TIME IN THE HOSPITAL IS NOT ALWAYS NORMAL: WHAT DOES ONE YEAR’S WORTH OF STUDIES TELL US?

Michael C Distefano, MD¹
Michael.distefano@ucdenver.edu
A. Chantal Caviness, MD MPH PhD²
alisonc@bcm.edu
Binita Patel, MD²
bxpatel@texaschildrens.org
Angelo P Giardino, MD, PhD, MPH³,⁴
apgiardi@texaschildrens.org

¹Section of Pediatric Emergency Medicine, University of Colorado
²Section of Pediatric Emergency Medicine, Baylor College of Medicine
³Section of Academic General Pediatrics, Baylor College of Medicine
⁴Texas Children's Hospital, Houston, Texas

Corresponding Author:
Angelo P. Giardino, MD, PhD
Professor and Section Chief, Academic General Pediatrics
Baylor College of Medicine
Senior Vice President/Chief Quality Officer
Texas Children's Hospital
6621 Fannin, Suite A135
Houston, Texas  77030
apgiardi@texaschildrens.org
832 824 1128
Background

Patient length of hospital stay (LOS) is an important indicator of patient care quality, particularly with regards to the efficiency and timeliness of care (Institute of Medicine, 2001; US Department of Health and Human Services, 2007). Although there is debate as to how LOS correlates with the quality of patient care (Clarke, 2002; Thomas, Guire, Horvat, 1997; Clarke & Rosen, 2001), it remains an important outcome measure in its own right. This is particularly true for interventions that may decrease or prolong patient treatment and hospital times. Duration of hospital stay has also been found to have an impact on patient satisfaction (Bursch, Breezy, Shaw, 1993; Thompson, Yarnold, Williams, Adams 1996; Bourdeaux, Autremont, Wood, Jones, 2004). Furthermore, since hospitalization carries risks of medical errors and hospital-acquired infection, LOS reflects the duration of exposure to these risks.

Given the focus on hospital LOS in many clinical research and outcome studies, it is critical that the correct statistical analyses are appropriately applied to the evaluation of this variable. As an example of the importance of analytic approach, our hospital evaluated the use of a specialized triage team (consisting of two nurses and a doctor) to help alleviate emergency department (ED) crowding. The hospital administration supported a two week trial of the new triage team and wanted to see improvements in patient wait times before supporting the permanent institution of the new program. When the data were first analyzed, it appeared that the work of the triage team had improved overall wait times but had not had any significant impact on departmental LOS for the sickest children. Specifically, when compared to a comparable two-week period, patients with critical acuity at triage during triage team times waited for evaluation only 7 minutes less (p=0.208, t-test) than patients with critical acuity conditions during non-atriage team times. It was not clear if these findings would jeopardize the continuation of the program. However, the results of the first analysis were carried out under the assumption that wait times were normally distributed. Further examination of the data indicated that the data were not normally distributed but positively skewed, making the mean an inappropriate measure of central tendency (Figure 1). Subsequent proper use of non-parametric tests, which assume no underlying distribution, led to the conclusion that the triage team had significantly decreased emergent patient wait times: the triage team had decreased median wait times for critical patients by 21 minutes (p=0.019, Mann Whitney U test). This example illustrates how the mean and assumption of normality in statistical testing can give an inaccurate picture of LOS data when the normality assumption does not hold true.

Since hospital LOS is an important measure for many outcome studies, and because analysis of this variable requires attention to the nature of its distribution, we sought to determine how the data have been approached and reported in published clinical studies. The purpose of this study was to describe the approach to evaluating the distribution, analysis, and reporting of hospital LOS results.

Methods

PubMed was searched for articles published between June 1, 2009 and May 31, 2010 using the MeSH term “length of stay” and keywords “length”, “stay”, “length of stay”. The search was limited to human subjects, the English language, and children. The principal investigator (ACC) reviewed all identified abstracts, or manuscripts if needed, to select the articles meeting study inclusion criteria. Studies where hospital LOS (any type) was an outcome (primary, secondary, or exploratory) were included if the following inclusion criteria were met:
published between June 1, 2009 and May 31, 2010, English language, human studies, and pediatric patients (aged 0-18 years). LOS was defined as any measure of time in the hospital, including the entire hospital stay, time in the pediatric intensive care unit (ICU), time in the ED, or duration of time in any other hospital unit.

The articles included in the study were reviewed independently by 2 of the authors (MCD and BP) using a standardized data collection instrument. Each article was reviewed to confirm that some type of hospital LOS was a study outcome and if so, to identify the statistical approach used to analyze LOS. It was noted whether: 1) the investigators examined the distribution of LOS before performing statistical analyses, 2) the LOS distribution was reported, and 3) information about LOS distribution was used to determine the analytic approach. In addition, the statistical methods (comparative or descriptive) and measures (i.e. medians, means) were recorded, along with the types of distributional assumptions made (i.e. parametric or non-parametric) by the investigators. The quality of studies was not otherwise assessed or graded. When complete manuscripts were unavailable for review, information was extracted from the abstract if possible; otherwise the information was classified as missing. The findings of these independent reviewers were checked against each other and discrepancies were resolved in consultation with the principal investigator.

Institutional Review Board approval was obtained for reporting of the triage study data but was not required for the review of published studies.

Data analysis was performed using SPSS 18 (SPSS Inc., Chicago, Illinois). The studies were described using frequencies of different pediatric disciplines (categorized primarily as surgical and medical), type of study (descriptive or comparative), LOS measures type (LOS in the ED, hospital, ICU, neonatal ICU), and unit of LOS measurement (minute, hour, day, week). The statistical approach to the LOS data was then described using frequencies of distribution evaluation, distribution types, measures of central tendency used, and methods of comparison (for comparative studies). In addition, we sought to compare the frequency of statistically significant LOS comparisons (defined as a p-value < 0.05) between studies using statistical tests that were appropriate or not appropriate for the reported LOS distributions.

3. Results

The initial PubMed search identified 795 articles for potential study inclusion and abstracts were reviewed for all of these. Of these, 581 (73%) articles did not meet the inclusion criteria for the following reasons: subjects not in the pediatric age range (407), length of stay not a study outcome (149), review article (19), case report (3), and commentary (3).

Published manuscripts were available for 212 of the 214 included studies (see appendix for a complete list of articles) and abstracts were reviewed for the 2 missing manuscripts. Of the 214 studies, 115 (54%) related to surgical issues and 99 (46%) to non-surgical issues (38 neonatology, 22 infectious disease, 14 critical care, 9 pediatric emergency medicine, 9 hospital medicine, and 7 other general pediatric issues). There was a single LOS outcome measure in 186 of the included studies: 157 LOS hospital, 13 LOS neonatal ICU, 11 LOS ICU, and 5 LOS ED. There were two LOS outcome measure in the remaining 28 studies: 23 LOS hospital and ICU, 3 LOS hospital and neonatal ICU, and 2 LOS ED and LOS hospital. There was one unit of LOS measurement in 211 studies: 1 LOS in weeks, 187 LOS in days, 18 LOS in hours, 1 LOS in minutes, and 4 not stated. Three studies used more than one unit of LOS measurement: 2 LOS in ICU hours and
hospital days, and 1 LOS in ED minutes and hospital days.

Of the 214 included articles, only 53 (25%) reported examining the distribution of LOS. Of these, 1 distribution was reported to be normal, 32 were reported to be non-normal, and the remaining 20 LOS variable distributions were not reported. The 1 study reporting a normal distribution of LOS was of neonates and LOS was measured in hours.

In 7 of the studies, LOS was categorized and not evaluated as a continuous outcome variable. None of these articles reported whether the LOS distribution was evaluated and none reported the nature of the uncategorized LOS distribution.

Of the 207 studies where LOS was not categorized, all reported some measure of hospital LOS central tendency (Table 1). The majority of reported LOS measures were means (49%) or medians (31%). Means were reported in 2 studies where LOS was reported to be non-normal in distribution. Twenty-two studies reported both means and medians despite 3 of these studies reporting a non-normal distribution of LOS. Seven studies reported ‘averages’ without defining whether these were means or medians.

Of the 207 included studies where LOS was not categorized, 142 of the studies were comparative and reported a statistical comparison of the LOS outcome (Table 2). Most statistical comparisons (43%) were non-parametric while 37% were parametric. Of the 61 studies using non-parametric methods, 24 reported that LOS was not normally distributed. One study reported results of parametric and non-parametric comparisons without reporting the distribution of LOS. In 18 of the comparative studies, the method of statistical comparison was unclear from the written report because the methods section reported that parametric methods would be used for normally distributed data and non-parametric methods for non-normally distributed data but the results did not specify the nature of the LOS distribution and method of statistical comparison. LOS was transformed to a normal distribution in 7 of the studies, 5 of which reported that LOS was not normally distributed.

Although we collected information about the statistical significance of the LOS comparisons, it was not possible to determine whether the appropriateness of the test performed had an influence on statistical significance. In the one study where the distribution was reported to be non-normal but a parametric test was used for statistical comparison, the result was significant. However, given that the majority of studies did not report the nature of the LOS distribution, it is not possible to determine how results compared between analyses that were and were not appropriate to the distribution.

4. Comment

In this review of pediatric studies published over a 1-year period with LOS as an outcome, it is clear that there is not a standard approach to the statistical analysis and reporting of this variable. Less than a quarter of studies reported that the distribution of LOS was evaluated. Even so, the majority of studies reported LOS means as measures of central tendency and compared LOS using parametric methods. In many studies it was not possible to determine which measures of LOS central tendency and comparative statistics were used for lack of clarity in the reported methods and results. It is clear that there is room for improvement in the analysis and reporting of LOS.

It is not possible to conclude from this study whether LOS is a variable that typically follows a normal or non-normal distribution. Although 32 of 53 LOS distributions
reportedly evaluated were non-normal in distribution, 1 of the 53 was normally distributed. The study reporting a normal distribution of LOS was performed in the neonatal population. Given that most prematurely born neonates are likely to spend considerable time in the hospital, it is possible that the duration of their stay follows a normal distribution. This is in contrast to patients who are admitted for conditions from which they typically recover quickly and go home after 1-2 days of hospitalization. The LOS for the latter situation is more likely to be skewed to the left, like ED wait time (Figure 1a). However, this is all speculative because it is difficult to draw conclusions about distributions that are rarely evaluated or reported in the literature.

The reliance on published reports for details of the analyses performed is a limitation of this systematic review. It is possible that the published manuscripts do not accurately reflect the statistical methods used by the investigators. Therefore, it is not possible to conclude with accuracy how the data were handled by all investigative teams. However, the very fact that the statistical methods were not always clearly reported in the published manuscripts is, of its own right, an important finding of this systematic review.

In addition, unpublished studies were not included in this systematic review, leaving the study open to publication bias. If unpublished studies were more likely to be negative (no statistical differences in LOS between groups) than published studies, and if differences were not found because LOS was not analyzed in a manner consistent with its distribution, the published studies could be biased toward studies using appropriate methods of analysis. Assuming this direction of publication bias, where published studies represent the most appropriate analyses and most transparent reporting, there is still plenty of room for improvement.

Furthermore, the reviewers were not blinded to the journals where the articles were published, to the authors of the studies, and to the outcomes of the studies when they extracted data for this review. However, it is unlikely that these factors had an influence on the data collected, given the quantitative (rather than qualitative) nature of the information extracted from each study. In addition, data extraction by 2 independent reviewers should have provided protection from the influence of journal, author, or outcome on the extracted data.

4.1 Recommendations

There are Uniform Requirements for Manuscripts Submitted to Biomedical Journals for the reporting of statistical data that were first published in 1998 by the International Committee of Medical Journal Editors (International Committee of Medical Journal Editors, 1988; Bailar & Mosteller, 1988). As related to statistical reporting, the requirements state the following: ‘Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results. When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals).’ Further, ‘Put general descriptions of methods in the Methods section. When data are summarized in the Results section, specify the statistical methods used to analyze them.’

At a minimum, these recommendations should be followed in the reporting of LOS data. In describing the statistical methods, authors should describe how they evaluated the distribution of any LOS variable. In addition, they should state how they used the information about the distribution to influence further analyses, both descriptive and comparative. In quantifying findings, authors should present measures of uncertainty that are appropriate to the statistical methods and LOS distribution. In
reporting measures of central tendency for LOS, means and standard deviations should be reported if LOS is normally distributed and medians and ranges if LOS is not normally distributed. Alternatively, if LOS is categorized, frequencies and 95 percent confidence intervals should be reported. In reporting comparisons of LOS between groups, preference should be given to reporting 95 percent confidence intervals for comparative measures. In addition, although general methods for describing or comparing LOS should be described in the methods, the authors should report the specific methods used in the results section. For example, the authors could state in the methods that they evaluated LOS distribution and used parametric or non-parametric methods based on those findings; and in the results section they would state how the LOS distribution was evaluated, state the actual distribution, and the specific statistical methods used to describe or compare the data. These steps should represent the minimal requirements for reporting the statistical analysis of LOS data.

In addition, LOS distribution must be evaluated before statistical analyses are performed. In evaluating LOS distribution, investigators can choose graphical or statistical methods. The most basic graphical method involves the LOS histogram, which can be visually compared to a normal probability curve. If LOS is normally distributed, its histogram should be bell-shaped and resemble the normal distribution. Using this method, our ED wait times and LOS are clearly not normally distributed, both being positively skewed, meaning that the right tail is longer than the left tail and that most values are to the left of the mean (Figure 1a and 1b). Alternatively, investigators can use a normal probability plot to assess whether LOS data are normally distributed. In this method, LOS data are plotted against a theoretical normal distribution where a straight line indicates that the LOS data are normally distributed. Most commonly used is the Q-Q plot, which plots the quantiles of two distributions against each other, allowing for their graphical comparison. Using Q-Q plots, it is clear that our ED wait time and LOS data are not normal in distribution because the plot line is not straight (Figure 2a and 2b). It is important to note that these methods of visual comparison can be difficult for small sample sizes, in which cases it is best to have an understanding of the distribution that would typically be expected for the variable in question.

LOS normality can also be evaluated using statistical tests, of which there are several options. The null hypothesis for these tests is that the sample being tested is from the same distribution as the normal distribution so that rejection of the null (typically with a p-value < 0.05) indicates that the sample in question does not follow a normal distribution. The most powerful options include the Shapiro-Wilk test (Shapiro & Wilk, 1965) and D’Agostino-Pearson test (D’Agostino & Belanger, 1990). Less powerful are the Anderson-Darling test (Anderson & Darling, 1952) and a modification of the Kolmogorov-Smirnov test, called the Lilliefors test (Lilliefors, 1967). Using the Lilliefors test on our ED wait time and LOS data, we obtained p-values of < 0.0001 for both, indicating that both are non-normal in distribution. With any of these statistical tests, it is important to note that the results are influenced by the sample sizes and as with any statistical test their results should be interpreted with caution if the sample is very small (less likely to find a significant deviation from normality) or large (more likely to find a significant deviation from normality).

Once the distribution of LOS has been determined, it should be used as a basis for determination of the correct statistical analyses. Further, the distribution should be reported in the published manuscript.
Statistical approaches to LOS data based on the underlying distribution should then be at the discretion of the investigators and should be conducted by or in consultation with a statistician. However, since we found in our review a preference for reporting of means, even when the data were known to be non-normal in distribution, we do recommend that the measure of central tendency should be appropriate to the actual distribution of LOS. That is, if the distribution is not normal, the measure of central tendency should be reported as a median or geometric mean, whereas an arithmetic mean can be used if LOS is normal in distribution. Specific recommendations for statistical approaches are otherwise beyond the scope of this article.

In summary, although used frequently as a measure of clinical efficiency, there is wide variability in the statistical approach to hospital LOS. We recommend that, at a minimum, the International Committee of Medical Journal Editors Uniform Requirements for Manuscripts Submitted to Biomedical Journals reporting of statistical data be followed. Furthermore, LOS distributions should be evaluated, reported, and used to inform the analyses performed. These steps should improve the quality of hospital LOS evaluation and reporting in healthcare quality studies.

Figure 1 – Histograms (and overlying normal curve) of wait time in minutes (1a) and emergency department length of stay in minutes (1b) for 2548 patients seen in the Texas Children’s Hospital Emergency Department.

---

1a – Wait time (minutes)

1b – Length of stay (minutes)
Figure 2 – Q-Q plots of wait time in minutes (2a) and emergency department length of stay in minutes (2b) for 2548 patients seen in the Texas Children’s Hospital Emergency Department

![Q-Q plots of wait time and length of stay](image)

Table 1: Measures of central tendency for 207 studies that reported evaluating hospital length of stay as a continuous variable

<table>
<thead>
<tr>
<th>Statistical Measure</th>
<th>Distribution Type</th>
<th>Total (N=207)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal (N=1)</td>
<td>Not Normal (N=32)</td>
</tr>
<tr>
<td>Mean</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>Mean &amp; Median</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Geometric Mean</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Average (not otherwise specified)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Not Stated</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

*Missing data from articles for which only abstracts were available for review
Table 2: Methods of statistical comparison for 142 studies comparing hospital length of stay as a continuous variable

<table>
<thead>
<tr>
<th>Statistical Comparison</th>
<th>Distribution Type</th>
<th>Total (N=142)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal (N=1)</td>
<td></td>
</tr>
<tr>
<td>Non Parametric</td>
<td>0</td>
<td>61</td>
</tr>
<tr>
<td>Parametric</td>
<td>1</td>
<td>53</td>
</tr>
<tr>
<td>Not Specified</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>Transformed</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Parametric and Non-Parametric</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Not Normal (N=30)</td>
<td></td>
</tr>
<tr>
<td>Non Parametric</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Parametric</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Not Specified</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Transformed</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Parametric and Non-Parametric</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Not Stated (N=109)</td>
<td></td>
</tr>
<tr>
<td>Non Parametric</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Parametric</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>Not Specified</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Transformed</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Parametric and Non-Parametric</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Missing* (N=2)</td>
<td></td>
</tr>
<tr>
<td>Non Parametric</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Parametric</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Not Specified</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Transformed</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Parametric and Non-Parametric</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

*Missing data from articles for which only abstracts were available for review
References


APPENDIX 1.
Additional articles use in the systematic review:


Appendix


54. elZein C, Hanhan Z, Massad M, et al. Surgical repair of anomalous coronary arteries arising from the opposite sinus of


92. Karsies TJ, Nicol KK, Galantowicz ME, Stephens JA, Kerlin BA. Thrombotic risk of recombinant factor seven in pediatric cardiac surgery: a single institution...


110. Letton RW, Worrell V. Delay in diagnosis and treatment of blunt intestinal injury does not adversely affect prognosis in...


130. Nankervis CA, Martin EM, Crane ML, Samson KS, Welty SE, Nelini LD. Implementation of a multidisciplinary guideline-driven approach to the care of the


169. Salas AA, Salazar HJ, Velasco VH. Haemophilus influenzae type b conjugate


186. Sola JE, Neville HL. Laparoscopic vs open pyloromyotomy: a systematic review


204. Villalona GA, Diefenbach KA, Touloukian RJ. Congenital and acquired


