Procedure to Normalize Data for Benchmarking

Robert L Chatburn RRT-NPS FAARC and Richard M Ford RRT FAARC

INTRODUCTION: The hospital billing system is usually the source for reporting activity counts used in benchmarking efforts. Because billing is associated with a specific procedure, benchmarking data are often reported as procedure-days, procedure-shifts, or procedure-hours. Normalizing (usually to procedure-days) is required when comparing data for benchmarking purposes. For an institution that uses hourly billing, simply dividing procedure-hours by 24 (or procedure-shifts by 2 or 3) will underestimate the procedure-days reported by a daily billing system, because daily billing systems use the convention that any fractional day of service is rounded up to the next higher day. The purposes of this study were: (1) to simulate sets of data and determine the expected error with conversion by simple division, (2) to derive a more accurate procedure for normalizing benchmarking data, and (3) to compare the new normalization procedure to simple division, using simulated and actual data. METHODS: A reference population of simulated patient data was created using a spreadsheet to generate random start times paired with actual procedure durations (eg, hours of mechanical ventilation) for 5,000 patients. The spreadsheet calculated "true" billable procedure-days and procedure-shifts from the simulated procedure-hours. Next, a resampling procedure was used to simulate the effect of submitting benchmarking data based on various numbers of patients. The resulting sets of data were used to examine the association between sample size and conversion error when converting from procedure-hours to procedure-days and to generate an alternative conversion procedure that uses linear regression to estimate procedure-days from procedure-hours. An additional regression equation was generated from actual patient data, using simultaneously recorded procedure-hours and procedure-days. The set of mean conversion errors for the 2 regression equations was compared using the Mann-Whitney rank sum test. RESULTS: In general, conversion errors (both systematic and random errors) were smaller with larger sample sizes and with longer service periods, approaching an asymptote at a sample size greater than about 20. Using division, the conversion errors for a sample size of 100 were −16% for hourly reporting, −11% for 8-hour shifts, and −8% for 12-hour shifts. The regression equations for conversion derived from simulated data were as follows. For hourly billing, procedure-days = −0.237 + (0.049) (procedure-hours). For 8-hour shifts, procedure-days = −0.205 + (0.372) (procedure-shifts). For 12-hour shifts, procedure-days = −0.114 + (0.541) (procedure-shifts). Using those regression equations, the conversion errors for a sample size of 100 were −1% for hourly reporting, −0.2% for 8-hour shifts, and −0.2% for 12-hour shifts. The regression equation (for hourly billing) derived from simulated data gave better results than did the equation derived from actual data (median error 0.39 vs −2.92, p = 0.013). *Key words: billing, benchmarking, procedure-days, procedure-shifts, procedure-hours, normalizing.* [Respir Care 2006;51(2):145–157.

Robert L Chatburn RRT-NPS FAARC is affiliated with the Respiratory Care Department, University Hospitals of Cleveland, and with the Department of Pediatrics, Case Western Reserve University, Cleveland, Ohio. Richard M Ford RRT FAARC is affiliated with the Department of Respiratory Care, University of California, San Diego, Medical Center, San Diego, California.

Correspondence: Robert L Chatburn RRT-NPS FAARC, Respiratory Care Department, University Hospitals of Cleveland, 11100 Euclid Avenue, Cleveland OH 44106. E-mail: robert.chatburn@uhhs.com.

> In service business accounting, we know how much money comes in and goes out. But we cannot relate expenditures to results; nobody knows how.
>
> —Peter Drucker[1]

Introduction

Benchmarking in health care has become a key tool used by hospital administrators to make decisions regarding the appropriate resources to employ in the provision of specific services. For example, productivity targets for respiratory care are often established through comparisons of worked or paid hours per unit of service. Units of service may be derived from counts of the activity performed by respiratory practitioners and reported as the number of procedures performed. The hospital billing system is usually the source for these activity counts. However, the use of billing data may present a problem because not all institutions use the same reporting period or billing increment for continuous therapies such as oxygen, oximetry, bi-level positive airway pressure, ventilator care, and continuous monitoring.

Because billing is associated with a specific procedure, benchmarking data are often reported as procedure-days, procedure-shifts, or procedure-hours. Normalizing, or converting data to a common format (usually procedure-days), is required when comparing data for benchmarking purposes. Yet the appropriate conversion factor is not obvious. For an institution that uses hourly billing, simply dividing procedure-hours by 24 (or procedure-shifts by 2 or 3) will underestimate the procedure-days reported by a daily billing system. This is because daily billing systems use the convention that any fractional day of service is rounded up to the next higher day.

What are the implications of this rounding convention? Let's begin by assuming that the preferred unit for reporting benchmarking data are procedure-days. Next, we compare the estimated procedure-days (using some conversion procedure, such as dividing procedure-hours by 24) to the "true value" defined in terms of the daily billing convention of counting fractional days as whole procedure-days. The error of estimating procedure-days from procedure-hours is defined as:

$$error\% = \frac{estimated\ value - true\ value}{true\ value} \times 100\% \tag{1}$$

Immediately we see that the only way conversion error will be zero when dividing procedure-hours by 24 is the specific case where the total procedure-hours are an integer multiple of 24 (eg, 48 procedure-hours converts to 2 d of service with 0% error).

Any conversion procedure that results in fractional days leads to conversion error. For example, suppose a patient was ventilated for 4 hours. In a daily billing system, the patient (or patients) would be billed at one day of service. In an hourly billing system there would be a total of 4 procedure-hours. Converting procedure-hours to procedure-days (4/24) results in 0.17 procedure-days. The error in converting is:

$$error\% = \frac{0.17 - 1}{1} \times 100\% = -83\% \tag{2}$$

When comparing a hospital that uses an hourly billing system to a hospital that uses a daily billing system, the latter appears to be much more productive (ie, more units of service for the same worked hours). Thus, any error in conversion will tend to invalidate benchmarking comparisons.

Looking closer at this example, we see that the error in converting from procedure-hours (or procedure-shifts) to procedure-days can also be affected by the actual procedure start and stop times. For example, suppose a patient is started on mechanical ventilation at 10 PM and then ventilation is stopped 4 hours later at 2 AM. For facilities that count by hours, those 4 hours convert to 0.17 procedure-days, as in the previous example. However, facilities that count per patient per day will report 2 procedure-days. That conversion of procedure-hours to procedure-days now yields an error of −92%.

Consider further, 2 people who both receive 4 procedure-hours that span 2 days, as above. In a daily billing system this would represent 4 procedure-days. In an hourly billing system it would represent 8 procedure-hours, which converts to 0.33 procedure-days. The error is again −92%. But suppose one person received all 4 hours in a single day. Now the daily billing system would record 3 procedure-days and the hourly billing system would still convert those 4 hours to 0.33 procedure-days. The resultant error is −89%.

We conclude from the foregoing analysis that conversion error can range from 0% to nearly −100% (always with procedure-hours underestimating days of service), depending on 3 factors:

1. Start time of procedure
2. Duration of procedure
3. Number of patients represented by the data

These factors interact randomly in any given sample of benchmark data. We would expect conversion errors to occur whenever we divide procedure-hours by 24 (or procedure-shifts by 3 or 2) to estimate procedure-days. The purpose of this study was threefold: (1) to simulate sets of data and determine the expected error with conversion by simple division, (2) to derive a more accurate procedure

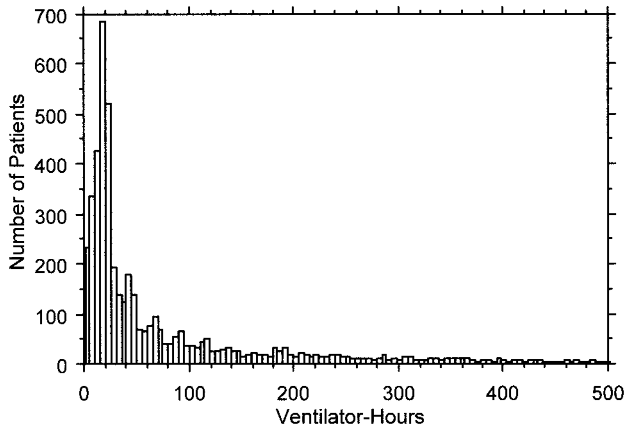

Fig. 1. Distribution of ventilator-hours (duration of mechanical ventilation). Data are clipped at a maximum of 500 hours, to show detail near the mode.

for normalizing benchmarking data that are generated from different reporting periods, and (3) to compare the new normalization procedure to simple division using actual and simulated data. The results of this study are intended to assist benchmarking efforts by providing a common basis for reporting productivity indices calculated with billed procedure-volume data.

Methods

There are 3 primary considerations that impact the process of converting procedure-hours or procedure-shifts to procedure-days. These considerations are the start time and durations of the procedures and the number of patients represented in the total billable units being used in the calculation. Therefore, our first task was to simulate a set of data to represent random distributions of procedure start times and durations. After that, we varied the sample size experimentally and evaluated the effect on the conversion error.

Random start times are easy enough to generate with a spreadsheet function, but simulating durations presented some problems. Specifically, what would be realistic minimum and maximum values, and what should the distribution look like? Rather than guess at the answers to these questions, we began with a set of actual billing data. We used the durations (in hours) of mechanical ventilation, derived from patient data from 5,000 patients admitted to the University of California, San Diego (UCSD) Medical Center during 2003. The distribution of ventilator-hours is shown in Figure 1 and a percentiles plot is shown in Figure 2. The data ranged from 1 hour to 4,967 hours, with a median of 34 hours and a mode of 18 hours.

A spreadsheet (Microsoft Excel, Microsoft, Redmond, Washington) model was created to calculate billable ventilator-days from individual ventilator-hours and randomly generated start times. Similar models were created to convert from ventilator-hours to ventilator-shifts (Table 1 and Table 2). Ventilator-hours were converted to ventilator-days by adding the ventilator-hours to the start hour, dividing the result by 24 hours, and rounding up to the nearest whole day. For example, if a patient received 4 hours of mechanical ventilation, starting at the 22nd hour of the day (ie, clock time = 22:00 h), then a ventilator-day billing system would credit a whole day for the first 2 hours (ie, 22:00 h to 24:00 h) and another whole day for the second 2 hours (00:00 h to 02:00 h). The spreadsheet mathematics for this example would be

$$\textit{ventilator-days} = \textit{roundup}\ (22 + 4)/24 = \text{roundup}\ (1.08) = 2 \quad (3)$$

The value for ventilator-days calculated in this way was used as the "true" number (ie, daily billing convention) when estimating the error of converting ventilator-hours to ventilator-days by dividing ventilator-hours by 24 *without* rounding up (ie, hourly billing convention). Similar equations were used to calculate ventilator-shifts from ventilator-hours, using a divisor of either 8 or 12 hours instead of 24 hours (Table 2).

With the set of simulated data at hand, our next step was to devise an experiment that would simulate the effect of submitting benchmarking data based on various numbers of patients (ie, sample sizes). What we needed was a table of simulated total ventilator-hours and associated ventilator-days for a set of different sample sizes. With this simulated data we could associate "true" ventilator-days with ventilator-hours, using regression analysis. This seemed (a priori) to be a more accurate way to convert from ventilator-hours to ventilator-days than using simple division.

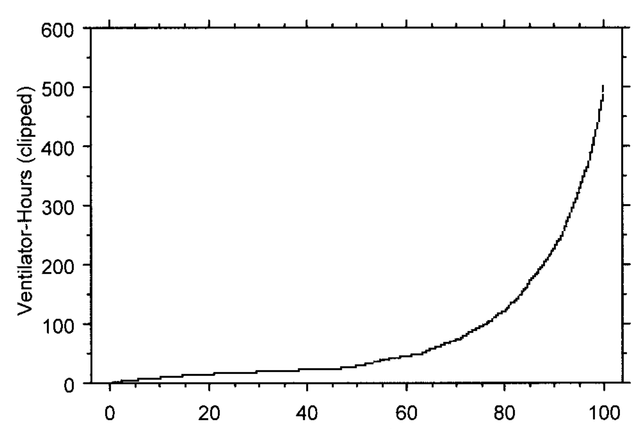

Fig. 2. Percentile plot of ventilator-hours. Data are clipped at a maximum of 500 hours to show detail. The graph allows interpretation of the data in terms of how often certain values occur. For example, 75 percent of the time, the duration of ventilation is ≤ 100 hours. Such plots may be useful in benchmarking, to help assure that the hospitals included in compare groups really are comparable.

Table 1A. Conversion From Ventilator-Hours to Ventilator-Days, With a 24-Hour Billing System, Using Randomly Selected Start Times

Ventilator-Hours	Random Start Hour (1–24)	Ventilator-Days
135	16.9	7
22	10.2	2
97	7.2	5
*	*	*
614,284†		30,397

*The asterisks represent et cetera additional ventilator-hours, start times, and ventilator-days.
†Sum of the ventilator-hours for 5,000 patients.

Table 1B. Conversion From Ventilator-Hours to Ventilator-Days, With an 8-Hour Billing System, Using Randomly Selected Start Times

Ventilator-Hours	Random Start Hour (1–8)	8-Hour Ventilator-Shifts	Ventilator-Days
135	7	18	7
22	1	3	2
97	8	14	5
*	*	*	*
614,284†		81,793	30,397

*The asterisks represent et cetera additional ventilator-hours, start times, shifts, and ventilator-days.
†Sum of the ventilator-hours for 5,000 patients.

Table 1C. Conversion From Ventilator-Hours to Ventilator-Days, With a 12-Hour Billing System, Using Randomly Selected Start Times

Ventilator-Hours	Random Start Hour (1–12)	12-Hour Ventilator-Shifts	Ventilator-Days
135	12.0	13.0	7
22	8.0	3.0	2
97	3.0	9.0	5
*	*	*	*
614,284†		56,219	30,397

*The asterisks represent et cetera additional ventilator-hours, start times, shifts, and ventilator-days.
†Sum of the ventilator-hours for 5,000 patients.

We used a Monte Carlo procedure (see Appendix)[2] to perform the experiment. Briefly, the sample of 5,000 ventilator-hours and associated ventilator-days was considered to contain all the information available about the true population of service-hours and service-days. This sample was then repeatedly resampled[3] with replacement (also known as the "bootstrap" procedure), using different sample sizes. The results of each "sample" from the simulated population were recorded and basic descriptive statistics were calculated. The experiment was conducted with a spreadsheet (using statistical software (a resampling add-in for Microsoft Excel from Resampling Stats, Arlington, Virginia) programmed to do Resampling Procedure A, *to evaluate the relationship between sample size and conversion error when using simple division*, as follows:

Resampling Procedure A

1. Select a random sample of 5 ventilator-hour durations and associated ventilator-days from the pool of 5,000 patients
2. Sum the ventilator-days to get the "true" total ventilator-days
3. Sum the ventilator-hours (or ventilator shifts) for the 5 patients to get total ventilator-hours (or ventilator shifts)
4. Convert ventilator-hours to ventilator-days, using simple division by 24, (or convert ventilator-shifts to ventilator-days using division by 3 or 2)
5. Calculate the conversion error, using Equation 1
6. Repeat steps 1–5 one thousand times
7. Calculate the mean and standard deviation for the conversion error

Resampling Procedure A was then repeated for sample sizes of 10, 15, 20, 25, 30, 40, 100, and 500, and the results were plotted as mean ± SD conversion error versus sample size.

We devised Resampling Procedure B *to derive a more accurate conversion procedure using linear regression instead of simple division*, as follows:

Resampling Procedure B

1. Select a random sample of 5 ventilator-hour durations and their associated ventilator-days from the pool of 5,000 patients
2. Sum the ventilator-days to get the "true" total ventilator-days
3. Sum the ventilator-hours (or ventilator shifts) for the 5 patients to get total ventilator-hours (or ventilator shifts)
4. Repeat steps 1–5 one thousand times.
5. Calculate the mean and standard deviation for ventilator-days and ventilator-hours (or ventilator-shifts)

Resampling Procedure B was repeated for sample sizes of 10, 15, 20, 25, 30, 40, 100, 150, 200, 300, and 500. Then mean ventilator-days were plotted against mean ventilator-hours (or ventilator-shifts) and a linear regression equation was generated using statistical software (StatView, SAS Institute, Cary, North Carolina) to predict ventilator-days from ventilator-hours (or ventilator-shifts).

An alternative regression equation was created using actual patient data obtained from UCSD Medical Center (recorded monthly from October 2004 to April 2005). Patient data were collected using clinical respiratory care operations software (CliniVision MPC, Puritan Bennett Melville, Kanata, Ontario, Canada), which was configured to simultaneously capture both the total procedure-hours

Table 2. Equations for Spreadsheet Models

Variable	Equation
Random start hour (1–24)	randbetween(1,24)
Ventilator days	IF(random start hour =24,ROUNDUP(vent hours/24,0),ROUNDUP((random start hour+vent hours)/24,0))
Random start hour (1–8)	randbetween(1,8)
8-hour ventilator-shifts	IF(random start hour =24,ROUNDUP(vent hours/8,0),ROUNDUP((random start hour+vent hours)/8,0))
Random start hour (1–12)	randbetween(1,12)
12-hour ventilator-shifts	IF(random start hour =24,ROUNDUP(vent hours/12,0),ROUNDUP((random start hour+vent hours)/12,0))

and the associated service-days (days of service were considered the sum of initial and subsequent days). Continuous services reported included not only mechanical ventilation but bi-level positive airway pressure, oxygen therapy, humidity therapy, and continuous pulse oximetry. Billing data were captured for a total of 3,218 patients with 265,645 procedure-hours, which corresponded to 13,533 days of service. The sample size (ie, number of patients included in each month's data) was estimated by assuming that the number of patients was equivalent to the number of initial days of service.

Conversion errors as a function of sample size for the regression procedures were evaluated using Resampling Procedure A (step 4 used regression instead of division). Both regression equations were used on the same simulated data set.

The conversion errors using division (by 24 h) were compared with the regression equation (from the simulated data), using the set of actual data from UCSD. For each actual-data pair (ie, total procedure-hours and total days of service) the one conversion error for each conversion procedure was calculated and plotted against the sample size (ie, number of initial service days).

The set of mean conversion errors for the 2 regression equations was compared using the Mann-Whitney rank sum test (because a normality test failed). Differences associated with p values ≤ 0.05 were considered significant. Statistical procedures were performed using statistical software (SigmaStat, Systat Software, Point Richmond, California; StatView, SAS Institute, Cary, North Carolina).

Results

The results of the experiment to determine the relation between sample size and conversion error using division by 24 (using Resampling Procedure A) are shown in Figure 3A. The error seems to approach an asymptote near −16%. Results for converting from 8-hour and 12-hour ventilator-shifts to ventilator-days is shown in Figures 3B and 3C. In general, conversion errors (both systematic and random errors) were smaller for larger sample sizes and for longer service periods, approaching an asymptote at sample sizes greater than about 20. For example, the mean errors were larger when converting from ventilator-hours (ie, 1-hour service periods) to ventilator-days than when converting from both 8-hour and 12-hour shifts (ie, 8-h and 12-h service periods).

The results of the experiment to generate data for a more accurate conversion procedure (Resampling Procedure B) are shown in Table 3. Note that the random error effects are larger for smaller sample sizes. For example, the coefficient of variation (ie, standard deviation divided by mean) for a sample size of 5 was about 10 times greater than for a sample size of 500. Plotting total ventilator-days against total ventilator-hours yielded a nearly perfect linear relationship ($r^2 = 0.999$), shown in Figure 4A. The resultant conversion equation was generalized as:

$$\textit{procedure days} = -0.237 + (0.049)\,(\textit{procedure hours}) \quad (4)$$

The plot of service-days against service-hours for the actual data from UCSD resulted in the general conversion equation:

$$\textit{procedure days} = -1.326 + (0.051)\,(\textit{procedure hours}) \quad (5)$$

which is shown in Figure 4B.

Figure 5 shows the conversion errors associated with the 2 regression equations. The regression equation derived from simulated data gave better results than the equation derived from actual data (median error 0.39 vs −2.92, p = 0.013). Based on these results, regression equations for converting service-shifts to service-days were generated using Resampling Procedure B above.

For 8-hour shifts:

$$\textit{procedure days} = -0.205 + (0.372)\,(\textit{procedure shifts}) \quad (6)$$

For 12-hour shifts:

$$\textit{procedure days} = -0.114 + (0.541)\,(\textit{procedure shifts}) \quad (7)$$

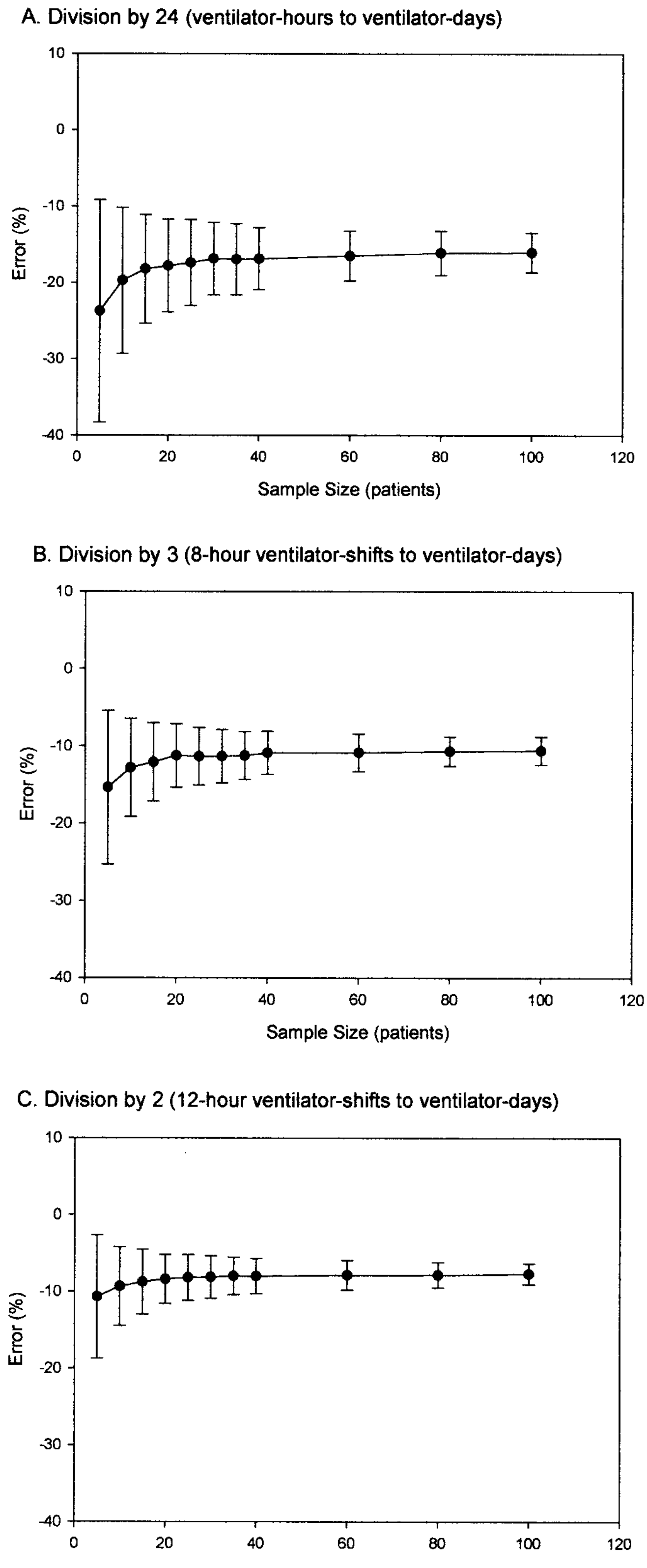

Fig. 3. Error as a function of sample size, using simple division, (A) when converting from ventilator-hours to ventilator-days, (B) when converting from 8-hour ventilator-shifts to ventilator-days, and (C) when converting from 12-hour ventilator-shifts to ventilator-days. Values plotted are mean error ± standard deviation of error.

Table 3. Results of the Experiment to Simulate Total Ventilator-Days and the Associated Total Ventilator Hours for Various Sample Sizes

Sample Size	Mean		Standard Deviation		Coefficient of Variation	
	Ventilator Days	Ventilator Hours	Ventilator Days	Ventilator Hours	Ventilator Days (%)	Ventilator Hours (%)
5	30.4	614.5	21.6	517.5	71	84
10	59.3	1,193.4	31.7	760.5	53	64
15	93.1	1,890.1	39.2	941.0	42	50
20	122.8	2,485.8	45.4	1,089.1	37	44
25	149.4	3,011.7	45.4	1,090.0	30	36
30	183.3	3,707.7	57.5	1,379.3	31	37
35	212.2	4,287.3	60.9	1,462.8	29	34
40	244.9	4,957.5	61.6	1,480.6	25	30
60	360.8	7,277.2	77.2	1,850.1	21	25
80	489.9	9,919.0	87.9	2,108.3	18	21
100	608.2	12,297.5	98.5	2,364.6	16	19
150	911.8	18,435.4	123.1	2,948.0	13	16
200	1,217.4	24,621.3	142.1	3,409.1	12	14
300	1,828.9	36,995.3	176.2	4,219.2	10	11
500	3,027.0	61,144.6	223.4	5,356.9	7	9

Mean error at asymptotes (ie, a sample size of 100) was smaller for regression than for division for every reporting period (Table 4).

Figure 6 compares the conversion errors using simple division by 24 versus the better regression equation. Figure 6 was generated using the UCSD data, by converting the actual procedure-hours to procedure-days and then calculating error on the basis of the actual procedure-days. Conversion errors with the linear regression equations are systematically lower than with simple division by 24. The average error across all sample sizes with conversion via regression was −6%, compared to an error of −20% with conversion via simple division. For both procedures, conversion errors were larger for smaller sample sizes. For sample sizes of 4–18, conversion error reached a maximum of −50% and −55% with regression and division, respectively (at sample size of 6). When the data were restricted to sample sizes of greater than 20, the average conversion errors were −4% and −18%, with maximum errors of −14% and −27%, respectively. The average error of −18% was similar to the simulated data plot asymptote of −16%, as were the general shape of the curves (see Fig. 3A). Indeed, the error curve for division (see Fig. 6) fit within the error bands of the curve for simulated data in Figure 3A, except at the very low sample sizes.

Discussion

Benchmarking can be simply defined as the process of identifying best practices and adopting or adapting them for your own use.[4] In health-care facilities, practice guidelines, care paths, and other forms of evidence-based medicine use the benchmarking idea. Benchmarking can be either internal or external. Internal benchmarking involves the comparison of similar processes within an organization, or the same process during different periods. External benchmarking requires an organization to compare its processes with those of other institutions either inside or outside its own industry.[5] But assuring comparability of benchmarking data can present some challenges.

Of particular interest to health-care managers is the comparison of productivity data. This interest has been stimulated by the continual need to cut costs, which is imposed by an economic environment of decreasing reimbursement. Unfortunately, department managers are often beleaguered by consultants who advocate productivity targets derived from "proprietary" sources. If the consultant is not intimately familiar with the derivation of the target data, serious miscalculations can occur, which will lead to dramatically incorrect management decisions.

Suppose that the goal is to compare 2 departments, using some type of efficiency index, such as the commonly used index calculated as worked hours per billed procedure. An immediate problem results from the fact that just about every department defines its billable procedures in a unique way, such that it is practically impossible to get 2 lists of billable procedures that match sufficiently for benchmarking purposes. However, consultants tend to overlook this complication in the belief that some sort of benchmark is better than nothing. But a more serious problem is that hospitals use different billing periods for some procedures. For example, ventilator usage and oxygen therapy may be reported as hours of use, shifts of use, or days of use. If

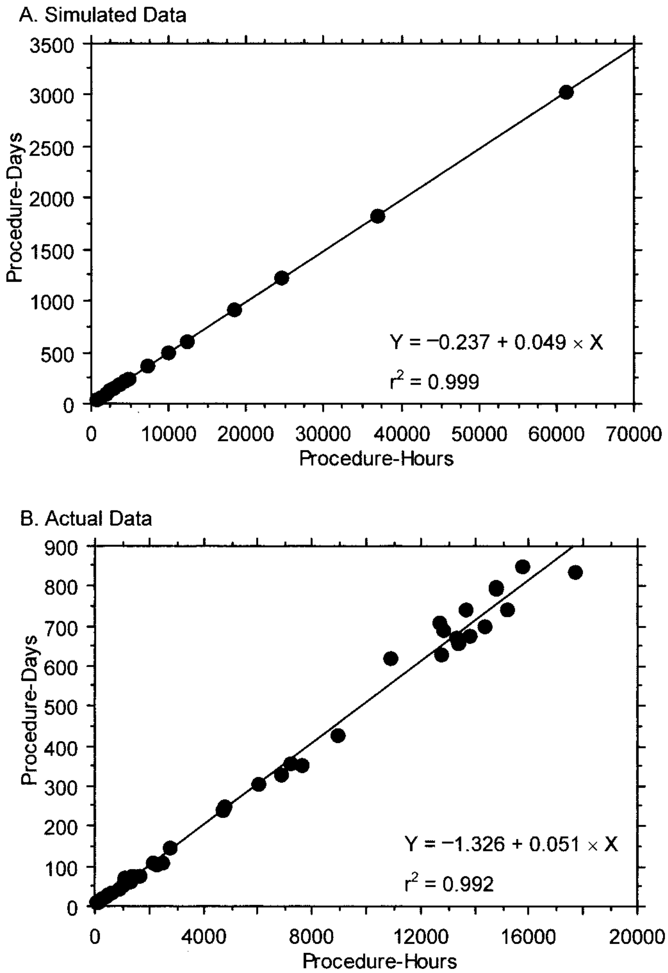

Fig. 4. Relationship predicting procedure-days from procedure-hours, using linear regression. A: Simulated data, with procedure-days calculated from actual procedure-hours using the billing convention of rounding up fractional days. B: Actual data of simultaneously recorded procedure-hours and procedure-days.

this difference is not accounted for, departments that use hours or shifts as the basis for reporting will tend to look less productive than departments that use days as the reporting period.

As an illustration of this problem, Table 5 shows actual data from 2 hospitals that use different periods to define billable procedures. The data set has been restricted to airway-clearance therapy, mechanical ventilation, and aerosol treatments because (1) every department bills for these procedures, (2) there is little difference in their definitions of what these procedures are, and (3) these 3 procedures often make up the majority of the work load (for example, at the University Hospitals of Cleveland these 3 procedures account for 66% of the work load). In Table 5A, notice that Hospital A seems to be considerably less efficient than Hospital B.

Though efficiency is generally defined as output divided by input (ie, worked hours represent input and procedure-volume represents output), the common practice in the consulting world is to invert the index. Thus, Hospital A's average index value of 1.25 indicates less productivity than Hospital B's average index value of 0.56, because Hospital B used less than half as many worked hours for the same volume of procedures. If this type of comparison was all that was offered by a consultant, Hospital A might feel pressured to reduce labor cost to come closer to the "benchmark" established by Hospital B. A reality check, however, is that Hospital A would have to cut its staff almost in half to realize the "opportunity" that might be suggested by a consultant. Given that most hospitals have already gone through decades of cost cutting, a staff reduction of this magnitude would have a noticeable effect on quality.

If we take into consideration the fact that Hospital B's volume for mechanical ventilation represents ventilator-shifts, while Hospital A's volume is ventilator-days, the differences between the 2 hospitals decreases (see Table 5B). However, there is still enough difference to cause a

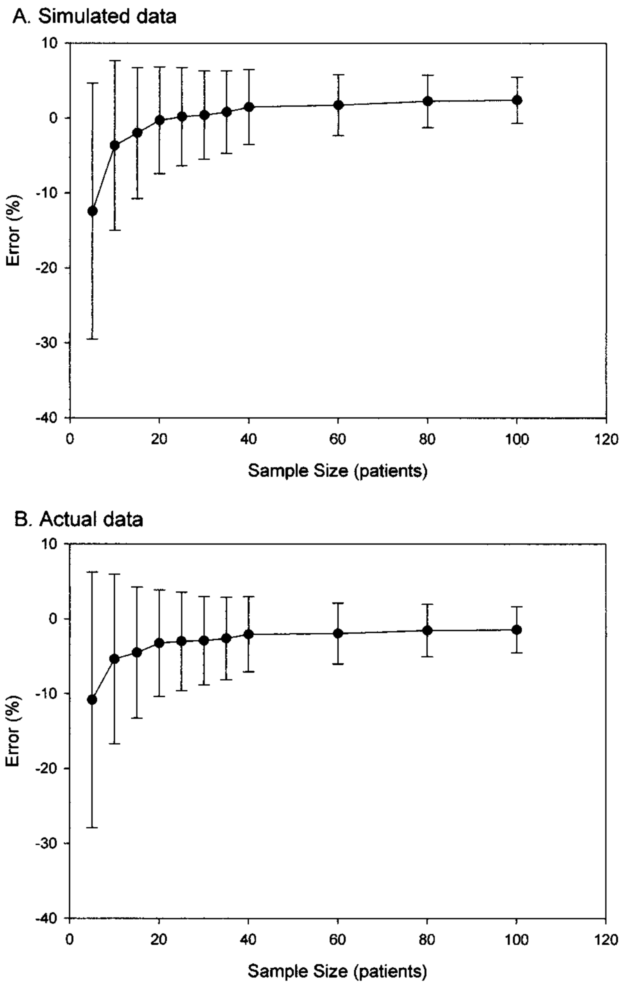

Fig. 5. Error as a function of sample size, using regression. A: Regression equation derived from simulated procedure-hour and procedure-day data. B: Regression equation derived from actual procedure-hour and procedure-day data. Values plotted are mean error ± standard deviation of error.

Table 4. Comparison of Errors Using Division and Regression for a Sample Size of 100 Simulated Patients

Reporting Period	Division Error (mean ± SD %)	Regression Error (mean ± SD %)
Hourly	−16.1 ± 2.6	−1.43 ± 3.09
8-h shifts	−10.6 ± 1.8	−0.20 ± 2.00
12-h shifts	−7.7 ± 1.4	−0.19 ± 1.54

misinterpretation of the data. A closer examination of Table 5 reveals that the discrepancy lies in the contribution of mechanical ventilation to the total work load. For Hospital A, mechanical ventilation (a much more lengthy procedure than either airway clearance or aerosol treatment) makes up over 16% of the total procedures in the data set, compared to less than 4% for Hospital B. Therefore, to make

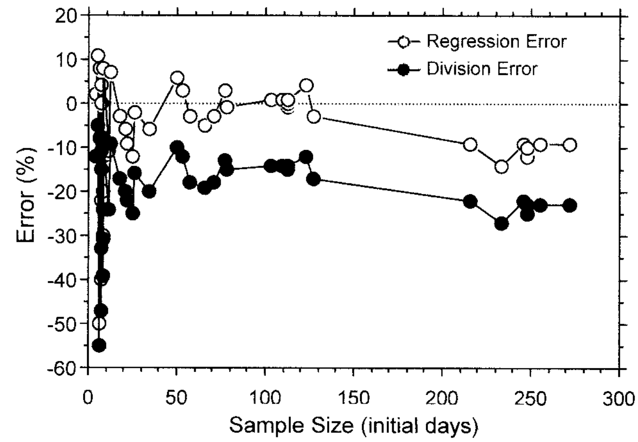

Fig. 6. Comparison of the conversion errors using simple division by 24 versus the better regression equation. Conversion errors were evaluated using actual procedure-hour and procedure-day data as the "true" values.

an accurate benchmarking comparison, we need to use standardized procedure-hours rather than procedure-volume. Now the efficiency index is defined as the expected output (standard hours) divided by the actual output (hours worked to perform the procedures).

Converting from procedure-volume to standard hours is a matter of multiplying each procedure-volume by a time standard for the procedure and summing across all procedures. We can look to the American Association for Respiratory Care (AARC) Uniform Reporting Manual for such time standards or we can use a similar standard, such as those formerly used by the University Hospitals of Cleveland Consortium. For this example, the Consortium's "Activity Time Value" standards are used for convenience, because they are expressed in hours per day, whereas the AARC's Uniform Reporting Manual reports minutes per ventilator assessment. The Consortium standard is 3 labor hours per ventilator per day (eg, ventilator checks every 4 h, at 30 min each). Of course, the conversion is never perfect, because hospitals differ in their policies regarding the frequency of ventilator checks, and thus the average hours per day devoted to managing mechanical ventilation per patient. Actual practice also differs, depending on types of ventilators used, patient illness-severity, et cetera. Added to this is the complication that a ventilator-day may represent any number of combinations of different procedures (eg, set-up vs ventilator-check, noninvasive vs invasive, continuous positive airway pressure) each with different time standards, which are billed under current procedural terminology (CPT) codes 94656, 94657, and 94660. But once we are finally comparing "apples with apples" (ie, standard-hours/worked-hours from both hospitals), the 2 hospitals look very similar (Table 5C). Indeed, a *t* test comparing the average efficiencies of the 2 hospitals shows no statistically significant difference.

Table 5A. Actual Data From 2 Hospitals That Use Different Periods to Define Billable Procedures: Data As Reported

Hospital A: Billable Volume Is Ventilator-Days					
Procedure	Jan	Feb	Mar	Apr	May
Airway-clearance treatments	1,173	1,181	1,343	815	1,038
Mechanical ventilation	1,384	1,204	1,060	923	905
Aerosol treatments	4,569	4,776	5,221	3,928	4,138
Total procedures	7,126	7,161	7,624	5,666	6,081
Worked hours	8,532	8,220	9,042	7,953	8,148
Worked hours/procedure	1.20	1.15	1.19	1.40	1.34
Average worked hours/procedure (across all months)	1.25				

Hospital B: Billable Volume Is 8-Hour Ventilator-Shifts*					
Procedure	Jan	Feb	Mar	Apr	May
Airway clearance treatments	1,836	1,700	1,727	1,104	1,168
Mechanical ventilation	1,864	1,678	1,357	1,226	1,132
Aerosol treatments	12,577	10,672	10,324	9,163	8,629
Total procedures	16,277	14,050	13,408	11,493	10,929
Worked hours	7,938	7,418	7,579	6,798	6,990
Worked hours/procedure	0.49	0.53	0.57	0.59	0.64
Average worked hours/procedure (across all months)	0.56				

*Mechanical ventilation data as ventilator-shift

Table 5B. Actual Data From 2 Hospitals That Use Different Periods to Define Billable Procedures: Hospital B Ventilator-Shifts Converted to Ventilator-Days

Hospital A: Billable Volume Is Ventilator-Days					
Procedure	Jan	Feb	Mar	Apr	May
Airway-clearance treatments	1,173	1,181	1,343	815	1,038
Mechanical ventilation	1,384	1,204	1,060	923	905
Aerosol treatments	4,569	4,776	5,221	3,928	4,138
Total procedures	7,126	7,161	7,624	5,666	6,081
Worked hours	8,532	8,220	9,042	7,953	8,148
Worked hours/procedure	1.20	1.15	1.19	1.40	1.34
Average worked hours/procedure (across all months)	1.25				

Hospital B: Billable Volume Is 8-Hour Ventilator-Shifts*					
Procedure	Jan	Feb	Mar	Apr	May
Airway clearance treatments	1,836	1,700	1,727	1,104	1,168
Mechanical ventilation	693	624	505	456	421
Aerosol treatments	12,577	10,672	10,324	9,163	8,629
Total procedures	15,106	12,996	12,556	10,723	10,218
Worked hours	7,938	7,418	7,579	6,798	6,990
Worked hours/procedure	0.53	0.57	0.60	0.63	0.68
Average worked hours/procedure (across all months)	0.60				

*Mechanical ventilation data converted from ventilator-shift to ventilator-day

Table 5C. Actual Data From 2 Hospitals That Use Different Periods to Define Billable Procedures: Efficiency Expressed As Standard Hours/Worked Hour Instead of Worked Hours/Procedure

Hospital A: Billable Volume Is Ventilator-Days*					
Procedure	Jan	Feb	Mar	Apr	May
Airway-clearance treatments	387	390	443	269	343
Mechanical ventilation	4,152	3,612	3,180	2,769	2,715
Aerosol treatments	1,142	1,194	1,305	982	1,035
Standard hours	5,681	5,196	4,928	4,020	4,092
Worked hours	8,532	8,220	9,042	7,953	8,148
Standard hours/worked hours	0.67	0.63	0.55	0.51	0.50
Average standard hours/worked hours (across all months)	0.57				
Hospital B: Billable Volume Is 8-Hour Ventilator-Shifts*					
Procedure	Jan	Feb	Mar	Apr	May
Airway clearance treatments	606	561	570	364	385
Mechanical ventilation	2,080	1,872	1,514	1,368	1,263
Aerosol treatments	3,144	2,668	2,581	2,291	2,157
Total procedures	5,830	5,101	4,665	4,023	3,805
Worked hours	7,938	7,418	7,579	6,798	6,990
Standard hours/worked hours	0.73	0.69	0.62	0.59	0.54
Average standard hours/worked hours (across all months)	0.63				

*All data converted to procedure-hours

The foregoing discussion is based on the observation that many hospitals report billable volume based on days versus hours or shifts. In using such data to compare performance, decision makers can compare themselves only to respiratory care departments that bill using the exact same methodology. This could greatly reduce the number of facilities within the compare group and diminish the opportunity to identify best practices in departments that do not use the same methodology to count and determine billable units. Rather than eliminating facilities within the compare group, *the key to establishing appropriate benchmarking comparisons is the ability to accurately convert to a common format (eg, standard procedure-hours). The basis for this conversion is the critical calculation of procedure-days from procedure-hours.* Of particular interest is the calculation of ventilator-days from ventilator-hours, because of the huge work-load proportion this procedure represents and because this service (like oxygen service) is one that may be billed on the basis of hours or shifts rather than days. This study has demonstrated that the intuitively obvious formula of simply dividing total ventilator-hours by 24 or ventilator-shifts by 8 or 12 results in fairly substantial and avoidable errors. Use of simple linear regression equations makes these errors negligible for sample sizes above approximately 20 patients. To our knowledge, this is the first study to address the issue of conversion error in benchmarking.

The general finding that sample size affects error is not too surprising. In the Introduction we saw how error can range from 0% to nearly −100%. The smaller the sample size, the bigger the potential influence of errors contributed by individual patients. Both random and systematic conversion errors caused by the random interaction of procedure start times, durations, and the number of patients involved are acceptably small when the number of patients represented by the data are above about 20 (see Fig. 5).

Figure 3 shows that error is also affected by the duration of the reporting period used in billing. The finding that conversion error is affected by the reporting period (ie, hourly vs 8-h shifts vs 12-h shifts vs daily) is interesting, but perhaps not very important. The smaller the reporting period, the greater the error, because there are more chances for a fractional period to occur. For example, when converting service-hours to service-days, there are 23 possible service durations (1, 2, 3 . . . up to 23 h) that all result in a fractional service-day. Recall that fractional service days result in conversion error, because of the convention used by daily billing systems that fractional days are rounded up to the next higher day of service. Similarly, there are 2 possible service durations that result in a fractional day with an 8-hour billing system (ie, 1 shift or 2 shifts duration) and only one in a 12-hour billing system. (We will assume that a billing system based on shifts would round up actual service hours to whole shifts.)

We compared the regression equation from simulated data to the regression equation from actual data, as a reality check (see Fig. 5). The 2 equations showed very high r^2 values (0.999 vs 0.992). Squaring the correlation coefficient (r) gives the coefficient of determination, which ranges in value from 0 to 1.0 and may be interpreted as the proportion of the variation in the dependent variable (in this case, procedure-days) that is explained by the variation in the independent variable (in this case, procedure-hours). The regression equation for the simulated data gave a higher r^2 value and a smaller median error, perhaps because it represented data from a larger total sample (11,000; 1,000 from each of 11 different sample sizes, vs 3,218 actual patients), and perhaps because there may have been some errors in the actual patient data.

Figure 6 clearly shows that, with actual data, using regression to convert procedure-hours to procedure-days results in less error than simply dividing procedure-hours by 24. Indeed, Table 4 shows that, with simulated data, converting with division results in errors 10–50 times larger than converting with regression, depending on the reporting period.

A potential limitation of this study is that it used simulated data sets. On the other hand, resampling is an established statistical procedure that provides information for decision making when sampling actual data is impossible or impractical.

Conclusions

Estimation of billable procedure-days from data generated as procedure-hours or procedure-shifts is a critical operation in the creation of benchmarking indices. Using the intuitively obvious procedure of dividing procedure-hours by 24 or procedure-shifts by 8 or 12 results in underestimation of billable service-days and resultant underestimation of productivity. Use of linear regression prediction equations reduces these errors dramatically. These equations are:

procedure days = −0.237 + (0.049) (*procedure hours*) (for hourly billing)

procedure days = −0.205 + (0.372) (*procedure shifts*) (for 8-h shifts)

procedure days = −0.114 + (0.541) (*procedure shifts*) (for 12-h shifts)

Data reported for benchmarking purposes should represent a minimum of 20 patients, to reduce conversion errors.

A key finding of this study is that there is no completely accurate method to convert billing data to a common reporting unit such as procedure-days. This is because of the variability caused by the random interaction of procedure start times, durations, and the number of patients involved.

ACKNOWLEDGMENTS

We thank Larry Ramer RRT, Department of Respiratory Care, Mercy Medical Center, Canton, Ohio, for his assistance in preparing and conducting this study.

REFERENCES

1. Drucker PF. Management in the next society. Burlington, Massachusetts: Butterworth-Heinemann; 2002.
2. Blank S, Seiter C, Bruce P. Resampling in Excel. Arlington: Resampling Stats Inc; 2001:2–3.
3. Good PI. Resampling methods: a practical guide to data analysis. Boston: Birkhäuser; 1999.
4. Tweet AG, Gavin-Marciano K. The guide to benchmarking in healthcare. New York: Quality Resources; 1998.
5. Benchmarking in healthcare. Oakbrook Terrace, Illinois: Joint Commission on Accreditation of Healthcare Organizations; 2000.

Appendix
A BRIEF INTRODUCTION TO RESAMPLING THEORY*

The conventional analytic approach to inferential statistics requires that you understand complex formulas, and too often you can find yourself selecting the wrong formula. In contrast, resampling proceeds in stages that are easy to understand. Most problems can be tackled using the following 3-stage process:

1. Specify the universe to sample from (eg, random numbers, an observed data set, 0s and 1s).
2. Specify the sampling procedure (number of samples, sizes of samples, sampling with or without replacement).
3. Specify the statistic you wish to keep track of (eg, mean, standard deviation).

Resampling methods are typically used to address questions of statistical inference:

1. How much sampling error might there be in an estimate based on limited data (eg, establishing confidence limits)?
2. How likely is it that chance sampling error might produce a sample result as extreme as the observed sample (ie, hypothesis testing)?

With resampling, you try to answer these questions by drawing simulated samples (or "resamples") from the data themselves, or from a reference distribution based on the data, and observing how the statistic of interest in these samples behaves.

Early in the 20th century, when computers were unavailable to do the hard work of drawing all these samples, statisticians found they were able to represent the distributions of many sample statistics with calculated theoretical distributions of random variables. William Gossett, the statistician better known by the pseudonym "Student" under which he published, repeatedly dealt out sets of randomly drawn cards with prisoners' data written on them to see how the means of these samples were distributed. He used these simulated data in deriving his now-famous *t* distribution (which is used, among other things, to construct confidence intervals and to perform *t* tests).

Suitable theoretical approximations to sampling distributions have been created for a variety of sample statistics. However, they are not available for all statistics in all circumstances. Approximations require assumptions about how the data are distributed, and are generally good for large samples, but less accurate for small and imbalanced samples.

Resampling methods, including the bootstrap and permutation methods, can be used with virtually any sample statistic and do not rely on assumptions about how the data are distributed. Permutation methods for significance testing have the added advantage that they produce "exact" p values.

*Adapted from Blank S, Seiter C, Bruce P. Resampling in Excel. Arlington: Resampling Stats Inc; 2001:2–3.