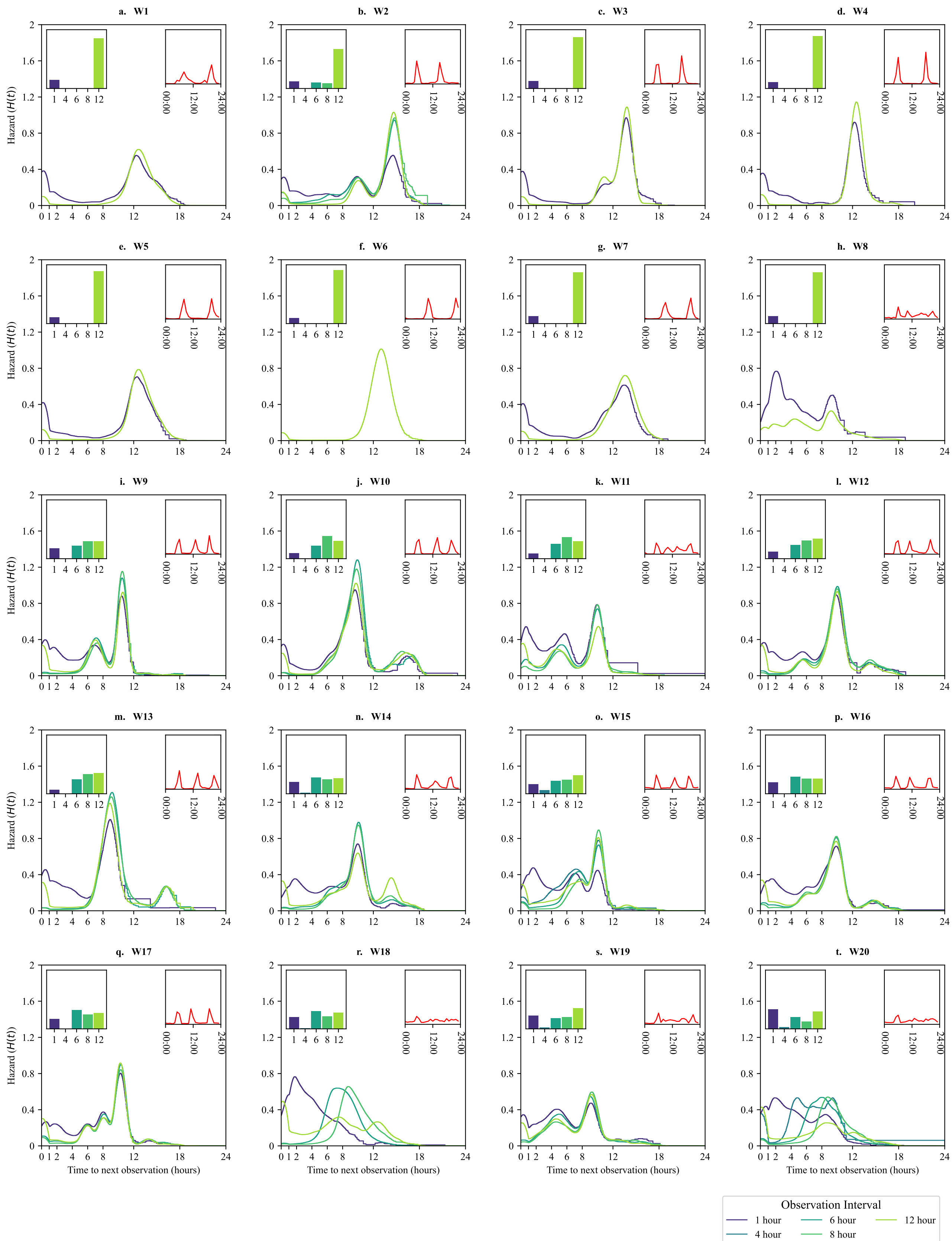

Fig. 3. Time to next observation 'hazard' distribution curves for vital sign observations stratified observation interval and by ward. Left inset axis show the proportion of each observation interval used and right inset axis show the proportion of observations taken each hour.

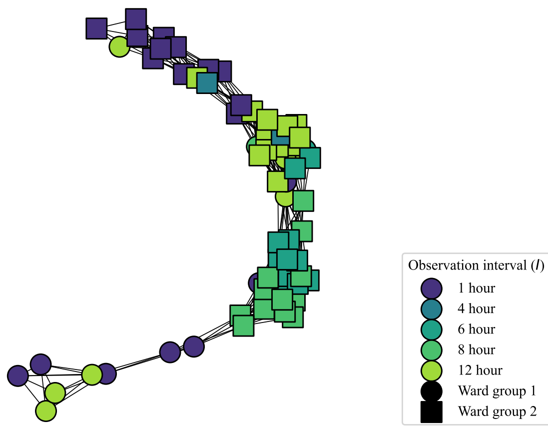

Fig. 4. Pairwise dRMST network representation.

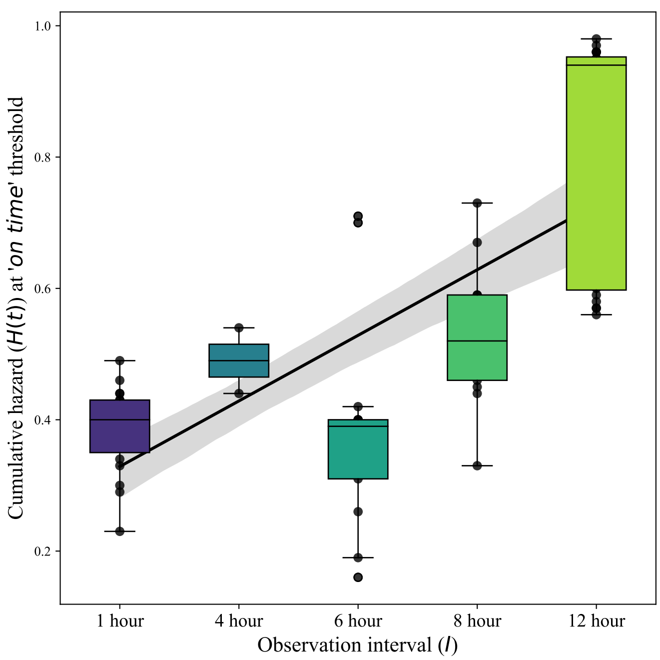

Fig. 5. Distribution of cumulative hazard probabilities that an observation for a given interval will be taken within its specified interval. Regression is scaled for visualisation purposes.

B. Covariant effects

A Cox proportional hazards model was fitted with primary strata for the ward (X_{ward}), and observation interval (X_I). Preliminary fitting showed the time of day for which an observation is taken (e.g., morning, afternoon, evening, or overnight) (X_{ToD}) to also be a notably non-proportional hazard. Reconsidering Figure 3, the influence of ward rounds create 'shoulders' in the hazard distributions. Stratifying hazard curves by time of day appears effective in separating these multimodal hazards into component distributions. Figure 6 presents four example observation schedules in different wards when stratified by the time of day. W1 and W3 are very similar wards when considering the metrics shown in Figure 3, but it is clear that the timing of ward rounds largely influences the TTNO hazard. Group 2 wards, such as W9, also show how different interval lengths appear to be treated very similarly. This covariate is best adjusted through stratification, as the categorical order for X_{ToD} is different between wards and ward round structures.

With the aforementioned stratifications the model was fitted for the remaining five covariants; X_{ews}, $X_{ews\ inc}$, X_{ic}, X_{oxy}, X_{sps}, for a sample size of $N_{obs} = 100,000$ (where 611 observations were right-censored). To determine if the chosen sample size is appropriate, we calculate statistical power for each individual covariate with a significance level of $\alpha = 0.05$. The expected probability is based on the frequency in which each covariate appears in the data and the minimum detectable effect to the hazard ratio that we want to determine. For this study, we suggest a 10% change to the hazard ratio, that is, $X_{covariate} = 1 \pm 0.1$, to be a 'clinically significant' effect. We selected a sample size to support 80% statistical power for the least frequent covariate, X_{ic}, which is only seen in about 2% of observations, but acknowledge this may lead to Type-I errors when testing for proportionality in more frequent covariates. As mentioned in Section IV, if the sample is too large, we may resample from the same patient too frequently and create biases. The power for each covariate based on each individual frequency is described in Table V.

Each covariant was checked individually against the proportional hazard assumption under a threshold set at $p < 0.05$, shown in Table IV All covariates appear to be non-proportional when using a Kaplan-Meier time transform (denoted as 'km' in Table IV), but only X_{oxy} failed under rank transformed time. Even under the null hypothesis of no violations, some covariates may appear non-proportional by chance, either because there are many covariates, or when there are lots of observations (where small deviances can be flagged, producing a type I error). We retain non-proportional hazard covariant results as a convenient summary of their effect, as in many cases, these may still be interpreted as a weighted average of the true hazard ratios over the entire follow-up period.

Martingale residuals are typically used against continuous covariates to detect nonlinearity, and although the covariates in this study are categorical, any patterns in Figure 7 can suggest where the model is not properly fit. A graphical inspection highlights clear time-dependent patterns and the presence of large negative values (corresponding to observations taken later than the model expects). The residual bands seemingly align to ward round scheduling, reaffirming their influence, suggesting that individual patient covariates lose power in dictating the TTNO during ward rounds.

The model is summarised in Table V. For every additional point attributed to a covariate ($X_{covariate}$), the 'hazard' is scaled by $\beta_{covariate}$. Where $exp(\beta_{covariate})$ is greater than 1, there is an increased likelihood that observation will be taken earlier than baseline. X_{ews} appears to be the most influential, particularly because its covariate value could increase by two points. In

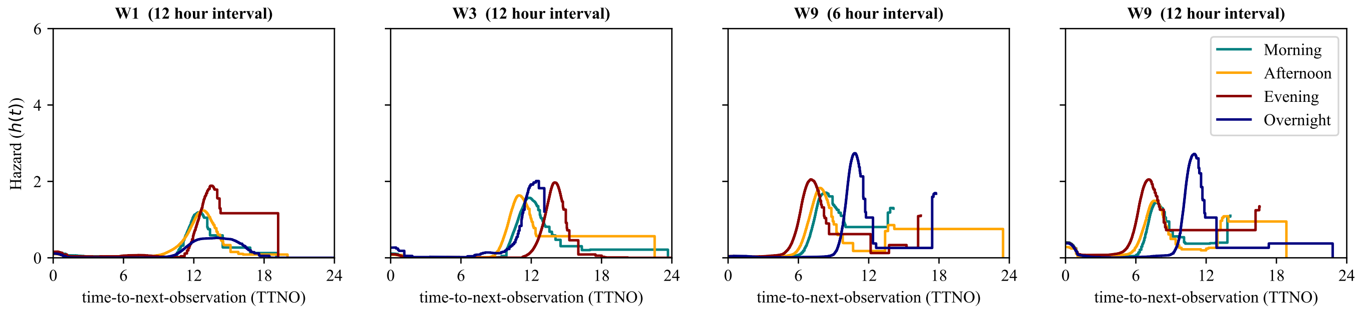

Fig. 6. TTNO hazard distribution examples stratified by the time of day of which the observation was taken.

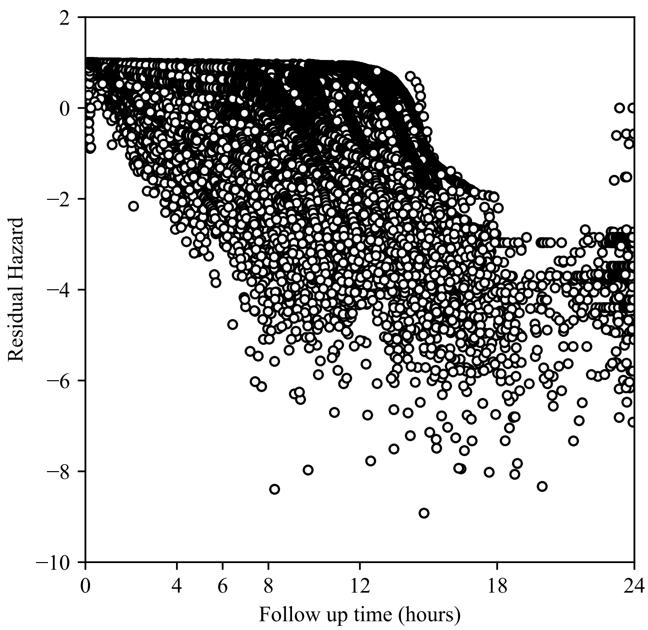

Fig. 7. Martingale residuals.

TABLE IV
INDIVIDUAL COVARIATE TESTS FOR PROPORTIONALITY.

Covariate	time transform	test statistic	p	$-log2(p)$
X_{ews}	km	51.54	< 0.005	40.38
	rank	0.39	0.53	0.91
$X_{ews\ inc}$	km	58.41	< 0.005	45.41
	rank	0.69	0.41	1.3
X_{ic}	km	18.98	< 0.005	16.21
	rank	0.41	0.52	0.94
X_{oxy}	km	995.04	< 0.005	723.08
	rank	73.95	< 0.005	56.8
X_{sps}	km	25.96	< 0.005	21.45
	rank	0.9	0.34	1.54

this data, elevated NEWS should only be present in short intervals, as a patient with NEWS over 6 is deemed 'critical' and should be reviewed within 30 minutes. Conversely, $X_{ews\ inc}$ does not seem to have as large an impact on observations management, despite also being an indicator of critical patient condition [4]. The response to observation labels was mixed; X_{ic} demonstrated a low effect, measuring close to the chosen determinable limit under our sample. X_{oxy}, although not proportional across either transformed time, measured a large effect. Lastly, X_{sps}, rather oddly, appears to encourage the pushing back subsequent observation. Sepsis labelling does only occur for 1 hour intervals, so this may instead reflect procedure for follow-up measures rather than severity.

VI. Discussion

Modelling 770,720 vital sign observations as time-to-event data has revealed new insights into the practical implementation of health policies for managing routine and non-routine vital sign observations across a broad range of ward sizes and specialities. The operational differences between wards that schedule two daily observation rounds to those scheduling three or four are substantial and clearly reflect higher-level ward requirements (e.g., specialism, patient flow, size). The curves in Figure 3 show how current clinical staff resources can manage vital sign observations alongside their other duties, such as drug administration, meal times, and staff handover. They indicate that implementing fewer varieties of observation intervals to align with ward round policy can result in short intervals being pushed back to the next ward round. Plainly offering more interval options would be an ineffective solution without additional resources or change in practice; this can be seen in many of the wards with more diverse scheduling, where 6 and 12 hour intervals are often consolidated with 8 hour schedules because it simply aligns with the daily ward round schedule.

In practice, there is a variety of potential reasons as to why vital sign observations are taken earlier or later than planned. Drops in vital sign observation timeliness and nurse staffing levels have been previously linked [13], but it is worth considering how regular ward round timings may also play a part when not aligned with the available standard available (or commonly used) observation intervals within individual wards. Looking back at what trends exist with respect to specified scheduling and the effects of individual patient covariates will at least help lead hospital management to understand the requirements for positive change in the hospital ward environment. We are motivated to offer short

TABLE V
COVARIATE COEFFICIENTS.

Covariate (X)	Power	β	$exp(\beta)$	$exp(\beta)$ lower 95%	$exp(\beta)$ upper 95%	z	p	$-log2(p)$
X_{ews}	0.845	0.57	1.78	1.69	1.87	22.36	< 0.005	365.47
X_{ewsinc}	0.880	0.12	1.12	1.07	1.18	4.79	< 0.005	19.2
X_{sps}	0.999	-0.19	0.83	0.79	0.86	-9.25	< 0.005	65.22
X_{ic}	0.792	0.14	1.15	1.09	1.21	5.28	< 0.005	22.89
X_{oxy}	1.000	0.44	1.55	1.52	1.58	43.28	< 0.005	inf

recommendations for health management policies that may improve the input quality and usefulness of the e-observations data for understanding the behaviour of patient vital sign management with respect to patient escalation structures in hospital wards.

A. Clinical recommendations

Firstly, for most observation interval stratifications, it is clear that clinical staff often rapidly repeat a patients' vital sign observations, demonstrated by the non-zero hazards at $t = 0$ in Figure 3. This has also been corroborated when considering individual staff sequences of observations [5], [20] and has been speculated to occur for patients that are either on the threshold of a 'high' NEWS score or for patients that have appropriately been put onto hourly intervals but are seen 'early' because of an 'initial review' policy (this is expected within 30 minutes for our study wards). Otherwise, it may simply be a low stakes route alteration caused for instance by a patient fall, a student observation, monitoring short-term medication effects, or in cases where a complete set of observations is required at the point of ward to ward transfers. Although the software used in this dataset does have capacity to set patients to 10, 15, and 30 minute observation intervals for patients in the appropriate condition, these short intervals are rarely utilised in this selection of study wards (less than 1000 cases seen in any ward, as mentioned in Section IV) and are more likely seen in intensive care wards not included in this study. Some of the aforementioned reasons for a rapid repeating of vital sign observations may be foreseeable for the next planned observation, and could therefore be scheduled appropriately to 15 or 30 minutes with an accompanying additional data field indicating the motivating reason why:

> **Recommendation 1.** Enable, or encourage, scheduling patients to short observation intervals (15 or 30 minutes). If possible, also using labels for contextualising the reason (e.g., 'initial review', student observations, short-term medication monitoring).

This would improve the data quality and add significant context in retrospective reviews to the proportion of rapidly repeated vital sign observations that are planned compared to those taken in response to patient or ward stimulus, which could lead to improvements in staff resource allocation.

Secondly, similarly to sub-1 hour intervals, not all wards consistently utilise medium length intervals (2, 4, 6, and 8 hours), especially wards that only schedule 2 routine vital sign observation rounds. In this dataset, medium length observation intervals are dictated by the software as a function of patient NEWS, but some wards may not have these intervals enabled and/or clinical staff may simply set all non-routine patients to hourly observation intervals manually. In practice, medium intervals may simply occur in a potentially niche space, where patients must present an elevated symptoms of deterioration without causing 'concern', otherwise they would be set to hourly observations. Figure 3 shows that many patients on hourly observation schedules get pushed back to the subsequent ward round, and that simply allowing (or encouraging when appropriate) staff to adjust patients on hourly scheduling that are likely stable to be on medium length intervals at a review point (which is 2 hours post initial escalation in our study wards) could aid staff resource allocation. It may be short-sighted to appease all 1 hour observations that are deemed late, but these results point to the need for giving staff more flexibility:

> **Recommendation 2.** Implement medium length patient observation interval scheduling selection in all wards at the point of outreach, medical team, or nurse review.

Thirdly, revisiting Table V, it is clear that patients with elevated NEWS or have additional labels (such as staff concern) are likely to be attended to earlier than baseline (e.g., a patient with NEWS of 10 will be seen before a patient with NEWS score of 4 even if they're both on 1 hour intervals). While this may seem intuitive, it suggests that either these patients might benefit from being placed on shorter observation intervals or that many patients on short intervals may not need to be (discussed for **Recommendation 2**). Adjusting the patient condition thresholds for selecting shorter intervals at an earlier stage could improve outcomes and could be easily implemented through an adjustment to the e-observations software in use:

> **Recommendation 3.** Lower the threshold for scheduling patients to shorter observation intervals (15 or 30 minutes) for those that feature both high NEWS and sepsis markers.

Overall, it is likely the recurring influence from routine ward round periods incurs the largest effect on the treatment of observations and largely overshadows any individual covariant effects. Whilst the listed recommendations have scope to improve vital sign observations management, their effectiveness may ultimately be limited by existing ward round procedures.

B. Limitations and considerations

We are aware that data gaps may exist for some cases of vital sign observations. This can occur when a patient is escalated to a doctor and the patient is under constant supervision [44] or when no charged and working devices are available at the time of the observation [5]. We also note that the difference in conformance between shorter ($I = 1$ hour) and longer ($I = 12$ hours) observation intervals could be influenced by to how the bands for on time, late and missed are considered. The definitions from the literature [13] have so far been an effective descriptor for vital sign observation timeliness [33], but, we note that it is somewhat easier for clinical staff to 'miss' a shorter interval in the case of pushing back and observation to the next scheduled ward round, which could be over three hours later. However, we understand that being late for a 1 hour observation interval may lead to worse patient outcomes, so intervals should scale accordingly.

C. Future work

In this paper we make two main assumptions; (**1**) there is a multiplicative relationship between the predictors and the hazard and that there is a constant hazard ratio over time, and (**2**) that patient care is individually managed with respect to their condition. There is opportunity for further work to determine whether the TTNO hazard distribution could be appropriately parametrised, and whether individual hazards are proportional to more specific stratifications, or if other methods like time-varying-coefficients could be applied. Secondly, its unclear how significant the underlying timeliness dependence is between patients during routine ward operations. Suppose two stable patients in neighbouring beds both have their vital signs taken during a ward round, it is not a stretch to imagine that both subsequent observations would occur in sequence during the next routinely scheduled ward round. Measuring what level of vital sign observation timeliness dependency exists in wards with limited staffing resources will unveil a novel and essential aspect of ward operations. Building on this, in cases where wards are from the same sites, it may be valuable to explore whether timeliness in one ward affects timeliness in other wards, be it though shared staff or concurrent crises.

VII. Conclusion

The management of vital sign observations in wards has previously been explored with respect to routine observations scheduling and timeliness. This study has expanded this by considering timeliness for specific observation intervals, as well as individual patient covariates such as NEWS, condition labelling (sepsis, or staff concern), and the time of day for which observation was taken. Using a dataset of 770,720 vital sign observations, this study has provided insight into: the differences in typical ward behaviour that can occur in vital sign observations management, the stratification of the probability for a TTNO for different observation intervals and wards, and the extent to which each of the observation intervals is influenced by various individual patient covariates. There is a path to smarter patient care by giving different hospital stakeholders, from staff on wards, to managers on wards and sites, to health boards and trusts, enhanced information sources that can be used as a basis of care policies designed for the individual needs of a ward. From this, we present recommendations for future care policy and supporting software to improve data quality and data analysis opportunities surrounding vital sign observations management.

Ethical Approval

The data request was approved by the Aneurin Bevan University Health Board research risk panel and received a favourable opinion from the Cardiff University School of Computer Science and Informatics research ethics committee.

Declaration of competing interest

All authors declare they have no competing interests.

Acknowledgements

The authors would like to acknowledge the assistance and cooperation of the Aneurin Bevan University Health Board for their provision of the supporting dataset and their informal commentary relating our findings to real ward operation.

Author Contributions

R.I-S. processed the experimental data, performed computations, designed figures, and drafted the manuscript. S.A., B.N., and L.D.T. supervised the study, verified the analytical methods, and extensively assisted in interpreting the findings and preparing the manuscript.

Data Availability

The dataset used in this study was provided by Aneurin Bevan University Health Board and not available for public release by the authors due to its sensitive context. Aneurin Bevan University Health Board should be contacted for access.

References

[1] ABUHB. Aneurin bevan university health board deteriorating patient policy., 2017.

[2] Ban Al-Sahab, Alan Leviton, Tobias Loddenkemper, Nigel Paneth, and Bo Zhang. Biases in electronic health records data for generating real-world evidence: An overview. *Journal of Healthcare Informatics Research*, 8(1):121–139, 2024.

[3] B. Armstrong, H. Walthall, M. Clancy, M. Mullee, and H. Simpson. Recording of vital signs in a district general hospital emergency department. *Emerg Med J*, 25(12):799–802, 2008.

[4] Jim Briggs, Ina Kostakis, Paul Meredith, Chiara Dall'ora, Julie Darbyshire, Stephen Gerry, Peter Griffiths, Jo Hope, Jeremy Jones, Caroline Kovacs, Rob Lawrence, David Prytherch, Peter Watkinson, and Oliver Redfern. Safer and more efficient vital signs monitoring protocols to identify the deteriorating patients in the general hospital ward: an observational study. *Health Soc Care Deliv Res*, 12(6):1–143, Mar 2024.

[5] Shannon Katie Costello. *Using mobile technology to record patient observations: Impact on care management and clinical practice*. PhD thesis, Cardiff University, 2024.

[6] David Cox and David Cox. Regression models and life-tables. *Journal of the royal statistical society series b-methodological*, 1972.

[7] C. Dall'Ora, P. Griffiths, O. Redfern, A. Recio-Saucedo, P. Meredith, and J. Ball. Nurses' 12-hour shifts and missed or delayed vital signs observations on hospital wards: retrospective observational study. *BMJ Open*, 2019.

[8] C. Dall'Ora, J. Hope, J. Bridges, and P. Griffiths. Development and validation of a methodology to measure the time taken by hospital nurses to make vital signs observations. *Nurse Res*, 28:52–58, 2020.

[9] Chiara Dall'Ora. *The association of nurses' shift characteristics, missed vital signs observations and sickness absence. Retrospective observational study using routinely collected data*. PhD thesis, University of Southampton, 12 2017.

[10] Kelly M. Derby, Natalie A. Hartung, Sherry L. Wolf, Sherry Wolf, Heather L. Zak, and Laura K. Evenson. Clinical nurse specialist-driven practice change: Standardizing vital sign monitoring. *Clinical Nurse Specialist*, 2017.

[11] Y. Eddahchouri, M. Koeneman, M. Plokker, E. Brouwer, T. H. van de Belt, H. van Goor, and S. J. Bredie. Low compliance to a vital sign safety protocol on general hospital wards: A retrospective cohort study. *Int J Nurs Stud*, 115:103849, 2021.

[12] O. Gale-Grant and H. Quist. Electronic recording of vital signs for mental health inpatients. *British Journal of Mental Health Nursing*, 2018.

[13] P. Griffiths, K. Ball, J.and Bloor, D. Böhning, J. Briggs, C. Dall'Ora, A. D. Iongh, J. Jones, C. Kovacs, A. Maruotti, P. Meredith, D. Prytherch, A. R. Saucedo, O. Redfern, P. Schmidt, N. Sinden, and G. Smith. Nurse staffing levels, missed vital signs and mortality in hospitals: retrospective longitudinal observational study. *NIHR Journals Library*, 2018.

[14] D. T. Ha, T. Q. Dang, N. V. Tran, N. Y. Vo, N. D. Nguyen, and T. V. Nguyen. Prognostic performance of the rapid emergency medicine score (rems) and worthing physiological scoring system (wps) in emergency department. *Int J Emerg Med*, 8:18, 2015.

[15] K. Han and I. Jung. Restricted mean survival time for survival analysis: A quick guide for clinical researchers. *Korean J Radiology*, 2022.

[16] C. Hands, E. Reid, and P. et al Meredith. Patterns in the recording of vital signs and early warning scores: compliance with a clinical escalation protocol. *BMJ*, 2013.

[17] N. Hernandez, L. Castro, J. Medina-Quero, J. Favela, L. Michan, and W. Ben Mortenson. Scoping review of healthcare literature on mobile, wearable, and textile sensing technology for continuous monitoring. *Journal of Healthcare Informatics Research*, 5(3):270–299, 2021.

[18] Joanna Hope, Alejandra Recio-Saucedo, Carole Fogg, Peter Griffiths, Gary B. Smith, Greta Westwood, and Paul E. Schmidt. A fundamental conflict of care: nurses' accounts of balancing patients' sleep with taking vital sign observations at night. *Journal of Clinical Nursing*, 2018.

[19] P. J. Huggan, F. Akram, B. H. Er, L. S. Christen, L. Weixian, V. Lim, Y. Huang, and R. A. Merchant. Measures of acute physiology, comorbidity and functional status to differentiate illness severity and length of stay among acute general medical admissions: a prospective cohort study. *Intern Med J*, 45(7):732–40, 2015.

[20] Rupert Ironside-Smith, Beryl Noë, Stuart M. Allen, Shannon Costello, and Liam D. Turner. Motif discovery in hospital ward vital signs observation networks. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 13(1):55, 2024.

[21] N. Jackson, J. Woods, P. Watkinson, A. Brent, T. E. A. Peto, A. S. Walker, and D. W. Eyre. The quality of vital signs measurements and value preferences in electronic medical records varies by hospital, specialty, and patient demographics. *Sci Rep*, 13(1), 2023.

[22] K. D. Johnson, C. Winkelman, C. J. Burant, M. Dolansky, and V. Totten. The factors that affect the frequency of vital sign monitoring in the emergency department. *J Emerg Nurs*, 40(1):27–35, 2014.

[23] Edward L. Kaplan, Edward L. Kaplan, E. L. Kaplan, Paul Meier, and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 1958.

[24] I. Kostakis, G. B. Smith, D. Prytherch, P. Meredith, C. Price, A. Chauhan, and Covid Portsmouth Academic ConsortIum For Investigating. Impact of the coronavirus pandemic on the patterns of vital signs recording and staff compliance with expected monitoring schedules on general wards. *Resuscitation*, 158:30–38, 2021.

[25] Y. S. Lee, J. W. Choi, Y. H. Park, C. Chung, D. I. Park, J. E. Lee, H. S. Lee, and J. Y. Moon. Evaluation of the efficacy of the national early warning score in predicting in-hospital mortality via the risk stratification. *J Crit Care*, 47:222–226, 2018.

[26] C. H. Leuvan and I. Mitchell. Missed opportunities? an observational study of vital sign measurements. *Crit Care Resusc*, 10(2):111–15, 2008.

[27] J. Liljehult and T. Christensen. Early warning score predicts acute mortality in stroke patients. *Acta Neurol Scand*, 133(4):261–7, 2016.

[28] F. McGain, M. A. Cretikos, D. Jones, S. Van Dyk, M. D. Buist, H. Opdam, V. Pellegrino, M. S. Robertson, and R. Bellomo. Documentation of clinical review and vital signs after major surgery. *Med J Aust*, 189(7):380–3, 2008.

[29] H. McGloin, A. Sk, and M. Singer. Unexpected deaths and referrals to intensive care of patients on general wards. are some cases potentially avoidable? *Journal of The Royal College of Physicians of London*, 1999.

[30] R. S. Miltner, K. D. Johnson, and R. Deierhoi. Exploring the frequency of blood pressure documentation in emergency departments. *J Nurs Scholarsh*, 46(2):98–105, 2014.

[31] Wayne Nelson. Hazard plotting for incomplete failure data. *Journal of Quality Technology*, 1(1):27–52, 1969.

[32] NICE. Sepsis: recognition, diagnosis and early management., 2023.

[33] B. Noë, A. Bullock, J. Frankish, and L. D. Turner. Temporal patterns in vital sign recording within and across general hospital wards. *Resusc Plus*, 10:100247, 2022.

[34] G. N. Oliveira, L. S. Nogueira, and Dalmd Cruz. Effect of the national early warning score on monitoring the vital signs of patients in the emergency room. *Rev Esc Enferm USP*, 56(spe):e20210445, 2022.

[35] R. Peto and J. Peto. Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society. Series A (General)*, 135(2):185–207, 1972.

[36] RCP. National early warning score (news) 2, 2012.

[37] RCP. News2: Additional implementation guidance., 2012.

[38] O.C. Redfern, P. Griffiths, A. Maruotti, A. Maruotti, A. Recio-Saucedo, A.R. Saucedo, and G.B. Smith. The association between nurse staffing levels and the timeliness of vital signs monitoring: a retrospective observational study in the uk. *BMJ Open*, 2019.

[39] G. B. Smith, O. Redfern, A. Maruotti, A. Recio-Saucedo, P. Griffiths, and Group The Missed Care Study. The association between nurse staffing levels and a failure to respond to patients with deranged physiology: A retrospective observational study in the uk. *Resuscitation*, 149:202–208, 2020.

[40] SystemC. Systemc careflow epr, 2023.

[41] L. S. van Galen, P. W. Struik, B. E. Driesen, H. Merten, J. Ludikhuize, J. I. van der Spoel, M. H. Kramer, and P. W. Nanayakkara. Delayed recognition of deterioration of patients in general wards is mostly caused by human related monitoring failures: A root cause analysis of unplanned icu admissions. *PLoS One*, 11(8), 2016.

[42] Franco van Wyk, Anahita Khojandi, Brian Williams, Don MacMillan, Robert L. Davis, Daniel A. Jacobson, and Rishikesan Kamaleswaran. A cost-benefit analysis of automated physiological data acquisition systems using data-driven modeling. *Journal of Healthcare Informatics Research*, 3(2):245–263, 2019.

[43] A. Watson, C. Skipper, R. Steury, H. Walsh, and A. Levin. Inpatient nursing care and early warning scores: a workflow mismatch. *J Nurs Care Qual*, 29(3):215–22, 2014.

[44] M. Yeung, S.E. Lapinsky, J. Granton, D.M. Doran, and J.A. Cafazzo. Examining nursing vital signs documentation workflow: barriers and opportunities in general internal medicine units. *Journal of Clinical Nursing*, 2012.