Using Type IV Pearson distribution to calculate the probabilities of under- and over-run of lists of multiple cases

Short Title: Under- and over-run probability estimation of lists of multiple cases

1. **Jihan Wang**
   - **Affiliation:** Department of Industrial and System Engineering, Wayne State University
   - **Email:** Jihan.Wang@wayne.edu
   - **Role:** This author helped design the study, conduct the study, collect and analyze the data, and write the manuscript
   - **Conflicts:** Jihan Wang reports no conflicts of interest
   - **Attestation:** Jihan Wang approved the final manuscript

2. **Kai Yang**
   - **Affiliation:** Department of Industrial and Systems Engineering, Wayne State University
   - **Email:** Kai.Yang@wayne.edu
   - **Role:** This author helped analyze the data and write the manuscript
   - **Conflict:** Kai Yang reports no conflicts of interest
   - **Attestation:** Kai Yang approved the final manuscript

**Institution:** John D. Dingell VA Medical Center

**Corresponding Author:**
Jihan Wang, MS
Department of Industrial and Systems Engineering
Wayne State University
4815 Fourth Street, Room 2033
Detroit, MI 48202
Phone: +1 313-577-0864
Fax: +1 313-577-8833
Email: Jihan.Wang@wayne.edu

**Reprints** will not be available from the authors
ABSTRACT

CONTEXT: An efficient operating room (OR) theater has both little under- and over-utilized OR time to achieve optimal OR cost efficiency. The probabilities of under- and over-run of lists of cases can be estimated by a well-defined duration distribution of the lists.

OBJECTIVE: To propose a method of predicting the probabilities of under- and over-run of lists of cases using Type IV Pearson distribution to support case scheduling.

DESIGN: Six years of data were collected. The first five years of data were used to fit distributions and estimate parameters. The data from the last year was used as testing data to validate the proposed methods. The percentiles of the duration distribution of lists of cases were calculated by Type IV Pearson distribution and t-distribution. Monte Carlo simulation was conducted to verify the accuracy of percentiles defined by the proposed methods.

SETTING: ORs in John D. Dingell VA Medical Center, United States, from January 2005 to December 2011.

PATIENTS OR OTHER PARTICIPANTS: None

INTERVENTIONS: None

MAIN OUTCOME MEASURES: Differences between the proportion of lists of cases that was completed within the percentiles of the proposed duration distribution of the lists and the corresponding percentiles.

RESULTS: Compared to the t-distribution, the proposed new distribution is 8.31% ± 0.38% (mean ± standard error) more accurate on average and 14.16% ± 0.19% more accurate in calculating the probabilities at the 10th and 90th percentiles of the distribution, which is of major concern of OR schedulers. The absolute deviations between the percentiles defined by Type IV Pearson distribution and those from Monte Carlo simulation varied from 0.20 ± 0.01 minutes to 0.43 ± 0.03 minutes. OR schedulers can rely on the most recent 10 with the same combination of surgeon and procedure(s) for distribution parameter estimation to plan lists of cases.

CONCLUSIONS: The proposed Type IV Pearson distribution is more accurate than t-distribution to estimate the probabilities of under- and over-run of lists of cases. However, since not all the individual case durations followed log-normal distributions, there was some deviation from the true duration distribution of the lists.
TRIAL REGISTRATION: None

KEYWORDS: health services research, operating room management, OR efficiency, OR scheduling
INTRODUCTION

The operating room (OR) is one of the most expensive units for hospitals. To control costs, it is important for OR management to plan the lists of cases in such a way that the allocated OR time is utilized as much as possible with little over-utilized OR time, i.e. maximizing the efficiency of use of OR time. A list consists of cases taking places in the same OR on the same day. When the duration of a list of cases (i.e. the OR workload) is greater than the allocated OR time, over-utilized OR time is observed. Too much over-utilized OR time may cause case cancellations. On the other hand, when the duration is smaller than the allocated OR time, under-utilized OR time is seen and OR capacity is wasted. When surgeons can schedule elective cases on any workdays, the maximum efficiency of use of OR time can be achieved by predicting the future OR workload and determining the optimal OR allocation. In most U.S. OR suites, the decisions on OR allocation are made every two to three months. For such a system, case duration prediction accuracy is less of a problem because the decisions on OR allocation take into consideration case duration prediction inaccuracy, as the actual workload is used for calculation. With more than two historical cases of the same combination of surgeon and procedure(s), the reduction in over-utilized OR time is negligible when people use a more accurate case duration prediction model rather than using the mean case durations of historical cases to assign scheduled case durations. People work late because of the workload rather than underestimation in case durations.

However, the OR allocation optimization approach does not apply any more to European OR suites. In these facilities, the demand for surgery is so high that patients have to enter the waiting list first and wait for months before their surgeries can be scheduled. In addition, the allocated OR time is usually static, and OR managers do not usually change the OR allocation. The way to maximize the efficiency of use of OR time is to avoid both under- and over-utilized OR time through scheduling versus decisions on OR allocation. A method to accurately estimate the durations of lists of cases is needed for this type of system.

Much research has been done in predicting the case duration of single cases. However, even with very complicated models, the improvements in case duration prediction accuracy are not significantly better than simpler methods to
reduce the tardiness of case starts.\textsuperscript{14,17,18,19} More accurate models only improve the prediction on the central tendency (i.e., the mean or median of case duration).\textsuperscript{18} As there is high variability associated with case duration,\textsuperscript{20} such improvements make limited differences in practice. Especially when there are multiple cases in a list, the cumulative effects of the high variability make it even more difficult to predict the duration of the list precisely.

The expected duration of a list of cases can be derived using the mean case durations of historical cases. However, the high variability of case durations makes the expected close time of the OR not a reliable estimate. It is not uncommon to see a list of cases under- or over-run by more than an hour. It is more meaningful for OR managers to estimate how reliably a list of cases can be completed within a pre-defined range with some tolerances in both under- and over-utilized OR time. For example, the allocated OR time is 8 hrs, and the OR manager wants to know how certain it is that a scheduled list of cases can be finished in between 7 hrs and 8.5 hrs. If it is highly likely, then the list can be finalized because it will not cause a large amount of under- or over-utilized OR time. Otherwise, the OR manager can rearrange the cases on the list so that both under- and over-utilized OR time will not be too much. This problem has barely been studied. Dexter et al.\textsuperscript{21} built an optimization model and concluded that using mean case durations of historical cases to schedule lists of cases actually generates the optimal efficiency of use of OR time. However, the problem Dexter et al. studied is different from ours because they assumed that surgeons had open access to the OR and the allocation of OR time had been optimized to maximize the efficiency of use of OR time. As explained above, this is not applicable for European systems. Alvarez et al.\textsuperscript{22} explored whether the second tertile cut-off point was better than the sum of mean case durations of historical cases in predicting over-utilization. They reached the same conclusion as Dexter et al. that mean case duration of historical cases is a good estimate for case scheduling. They only studied cardiovascular surgeries, and the results cannot be generalized to other specialties because case variability and duration distribution are different. In 2011, Pandit and Tavare\textsuperscript{10} proposed a scheduling algorithm that plans a list of cases based on the probability of the list being completed within pre-defined limits on under- and over-utilized OR time. This approach was much better than the ad-hoc scheduling approach in their facility. Although this approach generated promising results, the assumption that
the duration of a list of cases follows a t-distribution might not be valid because previous research have shown that individual case duration actually follows a log-normal distribution.\textsuperscript{11,15} Thus, the sum of log-normally distributed case durations does not follow a Gaussian distribution. The probabilities of under- and over-run of a list of cases calculated from t-distribution would deviate significantly from the true values. As a result, the conclusion drawn from the current science is not accurate and may cause inappropriate scheduling decision making.

In this study, we propose a new approach to approximate the probability of under- and over-run of lists of cases. First, we tested the hypothesis that individual case duration was log-normally distributed. Then, a new distribution was introduced to approximate the true duration distribution of lists of cases. We checked the accuracy of the proposed method using real data from one year’s lists of multiple cases.\textsuperscript{12,23,24} We compared the results from the proposed distribution with those generated by t-distribution as discussed by previous study.\textsuperscript{10} After this, we identified the optimal number of previous cases for the same combination of surgeon and procedure(s) to be included to derive reliable estimates of probabilities of under- and over-run.

**METHODS**

Among the surgeons who were working in 2011, some surgeons had worked for the studied facility since 2005. Thus, in order to include complete information, we extracted data all the way back to 2005. Data was collected for all surgeries performed from January 1, 2005, to December 31, 2011, in John D. Dingell VA Medical Center located in Detroit, MI.

The analysis was done in two phases. As our proposed method depended upon the fact that individual case duration within a list of cases followed a log-normal distribution, we first checked the validity of this log-normal assumption of individual case duration. Then, we introduced a new distribution to approximate the duration distributions of lists of cases so that the OR scheduler can rely to estimate the probability of under- and over-run of lists of cases. Data from January 1, 2005, to December 31, 2010, was used to estimate the case duration distribution parameters. The most recent data from January 1, 2011, to December 31, 2011 acted as testing data set to evaluate the performance of the proposed method.
Log-normality Tests of Individual Case Duration Distribution

From January 1, 2005, to December 31, 2010, there were 15646 cases, 0.8% (121 of 15646) of which had incomplete information of case duration. We excluded these cases from analysis. There were 2994 different combinations of surgeon and procedure(s). Previous studies show that type of anesthesia also is an important factor that impacts the case duration; however, when surgeons scheduled cases, they did not consult anesthesiologists and did not always know the type of anesthesia to be used, making this information incomplete and unreliable. Thus, we did not consider the type of anesthesia in our studies to group cases. This would not impact the analysis as the influence of anesthesia was reflected in the actual case duration. The variability in case duration caused by anesthesia was accounted by the variability of the fitted distribution.

Of the 2994 combinations of surgeon and procedure(s), 127 (7496 cases) had moderate to large sample sizes (n≥20). We used the case durations of these combinations to test whether individual case duration was log-normally distributed. We calculated the natural logarithms of the case durations for each combination. Then, Shapiro-Wilk tests were conducted for the log-transferred data sets. If the P-value of the test was greater than 0.05, then we failed to reject that the case durations of the combination followed a log-normal distribution. The analysis was done using R 2.15.0.

Duration Distribution of Lists of Multiple Cases

In 2011, there were 1463 lists of cases. The captured data was not able to identify when a case was moved to another OR. We could rely only on the final lists of cases that recorded the actual locations and times of completed cases versus the lists on the original schedules. As the methods to calculate the percentiles of single case duration distribution had been examined with high accuracy, and the focus of our study was on the durations of lists of multiple cases, we excluded 37% (547 of 1463) of the lists containing only one case. Among the remaining lists, 44% (401 of 916) consisted of at least one case had no or only one historical case. This made the estimation of case duration variance infeasible. We excluded these lists as well, leaving 515 lists for analysis.

If individual case duration follows a log-normal distribution, then the duration of a list of cases equals the sum of log-normal variables. The sum of log-normal variables has
been studied in the field of electronic engineering to analyze the performance of wireless communication systems.\textsuperscript{25,26,27} Most of the research assumes that the sum of lognormal variables follows a new log-normal distribution. However, this assumption has been proved to be false.\textsuperscript{28,29} In order to overcome this shortcoming, Nie and Chen\textsuperscript{28} proposed a new model using Type IV Pearson distribution to model the sum of log-normal variables. In the proposed method, a Type IV Pearson distribution was identified to approximate the duration distribution of a list of cases. We then obtained the 10\textsuperscript{th} to 90\textsuperscript{th} percentiles of the Type IV Pearson distribution corresponding to the duration of the list of cases.

For each list, the actual workload equaled the total actual case durations of all cases on the list plus the total turnover times. Whenever a turnover time was longer than 90 minutes, we rounded the turnover time down to 90 minutes. We used 90 minutes as the maximum because 90 minutes was the 90\textsuperscript{th} percentile of the turnover time in 2010.\textsuperscript{4} Longer turnover times might be due to gaps in the OR schedule (e.g. non-sequential cases).\textsuperscript{30} We did not want to consider using turnover times before 2010 because the further the data was away from the studied date range, the higher the risk that the distribution of turnover times had shifted.\textsuperscript{12,31}

If the proposed method is accurate, then the difference between the proportion of lists of cases that was completed within each percentile and the percentile value should be small (i.e. ideally, 10\% to 90\% of the lists of cases should be finished within the 10\textsuperscript{th} to 90\textsuperscript{th} percentiles, respectively). The following processes were conducted to calculate the proportions of lists of cases that was completed within the 10\textsuperscript{th} to the 90\textsuperscript{th} percentiles of Type IV Pearson and t-distributions:\textsuperscript{12,24}

1. A counter was set to zero.
2. For each list of cases:
   a. Was the actual workload smaller than the percentile of its corresponding Type IV Pearson distribution/t-distribution? If yes, then the counter was incremented by one. If no, nothing.
3. The proportion of lists of cases that was completed within each percentile equaled the value of the counter divided by the total number of lists of cases.

The proportion calculated by Type IV Pearson distribution was compared with the one derived using t-distribution by comparing the calculated absolute deviations between
the calculated proportions with the percentiles.\textsuperscript{10} Absolute deviation equals the absolute value of the difference between the proportion and percentile. It is an indicator of how close the proportion of cases that was completed within each percentile calculated from the distributions approximates the percentile; thus, the smaller the absolute deviation is the better.

OR schedulers only want to include the most recent data to calculate the percentiles of the duration distribution of lists of cases as surgeon case durations for procedures may change.\textsuperscript{11,31} The more previous cases are included, the higher the probability that the surgeons have become faster or slower in doing the procedures. To identify a reasonable decision point for the OR manager to determine how many previous cases to include for the calculation of the probability of under- and over-run of a list of cases, we repeated the above calculation steps by selecting different numbers of previous cases included for distribution parameter estimation, ranging from two to twenty. For combinations of surgeon and procedure(s) whose numbers of previous cases were smaller than the selected sample size, all the historical data was used for parameter estimation. After this was done, we investigated the problems in current scheduling practice of the studied facility. We identified the percentiles of the distribution of the scheduled duration of lists of cases with respect to the Type IV Pearson distributions corresponding to the lists.

In the research of Nie and Chen, the conclusion that Type IV Pearson distribution is a good representation of the sum of log-normal variables was based on the variability of the input variables being close to each other.\textsuperscript{28} This condition does not hold for surgery case durations. The standard deviation of the logarithms of case durations in decimal base (dB) for the combinations of surgeon and procedure(s) varied from 0.05 to 5.90. To test if Type IV Pearson distribution remained valid, we derived the empirical percentiles of the duration distribution of each list by doing 100,000 replications of Monte Carlo simulations, assuming that the duration of each case and turnover time in all lists followed a log-normal distribution. To use the Monte Carlo method, we first identified the log-normal distribution of the duration of each case and turnover time in a list. Then, in the Monte Carlo method, a random value was drawn from each of the identified distributions. The sum of these random values was considered a random value from the duration distribution of the list. By running large number of replications in simulation, the
shape of the generated distribution that consisted of the random values would be close to the true duration distribution of the list. It also ensured that the confidence intervals for the empirical percentiles would be so small that it converged almost to a point. The mean absolute deviation (i.e. the mean of the absolute value of the deviation) between the empirical percentiles and the ones calculated from Type IV Pearson distributions along with the standard errors were used for validation. A small mean absolute deviation indicates a good match between the hypothesized distribution and the true distribution. R 2.15.0 was used to calculate the percentiles of Type IV Pearson distribution for each list and perform Monte Carlo simulations. The standard errors for the proportion of lists of cases that was completed within each percentile were calculated by Clopper-Pearson methods.

RESULTS

Among the 127 combinations of surgeon and procedure(s) to test whether individual case duration followed log-normal distributions, 88 combinations (69%) passed the Shapiro-Wilk tests by having P-values greater than 0.05. The majority of the case durations were log-normally distributed, matching the findings of previous research.11,15 As a result, it is reasonable for us to proceed to the next step to define the duration distribution of lists of multiple cases assuming log-normal distribution of individual case duration.

From the 50th to 60th percentiles of the duration distribution of lists of cases, the differences between the estimated proportions of lists of cases that were completed within the percentiles calculated by Type IV Pearson and t-distributions were within 3% (Table 1). When we used 10 or more previous cases with the same combination of surgeon and procedure(s) for distribution parameter estimation, Type IV Pearson distribution provided reliable results between the 20th and 80th percentiles (the differences between the estimated and the true percentiles were within 6%). T-distribution, on the other hand, was not robust for as a wide range. It was only good for between the 40th and 70th percentiles. It performed significantly worse at identifying the probabilities at the tails of the distribution. Compared to t-distribution, the proposed methods is 8.31% ± 0.38% more accurate on average and 14.16% ± 0.19% more accurate at distribution tails (i.e. the 10th and 90th percentiles).
The empirical percentiles of the duration distribution of lists of cases generated by Monte Carlo simulation matched those calculated by Type IV Pearson distribution. The absolute deviation in the estimated percentiles from the two methods between the 10th and 90th percentiles varied from 0.20 ± 0.01 minutes to 0.43 ± 0.03 minutes (Table 2), validating that Type IV Pearson distribution provided a very accurate approximation of the sum of log-normal variables regardless of the ranges in mean and variance.

As indicated by Figure 1, the scheduled durations of lists of cases were very inaccurate in 2011. Based on the calculation of the percentiles of the scheduled durations of lists of cases with respect to associated Type IV Pearson distributions, there were significant under- and over-estimations (from less than 10% to more than 90%). The mean difference between the scheduled durations of lists of cases based on surgeons’ estimates and the actual durations of lists of cases was 5.9 ± 4.5 minutes with an absolute mean difference of 75 ± 3 minutes, whereas the mean difference was 1.0 ± 3.4 minutes with an absolute mean difference of 54 ± 2 minutes if the lists of cases were scheduled per the sum of mean case durations of historical cases (Figure 2).

DISCUSSION

Some previous studies have indicated the insensitivity of staffing to the workload distribution.32,33 For example, Panditt and Dexter found out that as long as the mean workload ≤ 8 hr 25 min, 8 hr of staffing has higher OR efficiency than 10 hr staffing for all combinations of standard deviation and relative cost of over- and under-run and vice versa when the mean workload ≥ 8 hr 50 min. He BB et al. concluded that the empirical distribution derived from historical daily workload generates the lowest long-term OR costs. However, the problem we studied in this paper is different from previous articles. Given a fixed OR staffing, the problem left is to decide how to fill up the allocated OR time with appropriate workload to ensure high utilization with little over-utilized OR time. As the day of surgery approaches, the incremental cost of an hour of under-utilized OR time becomes negligible relative to the cost of an hour of over-utilized OR time (e.g., because the staff have already been scheduled). Consequently, decision-making close to the day of surgery to reduce labor costs should be focused on reducing the hours of over-utilized OR time. To achieve this aim, it is not only the probability of over-run matters, but also the amount of over- and under-run. The proposed approach provides
OR schedulers with a tool to control the degree of over- and under-run. It answers the question that OR schedulers have when they have waiting list of surgeries, how can they tell that the surgeries they pull from the waiting list will be able to complete around the scheduled close time of OR. It complements the existing research in the allocation of OR time by monitoring the variability of OR workload.

Surgery case durations exhibit high variability. When multiple cases make up a list of cases, the variability accumulates. This makes the prediction of the close time of ORs more difficult. In order to ensure a good utilization of OR hours and low over-utilized OR time, the planning of lists of cases needs to consider risks at both low and high ends of the duration distribution of lists of cases. The accurate identification of the duration distribution of lists of cases assists the OR schedulers and managers in planning by examining if the current list has a low probability of falling outside of the pre-defined limits for under- and over-utilized OR time. The proposed method of using Type IV Pearson distribution to approximate the duration distribution of lists of cases is significantly better than assuming a t-distribution of the duration in previous studies, especially towards the tails of the distribution.

The accuracy in the estimation of the probabilities at the tails of distribution is of main concern for OR management. OR schedulers want to know if a list can be completed with little over-utilized OR time (high-end of the distribution) and no earlier than a desired duration (low-end of the distribution) with high certainty. The t-distribution provides reserved estimates of the probabilities of both under- and over-run. As shown in the Results, the use of Type IV Pearson distribution generates significantly more accurate reference points than t-distribution for case scheduling. The improvement is primarily from the under-utilized OR time. Given the OR schedulers use most recent 10 cases to estimate the probability of under- and over-run of lists of cases, up to 31% of days (i.e., 16% from the 10th percentile, and 15% from the 90th percentile) could have either scheduled more cases or had higher OR utilization. If the time savings from such improvement is proportional to the probability, then 149 minutes could have been saved for an 8-hr staffed OR. The same case scheduling algorithm proposed by Pandit and Tavare still applies. However, Type IV Pearson distribution replaces t-distribution to estimate the probabilities of under- and over-run. With more than 10 previous cases of the same combination of surgeon and procedure(s) (Table 1), the improvements in the
accuracy in the probability estimation were flattened. Thus, OR management can limit data collection to the 10 most recent case durations to plan lists of cases. This number matches previous work.\textsuperscript{12}

In the studied facility, the scheduled case durations were based purely on surgeons’ estimates. Although surgeons’ estimates were strong predictor for case durations,\textsuperscript{13,24} the estimates were subject to consistent biases.\textsuperscript{35} Significant under- and over-estimations were associated with surgeons’ estimates in the studied facility. In contrast, mean case durations of historical cases provided better estimates as they scattered more closely around the center of the distribution and were less subject to the bias of surgeons’ estimates. The absolute deviation between the scheduled durations of lists of cases and the actual durations could have been reduced by 20 minutes if the sum of mean case durations of historical cases was used compared to using surgeons’ estimates alone.

There were three major limitations in the proposed method. First, we assumed that each individual case duration and turnover time in the lists of cases followed a log-normal distribution. Although the majority of the case durations proved to be log-normally distributed, approximately 30\% of the cases in the lists of cases were not. This causes mismatches of percentiles at the tails of distribution. Then, to directly use the approach, at least two previous cases with the same combination of surgeon and procedure(s) had to exist to estimate the parameters for the Type IV Pearson distribution. As has been pointed out, there are many cases with inadequate historical case duration information.\textsuperscript{12,36,37} These infrequent cases cause huge uncertainty in OR decision making. Dexter et al. studied such a problem\textsuperscript{19,24} and show that it is appropriate to use the mean case durations of the same procedure(s) by other surgeons to schedule cases.\textsuperscript{19} A Bayesian model was employed to calculate the prediction bounds for individual case duration with little historical case duration information based on surgeons’ estimates.\textsuperscript{24} This solves the problem of the portion of the 401 unaddressed lists containing only one case with no or only one historical case. For the unaddressed lists of multiple cases, there are no such models to overcome this challenge. The proposed method could partially solve the problem by estimating the percentiles of total duration of the cases on the lists that had at least two historical points. By adding up the mean case durations of the same procedure(s) by other surgeons for cases on the lists
that could not be included in the proposed method (i.e., cases of no or only one historical duration), approximate percentiles can be calculated. Thus, the proposed method could potentially be applied to more lists of cases than those directly studied in this article. We did not consider tardiness at the beginning of workday. The tardiness of first cases can be easily incorporated in the analysis by considering it as an increase in turnover times.\textsuperscript{4} Due to incomplete information of the raw data set, we did not include the type of anesthesia, which has been defined as one of the major sources of case duration variability.\textsuperscript{14} For analysts who have access to detailed information on anesthesia and patient’s conditions, they can further divide the data into finer segments. Although this would potentially generate more accurate probabilities, it requires larger sample sizes which would limit its application.\textsuperscript{12,36,37}

In this study, we proposed and validated a new method to approximate the duration distribution of lists of multiple cases. It provides OR management with important information on what they would expect from the planned lists of cases. For a facility whose operations goals include not only the efficiency of use of OR time but also utilization, like the studied facility, after higher and lower limits on the durations of lists of cases have been set, OR managers can rearrange the lists of cases as needed to minimize the probabilities of the durations of the lists of cases falling beyond the set limits. Despite some limitations, our approach performs much better than the previously proposed methods that assume t-distributions,\textsuperscript{10} especially at the tails of the distribution.
ACKNOWLEDGEMENTS

Assistance with the study: none

Financial support and sponsorship: none

Conflict of interest: none
REFERENCES


5 Dexter F, Traub RD. How to schedule elective surgical cases into specific operating rooms to maximize the efficiency of use of operating room time. *Anesth Analg* 2002; 94:933 – 942.


8 Dexter F, Macario A, Qian F, Traub RD. Forecasting surgical groups’ total hours of elective cases for allocation of block time: application of time series analysis to operating room management. *Anesthesiology* 1999; 91:1501 – 1508.


12 Zhou J, Dexter F. Method to assist in the scheduling of add-on surgical cases-upper prediction bounds for surgical case durations based on the log-normal distribution. *Anesthesiology* 1998; **89**:1228 – 1232.

13 Ejkemans MJC, van Houdenhoven M, Nguyen T, Boersma E, Steyerberg EW, Kazemier G. Predicting the unpredictable: a new prediction model for operating room times using individual characteristics and the surgeon’s estimate. *Anesthesiology* 2010; **112**:41 – 49.


18 Zhou J, Dexter F, Macario A Lubarsky DA. Relying solely on historical surgical times to estimate accurately future surgical times is unlikely to reduce the average length of time cases finish late. *J Clin Anesth* 1999; **11**:601 – 605.

19 Macario A, Dexter F. Estimating the duration of a case when the surgeon has not recently performed the procedure at the surgical suite. *Anesth Analg* 1999; **89**:1241 – 1245.

22 Alvarez R, Bowry R, Carter M. Prediction of the time to complete a series of surgical cases to avoid cardiac operating room overutilization. *Can J Anesth* 2010; **57**:973 – 979.

23 Dexter F, Traub RD, Lebowitz P. Scheduling a delay between different surgeons’ cases in the same operating room on the same day using upper prediction bounds for case durations. *Anesth Analg* 2001; **92**:943 – 946.

24 Dexter F, Ledolter J. Bayesian prediction bounds and comparisons of operating room times even for procedures with few or no historic data. *Anesthesiology* 2005; **103**:1259 – 1267.


31 Dexter F. Application of prediction levels to operating room scheduling. *AORN J* 1996; **63**:607 – 615.


33 He B, Dexter F, Macario A, Zenios S. The timing of staffing decisions in hospital


37 Dexter F, Traub RD, Fleisher LA, Rock, P. What sample sizes are required for pooling surgical case durations among facilities to decrease the incidence of procedures with little historical data? *Anesthesiology* 2002; **96**:1230 – 1236.
FIGURE LEGENDS

1. Figure 1. Histogram of Percentiles of Scheduled Duration of lists of cases of Type IV Pearson Distribution. The horizontal axis represents the percentiles of the scheduled duration of each list of cases calculated from corresponding Type IV Pearson distribution of the 515 lists of cases. The vertical axis is for the frequency of lists of cases for each percentile.

2. Figure 2. Histogram of Percentiles of Sum of Mean Case Duration of Historical Cases of Type IV Pearson Distribution. The horizontal axis represents the percentiles of the sum of mean case duration of historical cases of each list of cases calculated from corresponding Type IV Pearson distribution of the 515 lists of cases. The vertical axis is for the frequency of lists of cases for each percentile.
Table 1. Proportion of Lists Completed within the Calculated Percentiles for Type IV Pearson Distribution and T-distribution

<table>
<thead>
<tr>
<th>No. of Previous Cases Used to Estimate Bound</th>
<th>Proportion of Lists Completed within the Calculated Percentiles</th>
<th>Type IV Pearson Distribution</th>
<th>t-Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>20%</td>
<td>30%</td>
</tr>
<tr>
<td>2</td>
<td>18%</td>
<td>26%</td>
<td>31%</td>
</tr>
<tr>
<td>3</td>
<td>14%</td>
<td>24%</td>
<td>30%</td>
</tr>
<tr>
<td>4</td>
<td>14%</td>
<td>23%</td>
<td>31%</td>
</tr>
<tr>
<td>5</td>
<td>13%</td>
<td>23%</td>
<td>30%</td>
</tr>
<tr>
<td>6</td>
<td>13%</td>
<td>25%</td>
<td>31%</td>
</tr>
<tr>
<td>7</td>
<td>14%</td>
<td>25%</td>
<td>31%</td>
</tr>
<tr>
<td>8</td>
<td>14%</td>
<td>23%</td>
<td>31%</td>
</tr>
<tr>
<td>9</td>
<td>15%</td>
<td>24%</td>
<td>31%</td>
</tr>
<tr>
<td>10</td>
<td>15%</td>
<td>24%</td>
<td>32%</td>
</tr>
<tr>
<td>12</td>
<td>15%</td>
<td>24%</td>
<td>32%</td>
</tr>
<tr>
<td>14</td>
<td>14%</td>
<td>22%</td>
<td>29%</td>
</tr>
<tr>
<td>16</td>
<td>15%</td>
<td>22%</td>
<td>30%</td>
</tr>
<tr>
<td>18</td>
<td>14%</td>
<td>22%</td>
<td>30%</td>
</tr>
<tr>
<td>20</td>
<td>14%</td>
<td>22%</td>
<td>30%</td>
</tr>
<tr>
<td>All</td>
<td>14%</td>
<td>23%</td>
<td>32%</td>
</tr>
</tbody>
</table>
The first column is the number of previous cases used to generate percentiles of Type IV Pearson distribution and t-distribution. Column 2 to Column 10 are for proportion of lists of multiple cases that was completed within each of the percentiles, from the 10th to 90th percentile, of Type IV Pearson distribution corresponding to the duration distribution of the lists. Column 11 to Column 19 are the proportion from the 10th to 90th percentile of t-distribution. For example, Row 5 Column 3 means 25% of the lists of multiple cases were completed within the 20th percentiles of the corresponding Type IV Pearson distribution when we used 6 previous cases to estimate the percentile value. The last row in the table was the results of when we included all the previous cases in the estimation of percentiles. The standard errors of the percentage values for both methods were 2%.
Table 2: Absolute Deviation between the Percentiles Calculated by Type IV Pearson Distribution and Empirical Percentiles from Monte Carlo Simulation

<table>
<thead>
<tr>
<th>No. of Previous Cases Used to Estimate Bound</th>
<th>Absolute Deviation Between the Percentiles of Type IV Pearson Distribution and Monte Carlo Simulation (Mean ± Standard Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>0.30 ± 0.03</td>
</tr>
<tr>
<td>20%</td>
<td>0.30 ± 0.01</td>
</tr>
<tr>
<td>30%</td>
<td>0.33 ± 0.02</td>
</tr>
<tr>
<td>40%</td>
<td>0.33 ± 0.02</td>
</tr>
<tr>
<td>50%</td>
<td>0.35 ± 0.02</td>
</tr>
<tr>
<td>60%</td>
<td>0.36 ± 0.02</td>
</tr>
<tr>
<td>70%</td>
<td>0.35 ± 0.02</td>
</tr>
<tr>
<td>80%</td>
<td>0.35 ± 0.01</td>
</tr>
<tr>
<td>90%</td>
<td>0.34 ± 0.01</td>
</tr>
<tr>
<td>All</td>
<td>0.40 ± 0.02</td>
</tr>
</tbody>
</table>
The first column is the number of previous cases used to estimate percentiles of Type IV Pearson distribution and Monte Carlo simulation. Column 2 to the last column are for the absolute deviations between the percentiles identified by Type IV Pearson distribution and Monte Carlo simulation, including the mean values and the standard errors. For example, Row 5 Column 3 says that when using 6 of previous case to estimate the 20th percentile of Type IV Pearson distribution and the empirical percentile of Monte Carlo simulation, the difference between the two estimates has a mean value of 0.38 minute with an standard error of 0.02 minutes. The last row in the table was the results when we included all the previous cases in the estimation of percentiles.