Determining Pre-procedure Fasting Alert Time Using Procedural and Scheduling Data

Health Informatics Journal
XX(X):1–15

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Abstract

Before a medical procedure requiring anesthesia, patients are required to not eat or drink non-clear fluids for 6 hours and not drink clear fluids for 2 hours. Fasting durations in standard practice far exceed these minimum thresholds due to uncertainties in procedure start time. The aim of this retrospective, observational study was to compare fasting durations arising from standard practice with different approaches for calculating the timepoint at which patients are instructed to stop eating and drinking. Scheduling data for procedures performed in the cardiac catheterization laboratory of an academic hospital in Canada (January 2020 to April 2022) were used. Four approaches utilizing machine learning (ML) and simulation were used to predict procedure start times and calculate when patients should be instructed to start fasting. Median fasting duration for standard practice was 10.08 hours (IQR 3.5) for both food and clear fluids intake. The best performing alternative approach, using tree-based ML models to predict procedure start time, reduced median fasting from food/non-clear fluids to 7.7 hours (IQR 2) and clear liquids fasting to 3.7 hours (IQR 2.4). 97.3% met the minimum fasting duration requirements (95% CI 96.9% to 97.6%). Further studies are required to determine the effectiveness of operationalizing this approach as an automated fasting alert system.

Keywords

Anesthesia, Machine Learning, Pre-procedure Fasting, Procedure Duration Prediction, Simulation

Introduction

Clinical practice guidelines recommend that patients should not consume clear liquids for 2 hours and not eat food or drink non-clear liquids for 6 hours before medical procedures requiring anesthesia care such as general anesthesia, regional anesthesia, or procedural sedation and analgesia [1, 2]. Despite the fact that similar recommendations have been in place for over 20 years, it is still common for patients to receive a standardized instruction such as "don't eat or drink anything after midnight". Such instructions are typically not changed regardless of alterations in scheduling even when there are significant delays in procedure start time. As a result, fasting durations far exceed the recommended requirement for most patients undergoing medical and surgical procedures [3, 4, 5]. A recent study, which was undertaken in departments where standardized fasting instructions are used, found that the mean duration of fasting from food and non-clear fluids was 12.7 hours (SD = 3.8) and clear fluids was 9 hours (SD = 4.5) [6]. Of great concern is that extended fasting periods in frail older adults can result in significant harm such as dehydration and malnutrition [7] and increased mortality [8, 9].

A solution is needed to properly address this difficult and long-standing operational problem that affects the very large number of people undergoing medical and surgical procedures with sedation or anesthesia. Prior attempts to more closely adhere to the 6 and 2-hour fasting duration recommendations that relied on additional workload for clinicians to communicate messages between departments have been ineffective (i.e., staff in the procedure department would need to contact a nurse on an in-patient ward who would then inform the patient about a change in fasting status) [10, 11]. Compounding the problem is the difficulty clinicians face in trying to anticipate procedure start time when devising appropriate cut-offs for pre-procedure fasting. It is common for there to be several changes in procedure scheduling, and there is considerable variation in durations for even similar types of procedures. With the procedural and scheduling data available in the hospital's system, it is possible to apply health informatics techniques to develop an automated solution where up-to-date fasting instructions are generated based on predictions of procedure start time. These instructions can be directly delivered to patients through mobile devices (e.g., sending the instructions as a text message to the patient's cellphone), which reduces clinician workload rather than increasing it. The aim of this study was to develop different data-driven approaches for providing pre-procedure fasting instructions that could be operationalized into an automated system that takes variations in procedure duration and changes in scheduling into account. Utilizing information technologies, we believe the automated system can improve patient experience related to pre-procedure fasting.

Materials and Methods

Design

A single-site retrospective observational design was used for this study. Approval from the University Health Network Research Ethics Board was received April 21, 2021 (ID: 21-25375). Machine learning and simulation techniques were used to predict procedure start times and the consequent fasting instruction alerts in an attempt to reduce overall

deviation from the 2-hour (clear fluids) and 6-hour (food) thresholds.

Outcomes

The outcomes for this study reflect the requirements for an automated fasting instruction system, which are to limit violations of the recommended fasting duration thresholds and minimize the total fasting duration. To evaluate the different approaches to generating fasting instruction alerts, we compared:

- Percentage of patients with a fasting duration below the 2-hour clear fluids and the 6-hour solid food fasting threshold.
- Median fasting duration from clear fluids and food.

Setting

This study was undertaken using routinely collected procedure and scheduling data for the cardiac catheterization laboratory (Cath Lab) in a large academic hospital in Canada. The Cath Lab has six procedure rooms and performs diagnostic and interventional coronary procedures, the full suite of electrophysiology and cardiac implantable electronic device procedures as well as complex interventions for valvular and congenital heart disease. Procedures are scheduled to be performed in one of the six rooms in a specific proceduralist's "session". The time that each procedure within a session is estimated to start is included in the schedule. The proceduralists do not provide input to this estimation based on characteristics of the exact procedure to be performed. Instead, the estimated start times for procedures are based on the order of procedures in the schedule, with a standardized duration allocated to similar types of procedures. For example, if a morning session begins at 8 a.m. and there are four coronary angiogram procedures scheduled to be performed, the procedure start times would be listed as 8 a.m., 8:45 a.m., 9:30 a.m. and 10:15 a.m. for the first to fourth procedure, respectively. If one of these procedures is completed early or is cancelled, the next procedure would be performed as soon as practically possible, even if this is before the scheduled start time. If a procedure takes longer to complete, the estimated times in the schedule will not be corrected. In this study, the standard fasting instruction is defined as follows. Patients who are scheduled for a procedure in a session that starts in the morning are instructed to not eat or drink anything from midnight before the procedure. If the procedure is scheduled for a session that starts after midday, patients are instructed to have a light breakfast before 6 a.m. and then not eat or drink afterwards. However, the exact instructions provided to each patient are not routinely documented at Cath Lab.

Data

The procedural data contained patient research IDs, basic demographic information, and procedure details such as procedure start time and participating staff members. Duplicate records with the same IDs were removed and missing values of patient heights and weights were filled by the means of these two features, calculated based on patient gender. The scheduling data contained 6,918 records from June 2020 to April 2022. Each record corresponded to a snapshot of the planned schedule for procedures in the Cath Lab on weekdays at every hour from 6 a.m. to 8 p.m.. The snapshots included a unique identifier, scheduled date and start time for the procedure, procedure type, name of the proceduralist, and room number as well as procedure notes made by the cardiac triage nurse. Entries in the schedules that did not have an assigned proceduralist were randomly assigned one based on an empirical proceduralist distribution for a specific procedure type.

To combine the procedural data and the scheduling data, we used a mapping file connecting the research IDs in the former dataset and the patients' medical IDs in the latter. Only records that exist in both sources were selected and the final dataset contains one-to-one mappings between the two datasets and 8,200 unique procedures.

Fasting Instruction Strategies

We compared the standard practice against the following four approaches for sending alerts directly to patients with instructions to stop either eating or drinking:

1. Fasting alerts are sent exactly 2 and 6 hours before the scheduled start times.
2. Fasting alerts are sent with a buffer added to the 2 and 6-hour time-points before the scheduled start times, where the buffer accounts for the deviation between scheduled and actual procedure start times observed in historical data.
3. Procedure start times are predicted with simulation and fasting alerts are sent 2 and 6 hours earlier.
4. Procedure start times are predicted with simulation and fasting alerts are sent with a buffer added to the 2 and 6-hour time-points before the predicted start time, where the buffer accounts for the deviation between scheduled and actual procedure start times observed in historical data.

Specific details about each approach are outlined below.

Predicting Start Times with Simulation There are several components to consider to predict procedure start times. First, there were six procedure rooms and 52 proceduralists included in the dataset. For procedures that are the first case for a given session, we defined their start times as the "session start times". Session start times were influenced by multiple factors, such as the duration of the previous procedure in the same room. For procedures within a session, the start times depended on their predecessor's duration and the changeover time. Therefore, the simulator consisted of three components: changeover times, session start times, and procedure duration predictions. Table 1 presents the fitted distributions for changeover and session start time deviations, which were selected based on minimizing the sum of square errors with the Fitter Python package [12]. We only considered the changeover times between two consecutive procedures in the same session that were less than 100 minutes. We divided the sessions into three groups based on the scheduled start time of the first case since we found that the start time deviations of procedures scheduled in different times of the day have different characteristics.

We split the data into training and test sets based on the dates of the procedures: procedures before October 23, 2021 are in the training set while the rest of the procedures are in the test set. There were 5,382 data points in the training set and 2,368 in the test set. We split the dataset temporally to be more aligned with the real-world scenario where we can only use past information to predict the future. We also tested the robustness of our approach on different data splits as shown in Appendix II.

Machine Learning Methods for Procedure Duration Prediction One key input to the simulation process was each procedure's duration. The duration of one procedure contributes to the start times of the subsequent procedures in the same session and, sometimes, in the following session. We used machine learning (ML) to predict procedure duration and included these predictions in the simulation approach.

The features selected for the machine learning model and their descriptions are shown in Table 2 with the feature importance information in Table 6 in Appendix III. A combination of patient and procedure-based features were selected either directly from the data or through feature engineering. All features describe information that can be obtained prior to procedures, such as the patient's age and weight. Proceduralist case-volume was created by counting the unique procedures performed by each proceduralist in the dataset and applying k-means clustering. We selected $k = 3$ using the elbow method. Finally, we created 9 procedure groups by inspection of procedure types.

Six ML models (linear regression, lasso regression [14], support vector regression [15], a decision tree regressor [16], a random forest regressor [17], and a gradient boosting regressor [18]) were evaluated. The root-mean-square error (RMSE) was calculated for the performance metric following different surgical duration prediction studies in the literature [19, 20, 21].

Procedure Schedule Simulation The simulation process is described in Algorithm 1. The simulation takes the schedule of $\mathcal{M}$ procedures to be performed in a day as input and generates a distribution of $\mathcal{N}$ predicted start times for each procedure. Therefore, the output of the simulation is $\mathcal{M}$ distributions of $\mathcal{N}$ predicted start times.

The *GetRandomVar*(Distribution) function (in line 4, 6, and 12) generates a random value based on the distributions described in Table 1. The end time of a procedure is calculated by adding the predicted duration to its simulated start time.

For each procedure in a session, there are three possible scenarios:

1. The procedure is the first case of the first session.
2. The procedure is the first case of a later session.
3. The procedure is not a session-first case.

The start times of procedures in Scenario 1 are calculated based on the generated deviations from the scheduled start times. For procedures in Scenario 2, the duration of the last procedure from the previous session is taken into account. Thus, the simulated start time is the maximum time between the end time of the previous procedure (*prev_ses_et*) plus changeover time, and the session's scheduled start time plus the deviation. For Scenario 3, the changeover time is added to the end time of the previous procedure (*prev_et*) to obtain the start time. The process is repeated $\mathcal{N}$ times where $\mathcal{N}$ is an arbitrarily large number (e.g., $\mathcal{N} = 1000$).

The simulation produces a start time distribution for each procedure in the snapshot and the p^{th} percentile of the distribution can be selected as the procedure's predicted start time. The p^{th} percentile start time is the time, t, such that $p\%$ of the simulated start times for the procedure are earlier than or equal to t. For example, if we want the expected proportion of fully fasted patients to be 98%, we would select the second percentile as the predicted start time.

Determining Fasting Alert Time After obtaining the predicted procedure start time, the timing of the fasting alert can be determined. One approach is to assume the predicted start times are accurate and subtract 2 and 6 hours directly from the start times. This approach is denoted as "no buffer". A second method to determine the fasting alert times is based on the historical differences between predicted and actual start times for procedures grouped by the scheduled hour j ($j \in \{8, 9, ..., 16\}$) and we denote these buffers as "historical buffers".

The method for calculating the historical buffers is described as follows. Let f be the minimum fasting duration requirements (2 or 6 hours) and b_j be the fasting requirements with an added buffer for scheduled hour group j, where group j represents the procedures that were scheduled to start within hour j. For each historical procedure, its prediction error e is the difference between the actual start time and the predicted start time. Thus, for each group j, we can obtain a prediction error distribution, $f_j(e)$. Next, a threshold, w, is defined as the minimum acceptable proportion of patients that would be fully fasted ($w \leq 1$) and we determine a set of historical buffers that satisfy this threshold by

$$b_j = f - P_{1-w}$$

where P_{1-w} represents the $(1-w)$ percentile of $f_j(e)$ and thus $-P_{1-w}$ is the buffer for group j according to our definition. The process is repeated twice for each scheduled hour j, once for calculating the buffers for clear liquids intake and once for solid food. Then, b_j hours is subtracted from the predicted start times based on the scheduled hour j to obtain the fasting alert times for each procedure.

Hourly Approaches

Recall that a new schedule snapshot is published every hour, thus, we can also use the fasting instruction strategies presented in this section with the later snapshots during the day. At each hour, we use one of the four approaches to obtain the fasting alert times. The final alert time comes from the last snapshot used that outputs a feasible time (i.e., we only update the procedure's fasting alert time if it is later than when the snapshot is captured). However, we did not choose to pursue this direction since the hourly approach is more complicated to implement and, as discussed later in this paper, did not result in significantly better performance.

Table 1. Distributions fitted to the changeover and the session start time deviations of the simulation.

Component		Distribution	Parameters (Inputs to the SciPy Python package [13])		
			Shape	Scale	Location
Changeover (Number of data points, N= 2,396)		Lognormal	0.49	35.81	-7.62
Session Start Time Deviations (N = 2,201)	Procedures scheduled between 8 and 9 a.m.	Lognormal	0.28	76.65	-56.68
	Procedures scheduled between 9 and 1 p.m.	Beta	$\alpha = 5.06$ $\beta = 1.9$	323.53	-220.2
	Procedures scheduled after 1 p.m.	Normal	N/A	68.33	-4.63

Table 2. Features used in the duration prediction ML model.

Feature	Type	Description
Age	Numerical	Age of the patient (years)
Weight	Numerical	Weight of the patient (kgs)
Gender	Categorical	Gender of the patient (i.e., Male or Female)
Height	Numerical	Height of the patient (cms)
Expected Procedure Duration	Numerical	The difference between the scheduled start times of two consecutive procedures within the same session (minutes)
Proceduralist Case-Volume Cluster	Categorical	Cluster created based on the performing surgeon's number of procedures performed
Day Of Week	Categorical	Day the procedure was performed (i.e., Monday to Friday)
Procedure Group	Categorical	Group the procedure belongs to (i.e., Implantable Loop Recorder, Pacemaker, Defibrillator, Cardiac Resynchronization, CATH, Biopsy, Structural Intervention, Coronary Intervention, Complex Coronary Intervention)

Results

Procedure Duration Prediction

A gradient boosting regressor [18] had the lowest RMSE (37.23 minutes) in 5-fold cross validation and was selected as the duration prediction model. The gradient boosting regressor has 64 estimators, a maximum depth of 5, a learning rate of 0.1, the minimum number of samples required to split an internal node is 10, the minimum number of samples required to be at a leaf node is 4, and the proportion of features to consider when looking for the best split is 30%. A summary of the results for the models that were evaluated is in Appendix I.

Comparison of Simulated Fasting Instruction Alert Strategies

We evaluated the performance of the proposed approaches for a test set of procedures performed between October 25, 2021 to April 22, 2022 and compared the results against standard practice. For the simulation-based approaches, we selected the second percentile from the procedure start time distribution as the predicted start time. For the historical buffer approaches shown in this section, we set the threshold defined in the Determining Fasting Alert Time section, w, to be 98%. That is, the historical buffers obtained ensured that 98% of the patients in historical data were fully fasted.

Table 3 shows the performance of each approach in the test set by the procedures' scheduled hour for clear liquids intake. For solid food, the percentage of fully fasted patients are the same with the fasting durations being simply an addition of 4 hours to the results for clear liquids.

Table 3 shows that the percentage of fully fasted patients for the historical buffer approaches was higher than the approaches without buffers. However, median fasting duration was higher, especially for procedures scheduled in the afternoon. To compare the two start time prediction

Algorithm 1 Simulation Process

Input: A snapshot of the procedure schedule
Output: Distribution of start times for each procedure in the snapshot

repeat
  for Session s in Snapshot **do**
    for Procedure pr in s **do**
      ChangeOver = *GetRandomVar*(Distribution)
      if pr is the first case of the first session of the day **then**
        FirstCaseDelay = *GetRandomVar*(Distribution)
        SimulatedStartTime = pr.ScheduledStartTime +
        FirstCaseDelay
      else if pr is the first case of the second, third etc. session **then**
        Calculate the end time of the last procedure in the previous
        session, *prev_ses_et*
        FirstCaseDelay = *GetRandomVar*(Distribution)
        SimulatedStartTime = max{pr.ScheduledStartTime +
        FirstCaseDelay, *prev_ses_et* + ChangeOver}
      else if pr is not a session first case **then**
        Calculate the previous procedure's end time *prev_et*
        SimulatedStartTime = ChangeOver + *prev_et*
      end if
      Record SimulatedStartTime
    end for
  end for
until Iteration $\mathcal{N}$

Table 3. Proportion of fully fasted patients (in percentages) and the median fasting duration (in hours) for clear liquids by scheduled hour.

Method		Initial Scheduled Hour								
		8 (N = 662)	9 (N = 325)	10 (N = 204)	11 (N = 231)	12 (N = 274)	13 (N = 191)	14 (N = 122)	15 (N = 65)	16 (N = 13)
Scheduled Start Time	No Buffer	86.23% (2.43 hr)	79.08% (2.82 hr)	83.82% (3.01 hr)	79.65% (3.12 hr)	76.64% (3.04 hr)	82.81% (3.32 hr)	82.79% (3.08 hr)	70.77% (2.65 hr)	76.92% (2.3 hr)
	Historical Buffer	97.88% (2.77 hr)	95.69% (3.71 hr)	95.59% (3.9 hr)	97.84% (3.98 hr)	97.81% (4.68 hr)	99.48% (5.38 hr)	98.36% (5.48 hr)	90.78% (4.8 hr)	100% (6.84 hr)
Simulated Start Time	No Buffer	89.43% (2.52 hr)	79.69% (2.7 hr)	83.82% (3.26 hr)	83.98% (3.38 hr)	79.93% (3.43 hr)	85.86% (3.75 hr)	77.05% (3.31 hr)	75.38% (3.26 hr)	76.92% (3.56 hr)
	Historical Buffer	97.43% (2.94 hr)	97.23% (3.78 hr)	98.03% (4.18 hr)	98.27% (4.21 hr)	95.62% (4.3 hr)	98.43% (5 hr)	98.36% (5.61 hr)	93.85% (5 hr)	100% (6.54 hr)

approaches, if we choose to not use any buffer to determine the fasting alert times, the simulated start time approach generally performed better than simply using the scheduled start time, though the simulated start time approach had longer median fasting durations for the afternoon procedures compared to scheduled start time. If we choose to use the historical buffers, both start time prediction approaches had similar fully fasted rates and median fasting durations.

Figures 1 and 2 illustrate the aggregated results of the percentage of patients fasting for more than x hour(s) ($x \in \{0, ..., 12\}$) for solid food and clear liquids intake, respectively. An ideal approach for would have 100% at $x = 6$ (or 2) followed by an immediate drop to zero.

Consistent with the results shown in Table 3, the approaches that applied a buffer using historical data had the best performance: 97.28% of the patients in the test set were fully fasted for the approach where simulated start times were used (95% confidence interval (CI) using a normal approximation: [96.92%, 97.64%]). For the no buffer approaches, 83.94% and 81.84% were fully fasted for simulated and scheduled start times, respectively (95% CI: [83.14%, 84.74%], [81%, 82.68%]). Note that the results from Figure 2 are essentially the same as Figure 1 but shifted 4 hours to the left because the only difference is the timing of the fasting alert for solid food and liquids.

Figures 3 and 4 compare approaches for both outcomes. The x-axis corresponds to each approach's percentage of underfasted patients at $x = 6$ or 2 and the y-axis shows the median fasting duration. The ideal points are (0,6) and (0,2), respectively (i.e., all patients exactly meet the requirements). Standard practice resulted in the longest fasting duration: 10.08 hours (Interquartile range (IQR) = 3.45) for both food and clear liquids intake.

The approaches that applied a historical buffer had shorter median fasting durations than standard practice. The average fasting from solid food duration was 7.77 hours (IQR = 2.43) for the scheduled start times, and 7.7 hours (IQR = 2.35) for the simulated start times. Clear liquids median fasting durations were 3.77 hours (IQR = 2.43) and 3.77 hours (IQR = 2.35), respectively. The approaches that directly used 2 and 6 hours resulted in shorter fasting duration but a higher

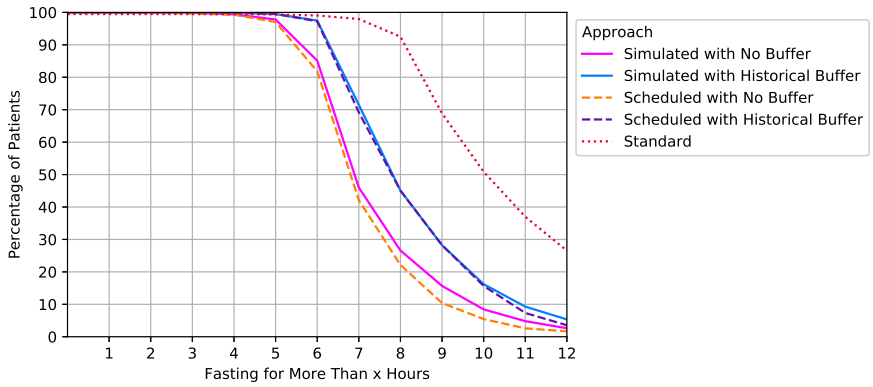

Figure 1. Percentage of patients fasting for more than x hours for solid food intake by approaches.

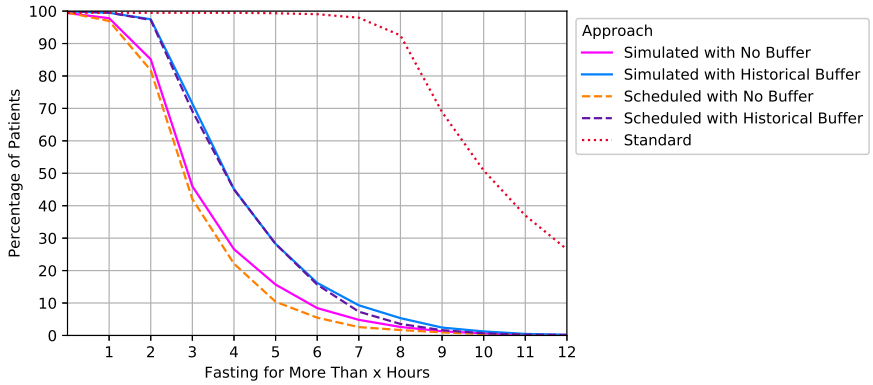

Figure 2. Percentage of patients fasting for more than x hours for clear liquids intake by approaches.

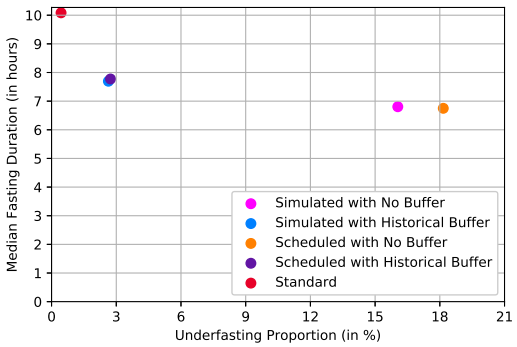

Figure 3. The median fasting duration and the proportion of underfasted patients for solid food.

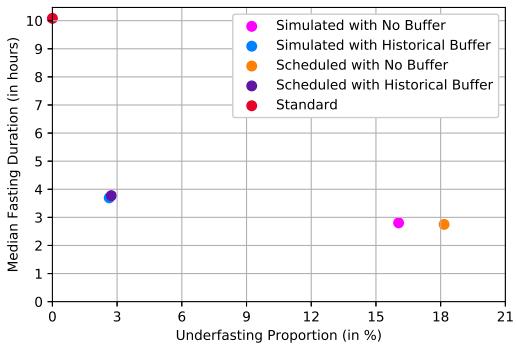

Figure 4. The median fasting duration and the proportion of underfasted patients for clear liquids.

proportion of underfasted patients, which is consistent with the results shown in Figures 1 and 2.

Performance of the Hourly Approaches

We also tested rerunning each of the four approaches presented above every hour to calculate the new fasting alert times and the results are shown in Appendix IV.

Since the final alert time is based on the last snapshot that produced a feasible time, we recorded the snapshot used for the fasting alert time of each procedure in the test set, and we aggregated the results by each snapshot. Figure 5 shows the frequency of snapshots used when fasting alerts were generated from simulated start times with no buffer at every hour. It shows that 84% of the solid food fasting alerts were generated from the initial snapshot of the day (i.e. the 6 a.m. schedule). Figure 6 shows the frequency of the snapshots used for clear liquids fasting alerts, which is less right-skewed compared to Figure 5. Although more procedures used the later snapshots to determine the fluid fasting alert times, Figure 8 (Appendix IV) indicates that using the hourly snapshots led to little or no improvement compared to only using the initial snapshot.

Discussion

All four approaches for generating times for fasting alerts shortened the total fasting duration considerably for both clear liquids and solid food intake compared to standard practice. The solution using scheduled start times and

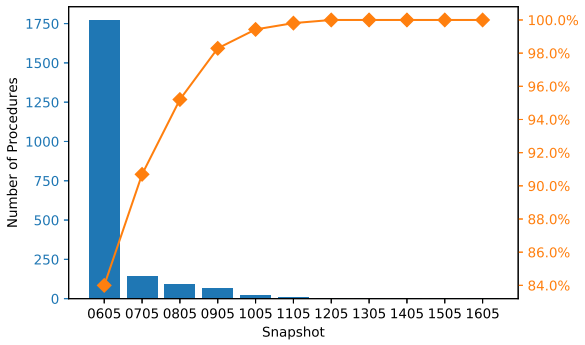

Figure 5. Pareto chart of the frequency of snapshot used for solid food fasting alert (hourly approach with simulated start times and no buffer).

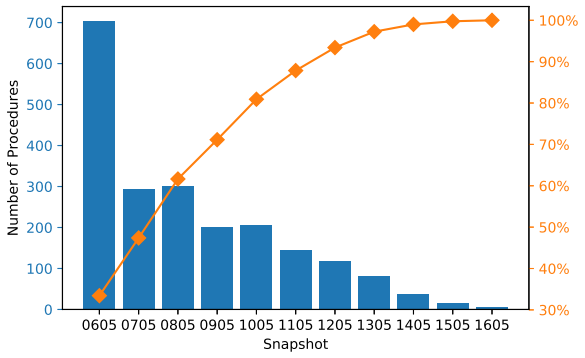

Figure 6. Pareto chart of the frequency of snapshot used for clear liquids fasting alert (hourly approach with simulated start times and no buffer).

historical buffers had the best overall performance in terms of limiting the number of patients who violate the 2 and 6-hour fasting recommendation and minimizing total fasting duration. In addition, as long as scheduled start times are readily available in a hospital's database, integrating this solution would be simpler than the simulation-based approaches because the algorithm computes the solution quickly and requires few computational resources. The logic behind the protocol (i.e. "subtract 2 or 6 hours with a historically calculated buffer time from the scheduled start time") is also more readily understood by patients and staff. Additional research would be needed to determine the best frequency for rerunning the buffer calculation algorithm with more historical data of the differences between scheduled and actual procedure start times (e.g. once every month or once every three months).

Instead of directly predicting the start times of procedures, we decided to use simulation to represent our knowledge of how the daily procedure schedule is executed. We did so by breaking the schedule down into several components using historical data. One of the components was to predict the durations of the procedures, for which there are two major lines of research in the literature. The first direction is fitting the data to a known distribution, such as normal [22, 23] or log-normal [24, 25, 26, 27]. While studies show that such models can produce accurate predictions and increase operating room efficiency [27], they did not offer insight into the factors that may impact the predictions. Therefore, we chose to focus on the second direction, which uses statistical and machine learning models to identify important features and produce predictions. The features selected for our prediction model are all known prior to the procedure and can be divided into two categories: patient factors and clinical factors. Basic information about the patients such as age and gender is often selected for the modeling process [21, 28, 29]. Some other features that are popular in the literature such as American Society of Anesthesiologists score (ASA class) [28, 30, 31] were not available in our datasets.

Common ML models used for procedure duration prediction include linear regression [32, 33, 30, 21, 34, 35, 36], tree-based algorithms such as regression trees, bagged trees, and random forests [34, 21, 31, 37, 19, 20], neural networks [34], and Support Vector Machines (SVM) [37, 20]. Of the features that were selected (Table 2), Table 6 in Appendix III indicates that the expected duration feature has higher impact on the tree-based models than on the linear regression models, while the procedure group feature is important for all models tested. In this study, we chose to focus on classical machine learning algorithms instead of more involved models such as neural networks due to the issue of interpretability. Since the target users for this application are clinicians and nurses, we believe the ML models should be easy to understand and apply. We chose a gradient boosting regressor because the model has the lowest RMSE, and combining tree-based models with boosting techniques is also fairly common in literature [31, 38, 28, 20, 39].

When using historical buffers, the more complicated simulation-based approach had nearly identical results with the scheduled start time approach. The exact cause of the comparative performance is not known since there were multiple components used in the simulation process. One hypothesis is that this lack of difference may be a result of little variability in procedure durations in different procedure groups. However, Table 7 in Appendix V shows that there is considerable difference in procedure duration means and standard deviations, indicating that the similar performance between the two approaches are not related to the duration variability. One observation is that the simulation process relied on the order of procedures in each snapshot. Thus, similar to the approach with scheduled start times, the simulation was built on the assumption that the procedures were executed in that exact order, which may not be true. While the simulation captured the uncertainty in procedure durations and changeovers, information such as a procedure being moved from the morning to an afternoon session is not available in the initial snapshot. Even if these changes were included in the later snapshots, since most of the fasting alerts were determined using the initial snapshot, neither of the approaches could use the new information when predicting procedure start times. Further, note that the two start time prediction approaches are not independent: to produce the start time distributions, we used the scheduled start time for each session's first procedure to start the simulation process. As such, one of the reasons for the similar performance may be that both methods depend on the schedule in the snapshots. To improve the simulation-based approach, we would need to develop an additional component for the system that can accurately predict, or,

actually specify, the order of procedures and use this prediction instead of the snapshot as the input to the simulation process.

The approaches that applied a historical buffer to generate fasting alerts reduced fasting duration, but it is important to consider the implications arising from the small proportion of patients in the test set who had not yet reached the 2 and 6-hour minimum fasting requirements before the "actual" procedure start time. When a patient has not fasted for the minimum recommended duration, there are two potential actions that can be taken: i) delay the procedure until the patient meets the minimum fasting period; or ii) change the schedule to start a procedure for a patient who has met the fasting requirement. These two approaches have their trade-offs in terms of costs, risks, and patient experience. For the first option, leaving the procedure room idle is costly and would delay subsequent procedures. Further, this option may not be applicable to urgent or emergency procedures as they may not need to be delayed to accommodate fasting requirements [40]. The second option may not be ideal for the hospital either, since changing the schedule may lead to logistical difficulties and confusion among staff and patients.

It is interesting that using schedules at each hour throughout the day to generate fasting alerts did not result in improved overall performance of the system in comparison to approaches using the first schedule. One reason why the hourly approaches did not outperform the single snapshot approach is that the majority of procedures were scheduled before noon. Therefore, most of the patients would start fasting for solid food early in the morning, which means that the alerts would be sent out based on the start times generated using the initial snapshot. Another reason is that the snapshots of each day remained mostly unchanged compared to the initial snapshot, meaning that even if the alert comes from a later snapshot in the day, it would be very similar or identical to the one from the initial snapshot.

Limitations

The data used in this study were extracted during the COVID-19 pandemic, which may have impacted how the Cath Lab schedules functioned during that time. As the hospital returns to its operation level before the pandemic, the procedure schedule may change as well. Additionally, as more data becomes available over time, the changeover time distribution may change as COVID-19 cleaning protocols related to ventilation may no longer be in place.

Nonetheless, the approaches to generating fasting instruction alerts in this study have the flexibility to be refined with new data and have the potential to adapt to changing trends. Other aspects of the context in which the study was conducted should also be considered. This study was conducted at a single-site in the cardiac catheterization laboratory setting. Further studies are required to confirm the effectiveness of the approaches we have evaluated in other similar settings that require pre-procedure fasting, such as peri-operative suites and diagnostic and interventional radiology departments.

Another limitation of our approach is the lack of use of more advanced ML models such as neural networks for the procedure duration prediction task. Although tree-based models are widely used in the literature and have achieved good results [34, 31, 20, 39], neural networks can model more complex relationships between variables and therefore potentially produce better predictions than the classical ML models implemented in this study. For neural networks, we can compute the features' Shapley values to gain some insights into how much each feature contributes to the model's predictions [41]. Future work includes further investigation and testing to determine the suitable neural architecture for procedure duration prediction.

Conclusion

In this study, we identified four approaches for generating fasting instruction alerts that would shorten pre-procedure fasting duration for Cath Lab patients in comparison to standard practice. We found that using scheduled start times and historical buffers had the best overall outcomes: this approach has the minimum total fasting duration while limiting the number of patients who did not meet the 2 and 6-hour fasting requirement. If implemented in a system that automates the delivery of these instructions to patients, the unnecessary discomfort and distress arising from symptoms associated with fasting could be reduced or avoided entirely. Utilizing the procedural and scheduling data available in the hospital's system, an implementation of our proposed approach has the potential to create a positive impact on patient experience and an improvement in the efficiency of healthcare delivery. Further studies are required to determine the effectiveness of operationalizing the proposed fasting alert system approach into practice.

Declaration of conflicting interests

The authors have no conflict of interest to disclose.

References

[1] American Society of Anesthesiology. Practice Guidelines for Preoperative Fasting and the Use of Pharmacologic Agents to Reduce the Risk of Pulmonary Aspiration: Application to Healthy Patients Undergoing Elective Procedures: An Updated Report by the American Society of Anesthesiologists Task Force on Preoperative Fasting and the Use of Pharmacologic Agents to Reduce the Risk of Pulmonary Aspiration*. *Anesthesiology* 2017; 126(3): 376–393. DOI:10.1097/ALN.0000000000001452. URL https://doi.org/10.1097/ALN.0000000000001452. https://pubs.asahq.org/anesthesiology/article-pdf/126/3/376/518874/20170300_0-00014.pdf.

[2] Dobson G, Chow L, Filteau L et al. Guidelines to the practice of anesthesia–revised edition 2020. *Canadian Journal of Anesthesia/Journal canadien d'anesthésie* 2020; 67(1): 64–99.

[3] de Aguilar-Nascimento JE, de Almeida Dias AL, Dock-Nascimento DB et al. Actual preoperative fasting time in brazilian hospitals: the bigfast multicenter study. *Therapeutics and clinical risk Management* 2014; 10: 107.

[4] Sorita A, Thongprayoon C, Ahmed A et al. Frequency and appropriateness of fasting orders in the hospital. In *Mayo Clinic Proceedings*, volume 90. Elsevier, pp. 1225–1232.

[5] Spitz D, Chaves G and Peres W. Impact of perioperative care on the post-operative recovery of women undergoing surgery for gynaecological tumours. *European Journal of Cancer Care* 2017; 26(6): e12512.

[6] Conway A, Chang K, Bittner M et al. Validating the peri-operative thirst discomfort scale for measuring thirst discomfort prior to procedures. *Journal of Radiology Nursing* 2021; 40(1): 75–79. DOI:https://doi.org/10.1016/j.jradnu.2020.10.006. URL https://www.sciencedirect.com/science/article/pii/S1546084320301814. A Culture of Safety in the Imaging Department.

[7] Arenas Moya D, Plascencia Gaitán A, Ornelas Camacho D et al. Hospital malnutrition related to fasting and underfeeding: is it an ethical issue? *Nutrition in Clinical Practice* 2016; 31(3): 316–324.

[8] El-Sharkawy A, Virdee A, Wahab A et al. Dehydration and clinical outcome in hospitalised older adults: a cohort study. *European Geriatric Medicine* 2017; 8(1): 22–29.

[9] Chan HY, Cheng A, Cheung SS et al. Association between dehydration on admission and postoperative complications in older persons undergoing orthopaedic surgery. *Journal of Clinical Nursing* 2018; 27(19-20): 3679–3686.

[10] Kyrtatos P, Constandinou N, Loizides S et al. Improved patient education facilitates adherence to preoperative fasting guidelines. *Journal of Perioperative Practice* 2014; 24(10): 228–231. DOI:10.1177/175045891402401003. URL https://doi.org/10.1177/175045891402401003.

[11] Rycroft-Malone J, Seers K, Crichton N et al. A pragmatic cluster randomised trial evaluating three implementation interventions. *Implementation Science* 2012; 7(1). DOI:10.1186/1748-5908-7-80. URL https://doi.org/10.1186/1748-5908-7-80.

[12] Cokelaer T, Kravchenko A, msat59 et al. cokelaer/fitter: v1.4.0, 2021. DOI:10.5281/zenodo.5394791. URL https://doi.org/10.5281/zenodo.5394791.

[13] Virtanen P, Gommers R, Oliphant TE et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 2020; 17: 261–272. DOI:10.1038/s41592-019-0686-2.

[14] Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 1996; 58(1): 267–288.

[15] Drucker H, Burges CJ, Kaufman L et al. Support vector regression machines. *Advances in neural information processing systems* 1996; 9.

[16] Loh WY. Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery* 2011; 1(1): 14–23.

[17] Breiman L. Random forests. *Machine learning* 2001; 45: 5–32.

[18] Friedman JH. Greedy function approximation: a gradient boosting machine. *Annals of statistics* 2001; : 1189–1232.

[19] Bravo F, Levi R, Ferrari LR et al. The nature and sources of variability in pediatric surgical case duration. *Paediatr Anaesth* 2015; 25(10): 999–1006. DOI:10.1111/pan.12709.

[20] Zabardast E. *Prediction of surgical operation durations using supervised machine learning techniques*. Master's Thesis, The Middle East Technical University, 2017.

[21] Martinez O, Martinez C, Parra CA et al. Machine learning for surgical time prediction. *Comput Methods Programs Biomed* 2021; 208: 106220. DOI:10.1016/j.cmpb.2021.106220.

[22] Barnoon S and Wolfe H. Scheduling a multiple operating room system: a simulation approach. *Health services research* 1968; 3(4): 272.

[23] Dexter F. Application of prediction levels to OR scheduling. *AORN J* 1996; 63(3): 607–615. DOI:10.1016/s0001-2092(06)63398-x. 8651672.

[24] May JH, Strum DP and Vargas LG. Fitting the lognormal distribution to surgical procedure times. *Decision Sciences* 2000; 31(1): 129–148.

[25] Robb DJ and Silver EA. Scheduling in a management context: uncertain processing times and non-regular performance measures. *Decision Sciences* 1993; 24(6): 1085–1108.

[26] Strum D, May J and Vargas L. Surgical procedure times are well modeled by the lognormal distribution. *Anesthesia & Analgesia* 1998; 86(2S): 47S.

[27] Stepaniak PS, Heij C, Mannaerts GH et al. Modeling procedure and surgical times for current procedural terminology-anesthesia-surgeon combinations and evaluation in terms of case-duration prediction and operating room efficiency: a multicenter study. *Anesthesia & Analgesia* 2009; 109(4): 1232–1245.

[28] Huang CC, Lai J, Cho DY et al. A Machine Learning Study to Improve Surgical Case Duration Prediction. *medRxiv* 2020; : 2020.06.10.20127910DOI:10.1101/2020.06.10.20127910. 2020.06.10.20127910.

[29] Ng N, Gabriel RA, McAuley J et al. Predicting Surgery Duration with Neural Heteroscedastic Regression. *arXiv* 2017; URL https://arxiv.org/abs/1702.05386v3. 1702.05386.

[30] Hosseini N, Sir MY, Jankowski C et al. Surgical duration estimation via data mining and predictive modeling: a case study. In *AMIA annual symposium proceedings*, volume 2015. American Medical Informatics Association, p. 640.

[31] Master N, Zhou Z, Miller D et al. Improving predictions of pediatric surgical durations with supervised learning. *Int J Data Sci Anal* 2017; 4(1): 35–52. DOI: 10.1007/s41060-017-0055-0.

[32] Edelman ER, van Kuijk SMJ, Hamaekers AEW et al. Improving the Prediction of Total Surgical Procedure Time Using Linear Regression Modeling. *Front Med* 2017; 0. DOI:10.3389/fmed.2017.00085.

[33] Schneider A, Wilhelm D, Schneider M et al. Laparoscopic cholecystectomy-a standardized routine laparoscopic procedure: is it possible to predict the duration of an operation? *Journal of Healthcare Engineering* 2011; 2(2): 259–269.

[34] Zhao B, Waterman RS, Urman RD et al. A machine learning approach to predicting case duration for robot-assisted surgery. *Journal of medical systems* 2019; 43(2): 32.

[35] Wu A, Weaver MJ, Heng MM et al. Predictive Model of Surgical Time for Revision Total Hip Arthroplasty. *J Arthroplasty* 2017; 32(7): 2214–2218. DOI:10.1016/ j.arth.2017.01.056.

[36] Wang J, Cabrera J, Tsui KL et al. Clinical and Nonclinical Effects on Operative Duration: Evidence from a Database on Thoracic Surgery. *J Healthcare Eng* 2020; 2020: 3582796. DOI:10.1155/2020/ 3582796.

[37] Davies SI. *Machine learning at the operating room of the future: A comparison of machine learning techniques applied to operating room scheduling*. PhD Thesis, Massachusetts Institute of Technology, 2004.

[38] Bartek MA, Saxena RC, Solomon S et al. Improving operating room efficiency: machine learning approach to predict case-time duration. *Journal of the American College of Surgeons* 2019; 229(4): 346–354.

[39] Master N, Scheinker D and Bambos N. Predicting pediatric surgical durations. *arXiv preprint arXiv:160504574* 2016; .

[40] Green S, Leroy P, Roback M et al. An international multidisciplinary consensus statement on fasting before procedural sedation in adults and children. *Anaesthesia* 2020; 75(3): 374–385.

[41] Molnar C. *Interpretable machine learning*. Lulu. com, 2020.

Appendix I Test Results of the ML Models

Table 4. Six ML models were selected and a GridSearch was performed for each model. Test results of ML models with the best result in bold.

Machine Learning Model	Hyperparameters	Best Parameters	5-fold CV RMSE (minutes)
Linear Regression	N/A	N/A	39.64
LASSO Regression	α = [0.01, 0.02 . . . , 1]	α = 0.01	39.64
Random Forest Regressor	max_depth: [1,2,3,4,5,6], max_features: [‘auto’, ‘log2’,‘sqrt’], min_samples_leaf: [3,4,5], min_samples_split: [8,9,10,11,12], n_estimator: [32, 64, 100, 128, 200, 256]	max_depth = 6, max_features = ’auto’, min_samples_leaf = 5, min_samples_split = 12, n_estimator = 128	38.1
Decision Tree Regressor	max_depth: [1,2,3,4,5,6], max_features: [‘auto’, ‘log2’,‘sqrt’], min_samples_leaf: [3,4,5], min_weight_fraction_leaf: [0,0.1,0.3,0.5], min_samples_split: [2,4,6,8], max_leaf_nodes: [3,5,7], splitter: [‘best’, ‘random’]	max_depth = 4, max_features = ‘auto’, max_leaf_nodes = 7, min_samples_leaf = 3, min_samples_split = 2, min_weight_fraction_leaf = 0, splitter = ‘best’	39.53
Gradient Boosting Regressor	Learning rate: [0.1, 0.2, 0.3, 0.4, 0.5], max_depth: [1,2,3,4,5,6], max_features: [0.2, 0.3, 0.4], min_samples_leaf: [3,4,5] min_samples_split:[8,9,10,11,12], n_estimators: [32, 64, 100, 128,200, 256]	Learning rate = 0.1 , max_depth = 5, max_features = 0.3, min_samples_leaf = 4, min_samples_split = 10, n_estimators = 64	**37.23**
Support Vector Regressor	kernel: [‘rbf’], C = [0.1,0.2,. . . ,10] , gamma: [1,2,. . . ,10]	C = 9.9, gamma = 1	51.45

Appendix II Sensitivity Analysis for Data Split

Table 5 shows the performance comparison among different splits of data. In the paper, we divided the dataset into a training set and a test set based on a specific date (i.e., October 25th, 2021) to be aligned with the real-life situation where we can only use historical data to predict the future. However, to test the robustness of our ML models to how the data is split, we conducted a sensitivity analysis with three additional training-test splits. Random Split 1 and 2 are two 70-30 splits with different seeds. The Odd vs. Even split represents a training set that only has odd months (52.2% of the data) while the test set contains data from even months (47.8% of the data).

Table 5. Sensitivity analysis for different data splits.

	Temporal Split	Random Split 1	Random Split 2	Odd vs. Even Month Split
RMSE of GBR (in minutes)	37.23	37.13	36.45	38.28
Median Fasting Duration (Clear Liquids) (in hours)	2.8	2.91	2.96	2.95
Underfasting Percentage	16.05%	11.83%	10.24%	11.48%

For the ML model for duration prediction, the performance across all four splits is quite similar. We also tested the different splits using the simulated start time with no buffer approach to compare the median fasting duration and the percentage of underfasting patients. For the fasting duration metric, the results among the splits are close to each other; for the underfasting percentage, however, we can see that the additional splits outperform the original temporal split used in the paper. Further analysis shows that a relatively small increase in the median fasting duration (e.g. 2.8 hours to 2.95) results in a non-trivial decrease in the underfasting percentage.

Appendix III Feature Importance of Models Tested

Table 6 presents the feature importance information for the ML models (except for SVR). The top three most important features are bolded. The procedure group features (e.g., Cardiac Resynchronization, Coronary Interventions) have a high importance in all models, while the expected duration feature has higher impact on the tree-based models.

Table 6. Feature importance of models tested (except for SVR).

	ML Model				
Feature Name	GRB	Linear Regression	LASSO	Decision Tree	Random Forest
Age	0.045	0.205	0.206	0.0	0.021
Height	0.035	0.021	0.024	0.0	0.015
Weight	0.037	0.047	0.048	0.0	0.01
Expected Duration	**0.345**	0.166	0.167	**0.535**	**0.465**
Proceduralist Case-Volume Cluster	0.035	2.473	2.466	0.0	0.011
Implantable Loop Recorder	0.0	-46.021	-18.449	0.0	0.0
Pacemaker	0.007	-21.075	-0.0	0.0	0.019
Defibrillator	0.004	-21.191	-0.0	0.0	0.015
Cardiac Resynchronization	**0.079**	**50.112**	**70.429**	0.0	0.001
Circulatory Support	0.0	0.0	0.0	0.0	0.0
Coronary Interventions	**0.086**	**14.961**	**35.853**	0.0	0.01
CATH	0.063	-35.715	-14.648	**0.202**	**0.187**
EPS	0.0	0.0	0.0	0.0	0.0
Biopsy	0.111	-56.546	-35.322	**0.263**	**0.207**
Structural Interventions	0.04	-1.352	19.539	0.0	0.0
Ablation	0.0	0.0	0.0	0.0	0.0
Complex Coronary Intervention	0.067	**116.826**	**135.019**	0.0	0.02
Gender_Female	0.003	-0.441	-0.781	0.0	0.001
Gender_Male	0.003	0.441	0.0	0.0	0.0
DayOfWeek_Mon	0.008	2.447	3.19	0.0	0.005
DayOfWeek_Tue	0.01	-0.795	0.0	0.0	0.001
DayOfWeek_Wed	0.004	2.814	3.595	0.0	0.002
DayOfWeek_Thur	0.013	-3.498	-2.663	0.0	0.009
DayOfWeek_Fri	0.003	-0.969	-0.101	0.0	0.002

Appendix IV Details of the Hourly Approaches

The simulation approach considered the start time prediction at one time-point. During the day, a new schedule is produced hourly, providing an opportunity to incorporate changes due to late arrivals or emergency procedures. Recalculating the fasting alert times using these updated schedules may lead to better performance since they better represent the anticipated schedules.

We tested two scenarios that utilize snapshots published between 6 a.m. and 17 p.m.: i) All Snapshots (AS): we use the newly published snapshots to calculate the fasting alert time; ii) Actual (Actl): we use the actual start times of the procedures that have started before the time the snapshot is generated.

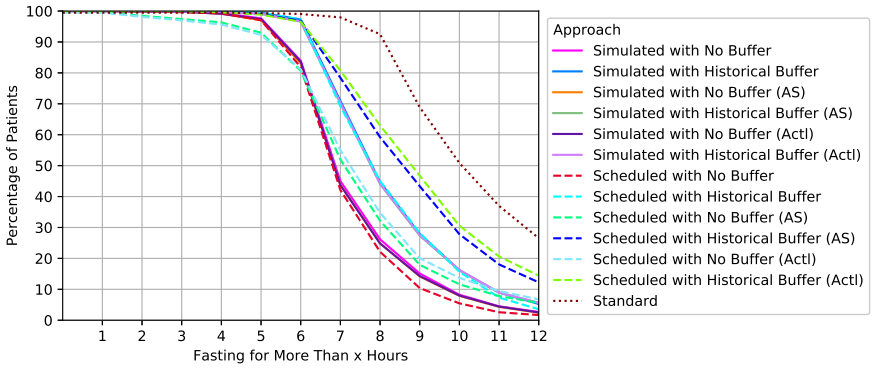

Figure 7. Percentage of patients fasting for more than x hours for solid food intake by all tested approaches.

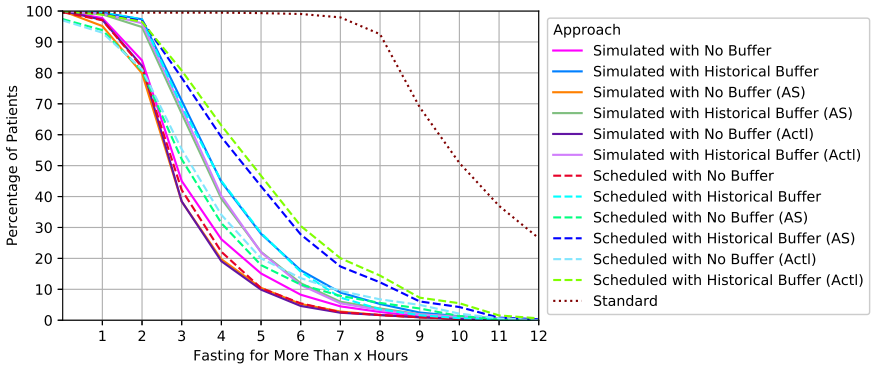

Figure 8. Percentage of patients fasting for more than x hours for clear liquids intake by all tested approaches.

Figures 7 and 8 show the performance of all strategies tested for solid food and clear liquids fasting. The results indicate that the hourly approaches have nearly identical performance as the single-schedule approaches presented in this study.

Appendix V Procedure Duration Means and Standard Deviations

Table 7 shows the procedure duration means and standard deviations for the different procedure groups. We can see that there is considerable difference in the durations and the standard deviation of each group is quite high, between 30% and 60% of the mean.

Table 7. Procedure duration means and standard deviations for different procedure groups.

Procedure Group	Procedure Duration Mean (minutes)	Procedure Duration Standard Deviation (minutes)
Biopsy	30.81	15.67
CATH	61.59	33.09
Cardiac Resynchronization	174.97	58.86
Complex Coronary Intervention	233.33	99.34
Coronary Intervention	115.66	70.53
Defibrillator	78.24	43.33
Implantable Loop Recorder	39.44	17.13
Pacemaker	77.07	41.26
Structural Intervention	89.55	47.86