

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/262777776>

Using type IV Pearson distribution to calculate the probabilities of underrun and overrun of lists of multiple cases

ARTICLE *in* EUROPEAN JOURNAL OF ANAESTHESIOLOGY · JULY 2014

Impact Factor: 2.94 · DOI: 10.1097/EJA.0b013e3283656ba4 · Source: PubMed

CITATIONS

6

READS

46

2 AUTHORS, INCLUDING:



Kai Yang

Wayne State University

108 PUBLICATIONS 1,148 CITATIONS

SEE PROFILE

ORIGINAL ARTICLE

Using type IV Pearson distribution to calculate the probabilities of underrun and overrun of lists of multiple cases

Jihan Wang and Kai Yang

BACKGROUND An efficient operating room needs both little underutilised and overutilised time to achieve optimal cost efficiency. The probabilities of underrun and overrun 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 underrun and overrun of lists of cases using Type IV Pearson distribution to support case scheduling.

DESIGN Six years of data were collected. The first 5 years of data were used to fit distributions and estimate parameters. The data from the last year were 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 Operating rooms in John D. Dingell VA Medical Center, United States, from January 2005 to December 2011.

MAIN OUTCOME MEASURES Differences between the proportion of lists of cases that were completed within the

percentiles of the proposed duration distribution of the lists and the corresponding percentiles.

RESULTS Compared with the *t*-distribution, the proposed new distribution is 8.31% (0.38) 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 a major concern of operating room schedulers. The absolute deviations between the percentiles defined by Type IV Pearson distribution and those from Monte Carlo simulation varied from 0.20 min (0.01) to 0.43 min (0.03). Operating room schedulers can rely on the most recent 10 cases with the same combination of surgeon and procedure(s) for distribution parameter estimation to plan lists of cases. Values are mean (SEM).

CONCLUSION The proposed Type IV Pearson distribution is more accurate than *t*-distribution to estimate the probabilities of underrun and overrun of lists of cases. However, as not all the individual case durations followed log-normal distributions, there was some deviation from the true duration distribution of the lists.

Published online 20 May 2014

This article is accompanied by the following Invited Commentary:

Macario A. Implementing operating room management science. From the bench to the scheduling office. *Eur J Anaesthesiol* 2014; 31:355–360.

Introduction

The operating room is one of the most expensive areas within a hospital.^{1,2} To control costs, it is important for operating room management to plan the lists of cases in such a way that the allocated operating room time is utilised as much as possible with little overutilised time, thereby maximising the efficiency of use.^{3–6} A list consists of cases taking place in the same operating room

From the Steward Health Care System, Boston, Massachusetts (JW) Department of Industrial and System Engineering, Wayne State University, Detroit, Michigan (KY), USA

Correspondence to Jihan Wang, PhD, Quality and Patient Safety Department, Steward Health Care System, 77 Warren Street 6th FL, Brighton, MA 02135, USA. Tel: +1 617 789 2031; fax: +1 617 789 5103; e-mail: Jihan.Wang@steward.org

on the same day. When the duration of a list of cases (i.e. the operating room workload) is greater than the allocated time, overutilised operating room time is observed. Too much overutilised operating room time may cause case cancellations.⁷ Meanwhile, when the duration is smaller than the allocated time, underutilised operating room time is seen and capacity is wasted. When surgeons can schedule elective cases on any workday, the maximum efficiency of use of operating room time can be achieved by predicting the future operating room workload⁸ and determining the optimal operating room allocation.^{3,4} In most operating room suites in the United States, the decisions on operating room allocation are made every 2 to 3 months. For a system such as this, case duration prediction accuracy is less of a problem because the decisions on operating room allocation take into consideration case duration prediction inaccuracy, as the actual workload is used for calculation.^{3,4} With more than 2 historical cases of the same combination of surgeon and procedure(s), the reduction in overutilised operating room time is negligible when people use a more accurate case duration prediction model rather than the mean case duration of historical cases to assign scheduled case durations. Theatre staff work late because of the workload rather than an underestimation in case durations.⁹

The operating room allocation optimisation approach, however, does not apply to European operating room suites. In these facilities, the demand for surgery is so high that patients have to enter a waiting list first and wait for months before their surgery can be scheduled. In addition, the allocated operating room time is usually static, and operating room managers do not usually change the operating room allocation.¹⁰ The way to maximise the efficiency of use of operating room time is to avoid both underutilised and overutilised operating room time through scheduling decisions rather than operating room allocation. A method to accurately estimate the durations of lists of cases is needed for use within this type of system.

Much research has been done in predicting the case duration of single cases.^{11–16} 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.^{14,17–19} More accurate models only improve the prediction on the central tendency (i.e. the mean or median of case duration).¹⁸ As there is a high variability associated with case duration,²⁰ such improvements make limited differences to practice. When there are multiple cases on 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 duration makes the expected close time of the operating room not a reliable

estimate. It is not uncommon to see a list of cases underrun or overrun by more than an hour. It is more meaningful for operating room managers to estimate how reliably a list of cases can be completed within a pre-defined range with some tolerances in both underutilised and overutilised operating room time. For example, if the allocated operating room time is 8 h, the operating room manager wants to know how certain it is that a scheduled list of cases can be finished between 7 and 8.5 h. If it is highly likely, then the list can be finalised because it will not cause a large amount of underutilised or overutilised operating room time. Otherwise, the operating room manager can rearrange the cases on the list so that both underutilised and overutilised time will not be excessive.

Little research has been done looking at this problem. Dexter *et al.*²¹ built an optimisation model and concluded that using mean case durations of historical cases to schedule lists of cases actually generates the optimal efficiency of use of operating room time. However, the problem Dexter *et al.* studied is different from ours because they assumed that surgeons had open access to the operating room and that the allocation of operating room time had been optimised to maximise its efficiency. As explained above, this is not applicable for European systems. Alvarez *et al.*²² explored whether the second tertile cut-off point was better than the sum of mean case durations of historical cases in predicting overutilisation. They reached the same conclusion as that by Dexter *et al.* that mean case duration of historical cases is a good estimate for case scheduling. However, they only studied cardiovascular surgeries, and the results cannot be generalised to other specialties because case variability and duration distribution are different. In 2011, Pandit and Tavaré¹⁰ proposed a scheduling algorithm that planned a list of cases based on the probability of the list being completed within predefined limits on underutilised and overutilised operating room time. This approach was much better than the *ad hoc* scheduling approach used 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 has shown that individual case duration actually follows a log-normal distribution.^{11,15} Thus, the sum of log normally distributed case durations does not follow a Gaussian distribution. The probabilities of underrun and overrun of a list of cases calculated from *t*-distribution would deviate significantly from the true values.

In this study, we proposed a new approach to approximate the probability of underrun and overrun 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 1 year's lists of multiple cases.^{12,23,24} We compared the results from the proposed distribution with those

generated by t -distribution as in the previous study.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 underrun and overrun.

Methods

Among the surgeons who were working in 2011, some 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 were collected for all surgeries performed from 1 January 2005 to 31 December 2011, in John D. Dingell VA Medical Center located in Detroit, Michigan, USA.

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 operating room scheduler could reliably estimate the probability of underrun and overrun of lists of cases. Data from 1 January 2005 to 31 December 2010 were used to estimate the case duration distribution parameters. The most recent data from 1 January 2011 to 31 December 2011 acted as testing dataset to evaluate the performance of the proposed method.

Log-normality tests of individual case duration distribution

From 1 January 2005 to 31 December 2010, there were 15 646 cases, 121 (0.8%) 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 have shown that type of anaesthesia is an important factor that influences the case duration;¹⁴ however, when surgeons scheduled cases, they did not consult anaesthesiologists and did not always know the type of anaesthesia planned, making this information incomplete and unreliable. We, therefore, did not consider the type of anaesthesia in our analysis. This would not affect the analysis as the influence of type of anaesthesia was reflected in the actual case duration. The variability in case duration caused by anaesthesia was accounted for 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 ($n \geq 20$). 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. Shapiro–Wilk tests were then conducted for the log-transferred datasets. 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 (<http://CRAN.R-project.org/doc/FAQ/R-FAQ.html>).

Duration distribution of lists of multiple cases

In 2011, there were 1463 lists of cases. The captured data were not able to identify when a case was moved to another operating room. 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,^{12,24} 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 that had no, or only one, historical case. This made the estimation of case duration variance infeasible. We excluded these lists as well, leaving a total of 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 analyse the performance of wireless communication systems.^{25–27} Most of the research assumes that the sum of log-normal variables follows a new log-normal distribution. However, this assumption has been proven to be false.^{28,29} In order to overcome this shortcoming, Nie and Chen²⁸ 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 10th to 90th percentiles of the Type IV Pearson distribution corresponding to the duration of the list of cases.

For each list, the actual workload equalled the total actual case durations of all cases on the list and the total turnover times. Whenever the turnover time was longer than 90 min, we rounded the turnover time down to 90 min. We used 90 min as the maximum time because this represented the 90th percentile of turnover time in 2010.⁴ Longer turnover times might be due to gaps in the operating room schedule (e.g. non-sequential cases).³⁰ We did not want to consider using turnover times before 2010 because the further the data were away from the studied date range, the higher the risk that the distribution of turnover times had shifted.^{12,31}

If the proposed method was accurate, then the difference between the proportion of lists of cases that were 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 10th to 90th percentiles, respectively). The following processes were conducted to calculate the proportions of lists of cases

that were completed within the 10th to the 90th percentiles of Type IV Pearson and t -distributions:^{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, then the count remained unchanged.
- (3) The proportion of lists of cases that were completed within each percentile equalled the value of the counter divided by the total number of lists of cases.

The proportion calculated by Type IV Pearson distribution was compared to the one derived using t -distribution by comparing the calculated absolute deviations between the calculated proportions with the percentiles.¹⁰ Absolute deviation equals the absolute value of the difference between the proportion and percentile. It is an indicator of how closely the proportion of cases that were completed within each percentile calculated from the distributions approximated the percentile.

Operating room schedulers only want to include the most recent data to calculate the percentiles of the duration distribution of lists of cases as case durations for procedures may change.^{11,31} The less recent the cases that 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 operating room manager to determine how many previous cases to include for the calculation of the probability of underrun and overrun 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 2 to 20. For combinations of surgeon and procedure(s) whose numbers of previous cases were smaller than the selected sample size, all the historical data were used for parameter estimation. After this was done, we investigated the problems in the 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,²⁸ 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. 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 whether 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. A random value was then 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 numbers 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. This also ensured that the confidence intervals for the empirical percentiles would be so small that it almost converged to a single 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 their standard errors were used for validation. A small mean absolute deviation indicates a good match between the hypothesised 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 were 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 was 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 as 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, however, was not robust for as wide a range, only showing good prediction between the 40th and 70th percentiles. T -distribution performed significantly worse at identifying the probabilities at the tails of the distribution. Compared with t -distribution, the proposed method 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).

Table 1 Proportion of lists completed within the calculated percentiles for Type IV Pearson distribution and *t*-distribution

Number of previous cases used in estimate	Proportion of lists completed within the calculated percentiles																	
	Type IV Pearson distribution									<i>t</i> -distribution								
	10%	20%	30%	40%	50%	60%	70%	80%	90%	10%	20%	30%	40%	50%	60%	70%	80%	90%
2	18	26	31	40	46	51	59	66	73	29	35	41	44	47	49	55	58	62
3	14	24	30	37	45	51	61	67	76	28	34	40	43	47	50	55	60	63
4	14	23	31	38	46	55	62	70	80	30	34	40	44	48	52	58	62	66
5	13	23	30	39	45	53	63	71	80	30	35	40	44	49	52	57	61	67
6	13	25	31	39	46	53	63	72	81	30	35	39	45	49	52	55	62	68
7	14	25	31	39	46	54	63	72	82	30	35	41	45	49	53	57	62	69
8	14	23	31	38	47	55	64	72	82	29	36	40	45	49	54	58	63	68
9	15	24	31	39	48	56	66	74	84	29	36	40	45	52	56	60	65	71
10	15	24	32	40	48	58	67	74	85	30	38	42	46	52	56	61	65	71
12	15	24	32	41	47	56	66	74	85	30	38	42	46	50	55	60	65	70
14	14	22	29	40	46	54	64	73	84	29	37	42	45	50	54	58	64	69
16	15	22	30	40	47	54	65	74	85	28	37	41	46	50	54	58	63	70
18	14	22	30	40	48	55	64	74	85	28	37	42	47	50	54	58	63	70
20	14	22	30	41	48	56	65	74	85	28	38	43	47	51	55	59	64	70
All	14	23	32	42	50	56	66	74	87	28	38	45	48	53	57	60	66	74

The first column shows 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 were 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 show 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 shows the results of when we included all the previous cases of the same combination of surgeon and procedure(s) in the estimation of percentiles. The standard error of the percentage values for both methods was 2%.

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) to 0.43 (0.03) min (Table 2), showing 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 Fig. 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 underestimation and overestimation (from less than 10% to more than 90%). The mean difference between the scheduled duration of lists of cases based on surgeons' estimates and the actual durations of lists of cases was 5.9 (4.5) min with an absolute mean difference of 75 (3) min, whereas the mean difference was 1 (3.4) min with an absolute mean difference of 54 (2) min if the lists of cases were scheduled per the sum of mean case durations of historical cases (Fig. 2). All values are mean (SEM).

Discussion

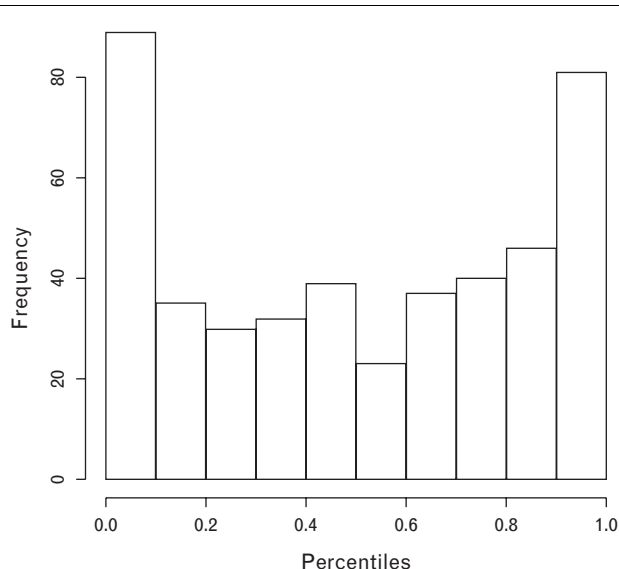
Some previous studies have indicated the insensitivity of staffing to workload distribution. For example, Pandit

Table 2 Absolute deviation between the percentiles calculated by Type IV Pearson distribution and empirical percentiles from Monte Carlo simulation

Number of previous cases used in estimate	Absolute deviation between the percentiles of Type IV Pearson distribution and Monte Carlo simulation (mean $\pm$ standard error)								
	10%	20%	30%	40%	50%	60%	70%	80%	90%
2	0.30 $\pm$ 0.03	0.28 $\pm$ 0.02	0.26 $\pm$ 0.02	0.24 $\pm$ 0.02	0.23 $\pm$ 0.02	0.26 $\pm$ 0.03	0.30 $\pm$ 0.02	0.32 $\pm$ 0.02	0.43 $\pm$ 0.03
3	0.30 $\pm$ 0.01	0.32 $\pm$ 0.02	0.28 $\pm$ 0.02	0.22 $\pm$ 0.01	0.20 $\pm$ 0.01	0.22 $\pm$ 0.01	0.27 $\pm$ 0.01	0.34 $\pm$ 0.02	0.38 $\pm$ 0.02
4	0.33 $\pm$ 0.02	0.31 $\pm$ 0.01	0.26 $\pm$ 0.01	0.22 $\pm$ 0.01	0.21 $\pm$ 0.01	0.23 $\pm$ 0.01	0.27 $\pm$ 0.01	0.32 $\pm$ 0.01	0.41 $\pm$ 0.02
5	0.33 $\pm$ 0.02	0.33 $\pm$ 0.02	0.27 $\pm$ 0.01	0.23 $\pm$ 0.01	0.20 $\pm$ 0.01	0.22 $\pm$ 0.01	0.28 $\pm$ 0.01	0.34 $\pm$ 0.02	0.39 $\pm$ 0.02
6	0.35 $\pm$ 0.02	0.38 $\pm$ 0.02	0.23 $\pm$ 0.02	0.24 $\pm$ 0.02	0.22 $\pm$ 0.02	0.25 $\pm$ 0.02	0.29 $\pm$ 0.02	0.33 $\pm$ 0.02	0.38 $\pm$ 0.02
7	0.36 $\pm$ 0.02	0.34 $\pm$ 0.02	0.28 $\pm$ 0.01	0.23 $\pm$ 0.01	0.22 $\pm$ 0.01	0.26 $\pm$ 0.01	0.30 $\pm$ 0.01	0.35 $\pm$ 0.02	0.39 $\pm$ 0.02
8	0.35 $\pm$ 0.02	0.33 $\pm$ 0.02	0.26 $\pm$ 0.02	0.21 $\pm$ 0.01	0.22 $\pm$ 0.01	0.26 $\pm$ 0.01	0.31 $\pm$ 0.01	0.37 $\pm$ 0.02	0.42 $\pm$ 0.02
9	0.35 $\pm$ 0.01	0.34 $\pm$ 0.01	0.27 $\pm$ 0.01	0.22 $\pm$ 0.01	0.21 $\pm$ 0.01	0.26 $\pm$ 0.01	0.31 $\pm$ 0.01	0.36 $\pm$ 0.01	0.43 $\pm$ 0.02
10	0.34 $\pm$ 0.01	0.32 $\pm$ 0.01	0.26 $\pm$ 0.01	0.22 $\pm$ 0.01	0.22 $\pm$ 0.01	0.26 $\pm$ 0.01	0.31 $\pm$ 0.01	0.35 $\pm$ 0.02	0.39 $\pm$ 0.02
12	0.35 $\pm$ 0.01	0.32 $\pm$ 0.01	0.26 $\pm$ 0.01	0.21 $\pm$ 0.01	0.21 $\pm$ 0.01	0.26 $\pm$ 0.01	0.33 $\pm$ 0.01	0.37 $\pm$ 0.01	0.38 $\pm$ 0.02
14	0.37 $\pm$ 0.01	0.36 $\pm$ 0.02	0.29 $\pm$ 0.01	0.23 $\pm$ 0.01	0.23 $\pm$ 0.01	0.27 $\pm$ 0.01	0.33 $\pm$ 0.01	0.38 $\pm$ 0.02	0.40 $\pm$ 0.02
16	0.38 $\pm$ 0.02	0.37 $\pm$ 0.01	0.29 $\pm$ 0.01	0.22 $\pm$ 0.01	0.22 $\pm$ 0.01	0.26 $\pm$ 0.01	0.32 $\pm$ 0.01	0.36 $\pm$ 0.02	0.41 $\pm$ 0.02
18	0.36 $\pm$ 0.02	0.35 $\pm$ 0.01	0.28 $\pm$ 0.01	0.22 $\pm$ 0.01	0.23 $\pm$ 0.01	0.27 $\pm$ 0.01	0.32 $\pm$ 0.01	0.37 $\pm$ 0.01	0.41 $\pm$ 0.02
20	0.34 $\pm$ 0.02	0.33 $\pm$ 0.01	0.26 $\pm$ 0.01	0.21 $\pm$ 0.01	0.22 $\pm$ 0.01	0.28 $\pm$ 0.01	0.33 $\pm$ 0.01	0.37 $\pm$ 0.01	0.37 $\pm$ 0.02
All	0.40 $\pm$ 0.02	0.39 $\pm$ 0.02	0.31 $\pm$ 0.01	0.23 $\pm$ 0.01	0.22 $\pm$ 0.01	0.30 $\pm$ 0.01	0.37 $\pm$ 0.02	0.42 $\pm$ 0.02	0.41 $\pm$ 0.02

The first column is the number of previous cases used to estimate percentiles of Type IV Pearson distribution and Monte Carlo simulation. The remaining columns show 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 shows that when using 6 of previous cases 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 min with a standard error of 0.02 min. The final row in the table shows the results when we included all the previous cases of the same combination of surgeon and procedure(s) in the estimation of percentiles.

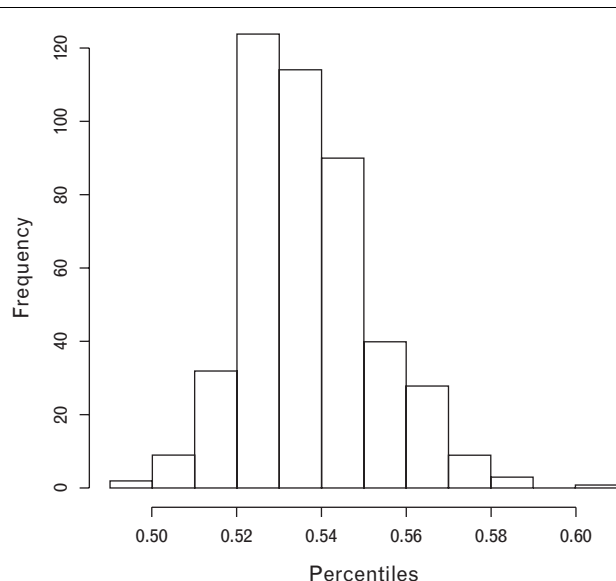
Fig. 1



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.

and Dexter³² found out that as long as the mean workload was 8 h 25 min or less, 8 h of staffing had higher operating room efficiency than 10 h staffing for all combinations of standard deviation and the relative cost of overrun and

Fig. 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.

underrun, and vice versa when the mean workload was at least 8 h 50 min, He *et al.*³³ concluded that the empirical distribution derived from historical daily workload generates the lowest long-term operating room costs. However, the problem we studied in this paper is different from that seen in previous articles. Given a fixed operating room staffing, the problem is to decide how to fill up the allocated operating room time with appropriate workload to ensure high utilisation with little overutilised operating room time. As the day of surgery approaches, the incremental cost of an hour of underutilised operating room time becomes negligible relative to the cost of an hour of overutilised operating room time (e.g. because the staff have already been scheduled). Consequently, decision-making close to the day of surgery to reduce labour costs should be focused on reducing the hours of overutilised operating room time. To achieve this aim, it is not only the probability of overrun that matters, but also the amount of overrun and underrun.

The proposed approach provides operating room schedulers with a tool to control the degree of overrun and underrun. It answers the question that operating room schedulers have when they have waiting lists of surgeries, namely, how can they tell that the surgeries they pull from the waiting list will be able to complete around the scheduled close time. It complements the existing research in the allocation of operating room time by monitoring the variability of operating room workload.

Surgery case durations exhibit high variability.²⁰ When multiple cases make up a list of cases, the variability accumulates. This makes the prediction of the finish time of operating rooms more difficult. In order to ensure a good utilisation of operating room hours and low overutilised operating room 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 operating room schedulers and managers in planning by examining whether the current list has a low probability of falling outside of the predefined limits for underutilised and overutilised operating room time. The proposed method of using Type IV Pearson distribution to approximate the duration distribution of lists of cases is significantly better than the assumption of a *t*-distribution of the duration used in previous studies,^{10,34} especially towards the tails of the distribution.

The accuracy in the estimation of the probabilities at the tails of distribution is of great concern for operating room management. Operating room schedulers want to know whether a list can be completed with little overutilised operating room 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

underrun and overrun. We have shown that the use of Type IV Pearson distribution generates significantly more accurate reference points than *t*-distribution for case scheduling. This improvement is primarily from the underutilised operating room time. Given the operating room schedulers use the most recent 10 cases to estimate the probability of underrun and overrun 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 operating room utilisation. If the time saving from such improvement is proportional to the probability, then 149 min could have been saved for an 8 h staffed operating room. The same case scheduling algorithm proposed by Pandit and Tavaré still applies.¹⁰ However, Type IV Pearson distribution replaces *t*-distribution to estimate the probabilities of underrun and overrun. With more than 10 previous cases of the same combination of surgeon and procedure(s), the improvements in the accuracy in the probability estimation were flattened. Thus, operating room management can limit data collection to the 10 most recent case durations to plan lists of cases. This number matches that seen in previous work.¹²

In the studied facility, the scheduled case durations were based purely on surgeons' estimates. Although surgeons' estimates were a strong predictor for case durations,^{13,24} these were subject to consistent biases³⁵ with significant underestimation and overestimation. In contrast, mean case durations of historical cases provided better estimates as they scattered more closely around the centre 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 min if the sum of mean case durations of historical cases was used compared with using surgeons' estimates alone.

There are some major limitations to 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. Second, 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.^{12,36,37} These infrequent cases cause huge uncertainty in operating room decision-making. Dexter *et al.*^{19,24} studied such a problem and showed that it is appropriate to use the mean case durations of the same procedure(s) by other surgeons to schedule cases.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.²⁴ This solved 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.⁴ Third, due to incomplete information of the raw dataset, we did not include the type of anaesthesia, which has been defined as one of the major sources of case duration variability.¹⁴ For analysts who have access to detailed information on anaesthesia 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 that would limit its application.^{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 operating room management with important information on what they would expect from the planned lists of cases. For a facility whose operation goals include not only the efficiency of use of operating room time but also utilisation, after higher and lower limits on the durations of lists of cases have been set, operating room managers can rearrange the lists of cases to minimise the probability of the durations of the list of cases falling beyond the set limits. Despite some limitations, our approach performs much better than the previously proposed methods that assume *t*-distributions,¹⁰ especially at the tails of the distribution.

Acknowledgements related to this article

Assistance with the study: none

Financial support and sponsorship: none.

Conflicts of interest: none.

Presentation: none.

References

- Macario A, Vitez TS, Dunn B, McDonald T. Where the costs in perioperative care? Analysis of hospital costs and charges for inpatient surgical care. *Anesthesiology* 2005; **83**:1138–1144.
- Denton B, Viapiano J, Vogl A. Optimization of surgery sequencing and scheduling decisions under uncertainty. *Healthcare Manage Sci* 2007; **10**:13–24.

- 3 Dexter F, Epstein RH, Marsh HM. A statistical analysis of weekday operating room anesthesia group staffing costs at nine independent managed surgical suites. *Anesth Analg* 2001; **92**:1493–1498.
- 4 McIntosh C, Dexter F, Epstein RH. The impact of service-specific staffing, case scheduling, turnovers, and first-case starts on anesthesia group and operating room productivity: a tutorial using data from an Australian hospital. *Anesth Analg* 2006; **103**:1499–1516.
- 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.
- 6 Strum DP, Vargas LG, May JH, Bashein G. Surgical suite utilization and capacity planning: a minimal cost analysis model. *J Med Syst* 1997; **21**:309–322.
- 7 Panditt JJ, Carey A. Estimating the duration of common elective operations: implications for operating list management. *Anesthesia* 2006; **61**:768–778.
- 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.
- 9 Dexter F, Epstein RD, Traub RD, Xiao Y. Making management decisions on the day of surgery based on operating room efficiency and patient waiting times. *Anesthesiology* 2004; **101**:1444–1453.
- 10 Pandit JJ, Tavare A. Using mean duration and variation of procedure times to plan a list of surgical operations to fit into the scheduled list time. *Eur J Anaesth* 2011; **28**:493–501.
- 11 Strum DP, May JH, Vargas LG. Modeling the uncertainty of surgical procedure times: comparison of the log-normal and normal models. *Anesthesiology* 2000; **92**:1160–1167.
- 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 Eijkemans MJC, van Houdenhoven M, Nguyen T, et al. Predicting the unpredictable: a new prediction model for operating room times using individual characteristics and the surgeon's estimate. *Anesthesiology* 2010; **112**:41–49.
- 14 Strum D, Sampson AR, May JH, Vargas LG. Surgeon and type of anesthesia predict variability in surgical procedure times. *Anesthesiology* 2000; **92**:1454–1466.
- 15 Strum DP, May JH, Sampson AR, et al. Estimating times of surgeries with two component procedures: comparison of the lognormal and normal models. *Anesthesiology* 2003; **98**:232–240.
- 16 Li Y, Zhang S, Baugh RF, Huang JZ. Predicting surgical case durations using ill-conditioned CPT code matrix. *IIE Trans* 2010; **42**:121–135.
- 17 Wachtel RE, Dexter F. Influence of the operating room schedule on tardiness from scheduled start times. *Anesth Analg* 2009; **108**:1889–1901.
- 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.
- 20 Macario A. Truth in scheduling: is it possible to accurately predict how long a surgical case will accurately last? *Anesthesiology* 2009; **108**:681–685.
- 21 Dexter F, Traub RD, Qian F. Comparison of statistical methods to predict the time to complete a series of surgical cases. *J Clin Monitor Comp* 1999; **15**:45–51.
- 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.
- 25 Fenton LF. The sum of lognormal probability distributions in scatter transmission systems. *IRE Trans Commun Syst* 1960; **CS-8**:57–67.
- 26 Schwartz S, Yeh Y. On the distribution function and moments of power sums with lognormal components. *Bell Syst Tech J* 1982; **61**:1441–1462.
- 27 Wu J, Nehta NV, Zhang J. Flexible lognormal sum approximation method. *Proc IEEE GLOBECOM '05* 2005; **6**:3413–3417.
- 28 Nie H, Chen S. Lognormal sum approximation with type IV Pearson distribution. *IEEE Commun Lett* 2007; **11**:790–792.
- 29 Beaulieu N, Xie Q. An optimal lognormal approximation to lognormal sum distributions. *IEEE Trans Veh Technol* 2004; **53**:479–489.
- 30 Dexter F, Epstein RH, Marcon E, Ledolter J. Estimating the incidence of prolonged turnover times and delays by time of day. *Anesthesiology* 2005; **102**:1242–1248.
- 31 Dexter F. Application of prediction levels to operating room scheduling. *AORN J* 1996; **63**:607–615.
- 32 Pandit JJ, Dexter F. Lack of sensitivity of staffing for 8-h sessions to standard deviation in daily actual hours of operating room time used for surgeons with long queues. *Anesth Analg* 2009; **108**:1910–1915.
- 33 He B, Dexter F, Macario A, Zenios S. The timing of staffing decisions in hospital operating room: incorporating workload heterogeneity into the newsvendor problem. *MSOM* 2012; **14**:99–114.
- 34 van Houdenhoven M, van Oostrum JM, Hans EW, et al. Improving operating room efficiency by applying bin-packing and portfolio techniques to surgical case scheduling. *Anesth Analg* 2007; **105**:707–714.
- 35 Dexter F, Macario A, Estein RH, Leolter J. Validity and usefulness of a method to monitor surgical services' average bias in scheduled case durations. *Can J Anesth* 2005; **52**:935–939.
- 36 Dexter F, Macario A. What is the relative frequency of uncommon ambulatory surgery procedures in the United States with an anesthesia provider? *Anesth Analg* 2000; **90**:1343–1347.
- 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.