Improving Planned Surgical Case Duration Accuracy by Leveraging the EHR and Predictive Modeling – A Randomized Control Trial

PROTOCOL FACE PAGE FOR
MSK NON THERAPEUTIC PROTOCOL

Principal Investigator/Department:	Christopher Stromblad, MS	Surgery, Strategy & Innovation
Co-Principal Investigator(s)/Department:	Roger Wilson, MD Nadeem Abu-Rustum, MD Julio Garcia-Aguilar, MD	Surgery Surgery/Gynecology Surgery/Colorectal
Investigator(s)/Department:	Marcia Levine, RN Shok-Jean Yee, RN Mary-Ann Connor, RN Stanley John Ryan Baxter-King Nick Kastango Josh Kunken Daniel Stein, MD, PhD Aaron Ostrovsky Alexia Iasonos, PhD	Nursing/Periop Nursing/Periop Nursing Informatics Nursing Informatics Surgery, Strategy & Innovation Strategy & Innovation Strategy & Innovation Health Informatics Health Informatics Epidemiology-Biostatistics
Consenting Professional(s)/Department:	Not Applicable	

Please Note: A Consenting Professional must have completed the mandatory Human Subjects Education and Certification Program.

OneMSK Sites
Manhattan

Memorial Sloan Kettering Cancer Center
1275 York Avenue
New York, New York 10065

Date:

Table of Contents

Date:

1.0 PROTOCOL SUMMARY AND/OR SCHEMA

Title: Improving Planned Surgical Case Duration Accuracy by Leveraging the EHR and Predictive Modeling – A Randomized Control Trial

Objectives: The objective is to study an effort to improve surgical case duration predictions. A predictive model has been developed for this purpose and shows an improvement in case duration accuracy of 25% through a retrospective review without actually using or implementing the output from the predictive model as the planned duration for the surgery. This study aims to test, in a real world setting, how an already developed predictive model will work compared to the current process by implementing the predictive model output as the planned duration value.

Human subjects: The primary subjects in this study is the OR team (i.e. surgeon, nursing, technical support, etc.). We are studying the duration it takes surgeons to complete their respective surgical cases. All Gynecology (GYN) and Colorectal (CRS) Surgeons at MSKCC will be included. Since it is not feasible to obtain consent from the entire OR team for each case, a waiver of consent and authorization is requested. The secondary subjects in this study are our patients as we are interacting with their PHI for research purposes. There will be no intervention with the patients and therefore, a waiver of consent and authorization is requested for patients and the entire OR team, justification is outline in Section 6.0, Recruitment.

Design: The intent is for GYN and CRS surgeons and other OR staff to perform their surgeries without being directly affected by the planned duration value. We propose a randomized control trial where the cases will be equally distributed to a control group and an intervention group. The surgeon, OR staff, and patient will not know which cases' duration was calculated using the predictive model or the current default calculation, and thus will not be able to, intentionally or unintentionally, skew the actual duration towards any bias. Furthermore, we will be able to compare the two groups over the same period of time mitigating any potential impact of trends or seasonality effects on outcome. We will control for surgeons and for surgical sites as each may consistently have some unique characteristics that may cause longer or shorter case durations.

Time to completion: About 14 weeks.

2.0 OBJECTIVES AND SCIENTIFIC AIMS

The main objective is to test if the predictive machine learning model, which was developed by Surgery and Strategy Analytics, can improve our overall operating room scheduling accuracy by implementing it for two services and evaluating the results in a controlled setting. The current process requires the scheduling manager to select between a default calculation process embedded in the electronic health record software (EPIC), a surgeon estimate, or provide her own estimate. The chief reason for this elaborate process is that the EPIC calculation is often not considered to be very accurate, and surgeons will at times ask that their estimate be used instead. At times there is no surgeon

estimate, nor a reasonable EPIC estimate, and the scheduling manager will provide a reasonable estimate. The developed predictive model has been tested using retrospective data and this study aims to test the model in an actual scheduling setting to ensure that all the input data presumed to be available is available, and to ensure that the model is robust in a real world setting. If the predictive model improves the scheduling accuracy it will enable us to anticipate the case work load more accurately, and we will be able to better allocate appropriate resources in advance. Higher level performance metrics may be impacted by the predictions and the indirect benefits and risks may include effects on delays, room idle time, utilization and extended operating room hours.

Specific Aim 1: To test the hypothesis that the developed surgical case duration prediction model compared to the current process of estimating surgical case durations, will show improved prediction accuracy, measured by mean absolute error.

Specific Aim 2: To assess the effectiveness of the predictive model across the different surgical sites, surgeons, and the two services, Gynecology and Colorectal at MSKCC.

Specific Aim 3: To evaluate the effect on the to-follow surgical case by using the developed prediction model as compared to the current process. Including to which extent underestimation of the case duration by the prediction model results in a delay on the following case and to which extent an overestimation of the case duration by the prediction model results in increased ‘room empty time’ before the following case actually begins.

3.0 BACKGROUND AND RATIONALE

Uncertainty is a natural component of hospital operations. Surgery, in particular, suffers from a high rate of variability, due to many factors such as procedure combination, surgeon performing the procedures, Relative Value Unit (RVU) of the procedures, surgical location, patient characteristics, unforeseen complications, and several other medical and operational elements. Operating Rooms (ORs) are considered as the main initial treatment option at MSK for new patients and the greatest source of revenue and cost for hospitals in general. As the most resource-intensive and high-demand section of the hospitals, efficient ORs are crucial in delivering hospitals' operational goals, surgeon and staff satisfaction, and better patient outcome and experience. To achieve higher performance in surgical rooms and to mitigate waiting and idling times for both patients and staff, we require reliable estimation of case duration. In 2016, 63% of Gynecology and Colorectal cases were predicted within 60 minutes accuracy and the mean absolute error was 66 minutes.

In a previous retrospective research study (IRB Protocol #: 16-1657) a prediction model was developed to address case duration prediction problem by leveraging the entire patient-hospital data available prior to the surgery. The performance of the model was tested using retrospective data and showed a relative improvement of 17% (from 63% to 74%) in predicting within 60 min and an improvement of 25% (from 66.4 minutes to 49.7 minutes) in mean absolute error.

This protocol focuses on the implementation aspects of the prediction model and the effect of applying advanced predictive modeling daily in a controlled setting. We propose a randomized control trial in order to prevent bias in planning and the completion of surgeries and to provide a better comparison group than a retrospective review; where one would risk ignoring duration trends or swings over the past years. The Gynecology and Colorectal surgeons will complete the surgeries without explicitly knowing if the planned duration was derived from the current process or the predictive model. If the expected improvements are evident from this study, the future work will look to integrate advanced predictive models across multiple surgical services and realize a broader impact across all surgical platforms in terms of planning, efficiency, patient experience and staff satisfaction.

4.0 OVERVIEW OF STUDY DESIGN/INTERVENTION

4.1 Design

The study is designed for each case to be randomly selected to one of two arms: (1) the control arm where each case will follow the standard scheduling process where the planned case duration value is derived using the current process (the default calculation from EPIC or estimate provided by a surgeon or scheduling office); (2) the intervention arm where each case is assigned a planned case duration value from the predictive model one day prior to surgery. Steps 1 and 2 explain the current scheduling process used at MSKCC. A case flow diagram can be found in Appendix A.

1. Case is scheduled from GYN or CRS clinic, and a case duration will automatically be assigned to the case, this value may at times be zero if the specific procedure combination did not occur in the recent past.
2. Scheduling office assigns start time and room for case and places case on schedule. At this point a default case duration is evaluated by the scheduling office, to see if the value is considered excessively short or excessively long. Depending on the assessment, the scheduling office will either keep the default value, use the value that the surgeon placed in the notes (if available), or the scheduling office provides their own estimation.
3. Two midnights before day T (day of surgery for relevant cases), all GYN and CRS cases will be randomized and land in one of the following groups:
   - **3.1.** Case randomized into control group:
     - **3.1.1.** Case duration will remain with the value derived from step 2.
   - **3.2.** Case randomized into intervention group:
     - **3.2.1.** Predictive model calculates new duration for case at 3AM the day before surgery, and the predictions are made available on a SecureShare-site. Model predictions are then read by scheduling manager sometime between 7am-10am from the SecureShare site, and the scheduling manager will in EPIC/OpTime, overwrite the current estimate with the new duration value that was generated by the predictive model.
4. Patient arrives day of surgery and the surgery will take place as planned.

4.2 Intervention

The intervention involves randomizing 50-50 to a control and an experimental arm. Patients in the control arm will follow the standard scheduling process where the planned case duration value is derived using the current process while patients in the experimental arm are expected to follow a model that will predict the planned case duration during the study period. The modified planned case duration is derived from a greater source of data with a more advanced predictive modeling approach and is hypothesized to significantly improve case duration accuracy and help with overall planning. A retrospective analysis using the predictive model has shown an improvement in case duration accuracy by 25%. The intervention is expected to be of minimal risk to the patient involved in the case. The surgeon and all other OR staff are expected to complete the surgery as she/he otherwise would.

We require both a control arm and an intervention arm for this study. This enables us to test the effectiveness of the model over the exact same time period so that trends and seasons apply similarly to both study arms. Furthermore, we mask which cases will be using the predictive model so that we can expect minimal to no associated behavior change by the surgeon or surgical staff to affect the case duration.

5.0 CRITERIA FOR SUBJECT ELIGIBILITY

Inclusion and Exclusion criteria will be built into the predictive model up front, and cases will be excluded as described below.

5.1 Subject Inclusion Criteria

- A surgeon or OR staff member in the Department of Surgery Gynecology and Colorectal service

5.2 Subject Exclusion Criteria

- Any new surgeon that starts their practice during the study
- Surgery will take place at a location other than the Main hospital or Josie Robertson Surgical Center
- Cases where input data was not available prior to the prediction generation including late add-on cases such as urgent and emergent cases that are placed on the schedule less than 24 hours before the surgery

6.0 RECRUITMENT PLAN

Surgeons and OR staff from the GYN and CRS services will be notified of this protocol, its procedures, and its importance, prior to initiation of the study. Staff will be notified that the first 568 cases that do not meet exclusion criteria will be used for this trial.

We are waiving the consent and authorization for patients. Justification for the waiver of consent/authorization is as follows:

- The research involves no more than minimal risk to the participants as will only be interacting with their PHI for predicting the case duration and for scheduling the OR case
- The waiver or alteration will not adversely affect the rights and welfare of the research participants
- The research could not be practicably be carried out without the waiver or alteration as we plan on carrying this research out on 568 cases and it is not feasible to obtain consent for the patient as the aim/intervention does not involve them outside of interaction with their PHI for surgical identification purposes
- The research is not regulated by the FDA.

For the surgeons and OR team in the Gynecology and the Colorectal services, we are waiving the consent and authorization. Justification for the waiver of consent/authorization is as follows:

- The research involves no more than minimal risk to the participants as we will not be carrying out an intervention on them or interacting with any of their PHI for the purposes of this research
- The waiver or alteration will not adversely affect the rights and welfare of the research participants
- The research could not be practicably be carried out without the waiver or alteration as we plan on carrying this research out on 568 cases and it is not feasible to obtain consent for the entire OR team who is being observed
- The research is not regulated by the FDA.

7.0 ASSESSMENT/EVALUATION PLAN

The first 568 cases that do not meet exclusion criteria will be used for this trial, with an estimated accrual time of 14 weeks. With approximately 62 cases per week and the possibility for excluding cases and removing cases, it should give us over the 568 cases needed.

The primary metric we will monitor and evaluate is the mean absolute error between planned and actual case duration in each of the study arms.

We will monitor the number of cases that will be excluded from the study and try to minimize this by coordinating with the service chiefs

We will monitor the # of cases within 15min, 30min, and 60min accuracy as secondary metrics

We will monitor the difference between the planned start time for the following case and the actual start time for the following case (start time defined as when the patient enters the operating room, “toes-in”)

Date:

We will monitor the time between toes-in for the following case and toes-out for the previous case, “turnaround” associated with the case

8.0 TOXICITIES/SIDE EFFECTS

Not applicable.

9.0 PRIMARY OUTCOMES

The primary metric (supporting specific aim 1) for the study is mean absolute error; which serves as an unweighted estimate of accuracy that has operational significance. Any of the below metrics will be investigated by service, site, and surgeon to address specific aim 2. To review of the impact on the next case we establish the metrics start delay for the next case, and turnaround for the next case; they will support specific aim 3.

Metric	Calculation	Control group	Intervention group
Mean absolute error (MAE)	Mean(|Planned – Actual|)		
Mean error	Mean(Planned – Actual)		
Root Mean squared error	Sqrt(Mean((Planned – Actual)^2))		
% cases with MAE <= 15min	# cases with MAE <= 15 min / # of cases		
% cases with MAE <= 30min	# cases with MAE <= 30 min / # of cases		
% cases with MAE <= 60min	# cases with MAE <= 60 min / # of cases		
% cases with MAE <= 3 sigma	# cases with MAE <= 3 sigma/ # of cases		
% cases with MAE <= 25% of actual case duration	# cases with MAE <= 25% of actual case duration/ # of cases		
Start delay for the next case	Actual start time - Planned start time (for next case)		
Turnaround for the next case	Toes in for next case – toes out for previous case		

10.0 CRITERIA FOR REMOVAL FROM STUDY

If some surgeons do not comply with the protocol and for example insists on overwriting the duration value for a case that was in the intervention arm, we will have to exclude those cases from the study. The compliance to the protocol will be addressed weekly in order to minimize the number of cases that we remove from the study.

11.0 BIOSTATISTICS

Specific Aim 1: To test the hypothesis that the developed surgical case duration prediction model compared to the current process of estimating surgical case durations, will show improved prediction accuracy, measured by mean absolute error.

The study is designed for each case to be randomly selected to one of two arms: (1) the control arm where each case will follow the standard scheduling process where the planned case duration value is derived using the current process (the default calculation from EPIC or estimate provided by a surgeon or scheduling office); (2) the intervention

arm where each case is assigned a planned case duration value from the predictive model one day prior to surgery.

The primary endpoint is the absolute mean error defined as the prediction from the model (arm 2) or EPIC combined with manual input (arm 1) minus the actual case duration. Preliminary data were obtained in a period of time where the prediction model was evaluated in a retrospective testing of the model over the time period May 2016 – September 2017 for all surgical cases that were either a Gynecology case or a Colorectal case. Data from a retrospective analysis of Gynecology and Colorectal surgical cases with the predictive model and the current process is shown in the table below:

	CRS	CRS
	Model	Current process
Arithmetic mean of absolute error	56.75	81.92
Standard deviation of absolute error	65.46	88.43
Number of cases (n)	446	446

	GYN	GYN
	Model	Current process
Arithmetic mean of absolute error	46.45	57.75
Standard deviation of absolute error	48.81	63.10
Number of cases (n)	754	754

	Experimental/Model	control
Weighted arithmetic mean of absolute error	50	67
Weighted standard deviation of absolute error	56	73

Date:

The power below provides a test for the differences in 2 sample means of 67 and 50 with sd as shown above.

Assuming 90% power, Type I error=10%, 2 sided test of means for 2 samples/arms, we need 568 cases in total (284 per arm). We expect that the same patient might be enrolled more than once on this protocol so the patient level data will not be independent. Based on our review of pilot data this might occur 3-4% of the time among GYN service and up to 15% in the colorectal cases. This might have an affect the power depending on the within patient correlation. The primary analysis will include patients with a single case (the earliest case in time) and exclude the secondary surgery from the same patients. A secondary analysis will serve as a sensitivity analysis will include all cases thus including multiple cases per patient. We will accrue enough cases so that when excluding secondary surgeries we will have 568 cases.

Specific Aim 2: To assess the effectiveness of the predictive model across the different surgical sites, surgeons, and the two services, Gynecology and Colorectal at MSKCC.

A subset analysis will be done within service (Gyn or CRS) and site (2 sites: main and, JR) to assess if there are differences in the error between the prediction and actual case duration by site, service and surgeon. This analysis will be descriptive and the aim is to identify outliers or source of biases that would lead to the model or the current process to be under or over estimating case duration compared to the actual time. No modeling will be performed as this analysis is exploratory.

Specific Aim 3: To evaluate the effect on the immediate following surgical case by using the developed prediction model as compared to the current process. Including to which extent underestimation of the case duration by the prediction model results in a delay on the following case and to which extent an overestimation of the case duration by the prediction model results in increased ‘room empty time’ before the following case actually begins.

We will first identify all cases regardless of the assigned arm where underestimation of the case duration was done either by EPIC (arm 1) or the model (arm 2) defined as actual time > planned time. We will do a similar procedure for cases that were overestimated. We will assess whether the following case was delayed as a result of underestimation of the previous case and whether the percentages of these delays were different in the two arms. Similarly, we will assess whether there was significant ‘room empty time’ before the next case. We will use 2x2 contingency tables to tabulate these (arm 1 vs 2, against delay: yes/no) among the subset of cases that were underestimated. We will also do a 2x2 contingency tables to tabulate these (arm 1 vs 2, against ‘room empty time’: yes/no) among the subset of cases that were overestimated.

12.0 RESEARCH PARTICIPANT REGISTRATION AND RANDOMIZATION PROCEDURES

12.1 Research Participant Registration

Not applicable.

12.2 Randomization

Cases will be randomized (1:1) to one of 2 arms. Arm 1 is the control arm where each case will follow the standard scheduling process where the planned case duration value is derived using the current process (the default calculation from EPIC or estimate provided by a surgeon or scheduling office); Arm (2) is the intervention arm where each case is assigned a planned case duration value from the predictive model one day prior to surgery.

After all eligibility criteria are established, cases will be randomized two midnights before the scheduled surgery date. Biostatistical representatives will work with Strategy and Health Informatics to ensure appropriate randomization, timing, and integration to the workflow. Randomization will be accomplished by the method of random permuted block, and cases will be stratified by site (main hospital, Josie Robertson) and surgeon (13 Gyn surgeons operate at both, plus 8 CRS surgeons at main). Not all surgeons operate at all sites so the available number of strata is 34.

13.0 DATA MANAGEMENT ISSUES

A detailed description of the data to be collected.

- A variety of patient-level data from the Institutional Database (IDB) including:
  - Patient characteristics
    - Age, race, sex, height, weight, BMI
  - Previous treatment at MSK
    - Tests, surgeries, treatments (e.g. chemo, radonc)
  - Patient diagnoses and comorbidities
  - Information about the posted surgery
    - Planned procedures
    - Date and time of the surgery
    - Number of other cases being performed on that day / in that room
  - Information about the attending surgeons for the upcoming surgery
    - Information about surgeons' past cases
- **Process of data collection**
  - There will be no RSA assigned to this study. All data management procedures will be performed by the PI or an analyst from the strategy & innovation department listed on the facesheet.
  - Data will be collected via SQL from the IDB server, the Darwin server and DHADTQUANTSTRAT server
  - Training: Query to gather training data manually run to generate model
  - Prediction: Automated script will run every morning to score cases for the following weekday
- **For data collection and storage we have set up a MS SQL Server database.**
  - Training Data, used to build the prediction model to be stored on the DHADTPERIOPCDA database

Date:

- Prediction data, that includes both inputs and outputs to be stored on the DHADPPERIOPCDA database
- Data made available to Scheduling Office through a SecureShare site only available for the investigators of in this protocol.

- **Data management resources should be addressed if the study is considered high profile, fast accruing and/or time consuming based on the stated objectives**
  - Health informatics will participate in sharing prediction data with the scheduling office. Otherwise, the data will be managed internally by Strategy Analytics

13.1 Quality Assurance

- **Protocol compliance:**
  - If there are cases that are randomized into the intervention arm, and the prediction generated by the model is not used, we will monitor this, understand the underlying reason, and address it on a weekly or daily basis, depending on the issue
    - A suspected reason that this may happen, could be that the surgeon's office calls to change the duration and insists that this is done. We will note the reason for this and evaluate at the end of the week if we need to ask the Chief of the respective service to address this issue.
    - If it is due to a modeling error, we will evaluate this daily and address as soon as possible.
- **Eligibility verification:**
  - In the randomization process we will only include the cases that are planned to be performed at Main Campus or at Josie Robertson Surgical Center where the primary surgeon is either from Gynecology or the Colorectal service
- **Informed consent procedure**
  - Verbal
- **Data accuracy**
  - Quality of Input Features
    - Missingness checks
  - Quality of Input Observations
    - Waterfall analysis of cases are making it into the study (those cases intentionally left out vs accidentally left out including add-ons)
  - *Quality of Output*
    - Checking if scheduling office successfully changed time correctly

13.2 Data and Safety Monitoring

1) **How trials are monitored, including frequency and data elements to be reviewed**

Date:

a. We will review the implementation closely in the first two weeks, after which we will do a weekly review to ensure that the correct cases are successfully receiving a prediction.

2) Plans for assuring compliance regarding adverse event reporting

a. Not applicable

3) **Plans for assuring appropriate action if a monitored trial results in a temporary or permanent suspension**

a. It will be simple to terminate the study if need be. The implementation team will be informed and we will stop sending predictions, and ask the scheduler to resume the workflow that was present prior to the study.

4) Plans for assuring data accuracy and protocol compliance

a. Through weekly review of case scheduling, we will be able to determine if case durations are overwritten or if there is an excess of cases that are scheduled late. We will review those cases and reach out to the service chief to enforce better protocol compliance.

5) Description of the process and means to avert potential conflicts of interest.

a. Not applicable

14.0 PROTECTION OF HUMAN SUBJECTS

We will protect the rights of human subjects by ensuring all the PHI data that is accessed and applied as input to the prediction model will be stored at an MSK secure server maintained by information security and strategy analytics.

14.1 Privacy

MSK's Privacy Office may allow the use and disclosure of protected health information pursuant to a completed and signed Research Authorization form. The use and disclosure of protected health information will be limited to the individuals described in the Research Authorization form. A Research Authorization form must be completed by the Principal Investigator and approved by the IRB and Privacy Board (IRB/PB).

Any results from this study will be shared in the aggregate. Though PHI will be accessed by the prediction model and used as inputs, no PHI will be disclosed in this study.

14.2 Serious Adverse Event (SAE) Reporting

Not applicable

14.2.1

Not applicable

15.0 INFORMED CONSENT PROCEDURES

The consent and research authorization for conduct of this study will be waived as per 45 CFR 46.116(d) and 45 CFR 164.512(i)(2)(ii). Justification for the waiver is written under the Recruitment section

Date:

16.0 REFERENCES

Not applicable

17.0 APPENDICES

Appendix A: Design of randomized control trial