The use of hospital admission data as a measure of outcome in clinical studies of heart failure

C. Metcalfea*, S.G. Thompsona, M.R. Cowieb, L.D. Sharplesa

a Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge, UK
b Department of Cardiac Medicine, National Heart and Lung Institute, Faculty of Medicine, Imperial College, London, UK

Received 20 May 2002; accepted 22 June 2002

Aim To describe and evaluate the methods currently used to summarize, depict, and analyse hospitalization data in clinical studies of patients with heart failure.

Methods and Results Studies published during 1999 and 2000 were reviewed if they reported hospitalization data collected during the follow-up of a defined group of adults who had been diagnosed as having heart failure. Of 130 studies identified, the majority considered only the first of any hospital admissions for each individual. When presented, statistical analyses in 70% of papers, and graphics in 75% of papers, were based purely on the first hospitalization. A minority of papers based analyses on the number of admissions (23%), or the amount of time spent in hospital by patients (14%) during the follow-up period.

Conclusions Data on the hospitalizations undergone by patients with heart failure are potentially informative about the disease burden to individuals and health services. The widespread practice of focusing a statistical analysis on the first admission a patient undergoes discards those aspects of hospitalization that are informative about disease burden: multiple admissions and length of stay. Statistical methods that incorporate these aspects appeared in a handful of papers, and the more widespread adoption of these is desirable.

© 2002 The European Society of Cardiology. Published by Elsevier Science Ltd. All rights reserved.

KEYWORDS Heart failure; hospitalization; mortality; disease burden; statistical methods; systematic review

Introduction

Heart failure imposes a heavy burden on individuals through a reduced tolerance of physical exertion and, in many cases, lengthy hospital admissions and short life expectancy. A sizeable burden on health services also arises through the cost of that hospital care.1,2 The hospitalizations undergone by patients with heart failure are frequently assumed by researchers to be a surrogate for worsening disease, for health related quality of life, and for demands on health service resources. This has made hospitalization a popular outcome for research into heart failure. An additional motivating factor is the frequent availability of hospitalization data from the databases of healthcare providers. The aim of this paper is to investigate whether the statistical methods currently employed by researchers are making full and appropriate use of hospitalization data.

This paper presents a systematic review of the presentation and analysis of hospitalization data, when used as an outcome, in recently published clinical research into heart failure. Hospitalization data are complex. Firstly, not all hospitalizations...
are necessarily relevant to heart failure progression. Secondly, hospitalization can occur repeatedly or not at all for each patient. Thirdly, each hospitalization has a duration. Fourthly, death may intervene preventing further hospitalizations. Lastly, there may be variable length of follow-up across the patients in the study. These difficulties in interpreting analyses of hospitalization data are discussed in the context of the reviewed studies.

Methods

Selection of study articles

The aim was to identify papers in the recent medical literature that were based upon data describing hospitalizations in patients with heart failure. Consequently, the search for relevant articles was limited to those published during the years 1999 and 2000. The Medline database was searched, with the May 2001 update in place. Table 1 describes the search strategy in detail.

Abstracts were inspected to identify those papers reporting hospitalization data, collected during the follow-up (prospective or retrospective) of a defined group of adults who had been diagnosed as having heart failure. If several papers described different analyses of the same data, all were included in the review. If they described the same analysis, only the paper giving the greatest amount of detail was included. Papers were excluded from consideration if they were not written in a language using the Roman alphabet (three papers). One hundred and thirty articles were retained for inclusion in the review (Table 1); a complete list is available from the authors.

It is likely that the search strategy in Table 1 has missed a small number of relevant papers. This does not limit the conclusions that can be drawn from this review, as there is no reason to suppose that any unidentified papers would be qualitatively different from those that have been included in the review. In contrast to reviews of treatment effects, reviews of methodology need to be systematic and representative, but not necessarily exhaustive.

Information collected

For each article information was collected onto a 37-item checklist covering the type of study, the average follow-up time, the type of hospitalizations recorded, the number of participants, whether information on mortality was included, and the methods of summarizing, depicting, and analysing the hospitalization data. Only unusual techniques required free-text recording.

The description of statistical methodology and the presentation of statistical analyses in the published papers sometimes allowed room for doubt in deciding what statistical techniques had been used in a study. Efforts were made to ensure that the results of this review could be replicated. Firstly, the first author (C.M.) read all the papers twice, the papers being read in a different order after an intervening period of 2 weeks. Any discrepancies were resolved by a third reading. Secondly, using the checklist, a random sample of 15 papers were re-assessed by a colleague who was blind to the initial assessment and otherwise not involved with the review. Agreement was generally high, with perfect agreement on 17 of the 37 checklist items, and disagreement on only one or two papers for a further 18 items. There was disagreement on five
papers with regards to whether total time spent in hospital had been reported, or just the length of the initial admission. Consequently there was disagreement on four of those papers as to whether a t-test had been used to compare total days spent in hospital between patient groups. A further re-reading of the five papers did not clarify the matter, and it has been assumed here that length of the first admission was being reported.

**Results**

**Description of papers**

The papers came predominantly from cardiovascular journals, with a number from general medical journals. Fourteen such journals accounted for 78 (60%) of the 130 papers. One hundred and twenty five of the papers were published in English, four in Spanish, and one in German.

**Description of studies**

Amongst the 130 papers, 39 described randomized trials, 18 intervention studies without random allocation, 55 observational studies, and a further 18 studies were purely descriptive of the hospitalization of heart failure patients within a particular context. Common choices of outcome were hospitalizations for worsening heart failure (84 studies), all hospitalizations (72 studies), hospitalizations for cardiac reasons (34 studies), and unplanned hospitalizations (12 studies). The number of patients in the studies varied from six to more than 90 000, the median being 244. Follow-up varied from 2 weeks up to 7 years, with a median of 11.5 months. In 65

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Descriptive statistics used</th>
<th>Number of papers (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No summary statistics for hospitalization data</td>
<td>7 (5)</td>
<td></td>
</tr>
<tr>
<td>Admissions*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion of patient with one or more admissions</td>
<td>98 (75)</td>
<td></td>
</tr>
<tr>
<td>Proportion of patients with one or more admissions or who have died</td>
<td>41 (32)</td>
<td></td>
</tr>
<tr>
<td>Proportion of patients with given numbers of admissions</td>
<td>11 (8)</td>
<td></td>
</tr>
<tr>
<td>Mean number of admissions</td>
<td>47 (36)</td>
<td></td>
</tr>
<tr>
<td>Geometric mean number of admissions</td>
<td>1 (1)</td>
<td></td>
</tr>
<tr>
<td>Median number of admissions</td>
<td>3 (2)</td>
<td></td>
</tr>
<tr>
<td>Time in hospital*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean days in hospital</td>
<td>22 (17)</td>
<td></td>
</tr>
<tr>
<td>Proportion of time alive spent in hospital</td>
<td>4 (3)</td>
<td></td>
</tr>
<tr>
<td>Geometric mean days in hospital</td>
<td>1 (1)</td>
<td></td>
</tr>
<tr>
<td>Median days in hospital</td>
<td>4 (3)</td>
<td></td>
</tr>
<tr>
<td>Other approaches*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean time to first admission</td>
<td>5 (4)</td>
<td></td>
</tr>
<tr>
<td>Mean length of individual admissions</td>
<td>5 (4)</td>
<td></td>
</tr>
<tr>
<td>Median time to first admission or death</td>
<td>3 (3)</td>
<td></td>
</tr>
<tr>
<td>Median length of individual admissions</td>
<td>3 (3)</td>
<td></td>
</tr>
<tr>
<td>Other*</td>
<td>2 (2)</td>
<td></td>
</tr>
</tbody>
</table>

*Studies often present more than one summary statistic.
studies the length of follow-up was fixed for all patients. Twenty-eight studies did not present information on mortality.

**Descriptive statistics**

Table 2 presents the number of papers making use of different descriptive statistics. The most common way of summarizing hospital admissions data was to present the number or proportion of individuals experiencing one or more admissions. Summaries of this ‘any admission’ outcome were presented in 98 papers. Forty-one papers presented a ‘composite’ outcome, which identified individuals who had either been admitted or had died during the follow-up period. In total, 104 papers presented one or both of these outcomes. Eleven studies presented at least some information on how many patients experienced particular numbers of admissions, hence potentially preserving more information than a simple dichotomy. More commonly, 47 papers presented the mean number of admissions experienced per patient over the follow-up period or, equivalently, the total number of admissions.

Fewer studies presented measures of the time spent in hospital. Twenty-two studies presented the mean number of days spent in hospital per patient over the follow-up period, or equivalently presented the total number of days spent in hospital. On the same lines, four studies presented the proportion of days alive spent in hospital.

**Modifications of the above methods**

A small number of papers used modifications of the above statistics, generally for one of two reasons. Firstly, in a given study patients are observed for different amounts of time due to staggered recruitment or mortality. To counter this, two of the 41 papers presenting the proportion of individuals succumbing to hospitalization or death also estimated the proportion per year. Similarly, 10 out of the 47 papers presenting the mean number of admissions estimated the mean number per patient year.

Secondly, in a group of individuals the majority will experience very few hospital admissions and few days in hospital, while a small number will experience repeated admissions or admissions of long duration. In acknowledgement of this a few studies used summary statistics based upon geometric means or medians, these giving less weight to extreme high values compared to the usual arithmetic mean (Table 2).

**Graphical presentation**

Table 3 presents details from the 51 of the 130 papers presenting their hospital admissions data graphically. Twenty-six of these 51 papers (51%) presented Kaplan–Meier curves for the cumulative probability of surviving and remaining free of hospital admissions over the follow-up time. Nine papers plotted the Kaplan–Meier curves for the cumulative probability of remaining free of hospital admissions, with death treated as censoring. One or both of these graphics appeared in a total of 33 papers, 65% of the 51 papers presenting a graphic of hospitalization data. Four papers used a bar chart, or similar, to depict the risk of any hospitalization during the follow-up period. The number of patients having specific numbers of admissions was presented in six
papers. The mean number of admissions per patient was presented graphically in four papers. In two cases this latter graphic was given as the accumulating total of admissions for each risk group over the follow-up period. These three groups of graphics were used to show how the risk or means varied over different treatment or patient groups, or in the same patient group over time.

In only four papers was information on the days a patient spent in hospital over the follow-up period presented graphically.

**Inferential statistics**

As shown in Table 4, of the 105 papers that did present statistical analyses, 52 (50%) employed Cox’s proportional hazards regression. Of these, 33 used a composite outcome of time to first hospital admission or death, and 31 used the time to the first hospital admission as the outcome.

Thirty-seven studies compared the proportions of individuals experiencing one or more hospitalizations using either a chi-squared test or logistic regression. In four of these studies logistic regression was also used to compare the proportion of individuals with multiple admissions. Eleven studies used logistic regression to compare the proportion of individuals experiencing the composite outcome of hospital admission or death. In total, 74 studies based their analysis on an outcome incorporating only the first observed admission, with or without combination with death (70% of the 105 papers including any analysis).

Mean number of hospital admissions per patient, and mean number of days spent in hospital per patient, were compared between groups using parametric tests (e.g. t-tests and ordinary regression) in 16 and 10 studies respectively. Cognisant of the skewed distributions of these hospitalization measures, two of these studies log-transformed their data before using these tests that assume a normal distribution. An alternative approach, distribution free tests (e.g. the Mann–Whitney U test or the Wilcoxon Signed Ranks Test), were used to compare the number of admissions undergone by each individual in six studies, and the number of days spent in hospital by each individual in six studies.

Less common techniques were, for the most part, focussed upon just the first admission of a patient. Two included more information: a comparison of the mean number of admissions in different risk groups using a mean difference with bootstrap confidence interval, and the duration of hospital stay with analysis based upon the Kaplan–Meier corrected days in hospital.

**Discussion**

**Findings of the review**

The studies included in this review encompass a broad range of aims, and vary in the data collected to address those aims. Different types of hospitalization are recorded, and the follow-up time varies enormously. In contrast, a degree of uniformity is
observed in the approaches taken to present and analyse the hospitalization data. Despite patients diagnosed with heart failure being potentially subject to more than one hospital admission during the course of their illness, only a minority of researchers attempt to incorporate this recurrence in the presentation of summary statistics, in the graphical presentations, or in the statistical analyses of their data. Similarly, the duration of those hospitalizations is neglected in the vast majority of studies.

Which hospitalizations should be examined?

Studies vary in the hospitalizations considered. If all hospitalizations are considered, the observed effects of treatment or risk factors may be diluted due to hospitalizations that occur independently of heart failure progression or treatment. However, if only hospitalizations recorded as due to heart failure or cardiac causes are considered, a number of admissions in fact related to the disease process or treatment side effects may be omitted. Many studies side step this debate by including analyses both of all hospitalizations and of hospitalizations due to worsening heart failure.

Incorporating the recurrence and the duration of hospitalization

This review found that statistical and graphical methods based on the patient’s first experience of an event during the follow-up period are the most popular for analysing hospitalization data in studies of heart failure. Consequently, both the recurrence and duration of admission are being discarded in the majority of published analyses. Yet it is these aspects that are likely to relate to the disease-burden on the individual and on health services.2

Information on multiple admissions and length of stay can be incorporated into the analysis of hospitalization data. Two useful summary statistics appeared in a small number of the reviewed papers: the mean number of admissions, and the mean number of days in hospital. Together these two statistics are sensitive to the occurrence of all admissions and the duration of those admissions. When comparing the mean number of admissions, or the mean number of days in hospital between treatment or risk groups, the skewed distributions of these measures prevent a blind reliance on t-tests and related techniques. These assume the data are normally distributed in order to calculate accurate significance tests.

Similar concerns with cost data have caused the increasing adoption of bootstrap techniques in economic evaluations of treatments. Bootstrapping is a technique whereby multiple samples are randomly selected from the observed data and the statistic of interest calculated for each sample. The variation of these sample statistics is taken as an estimate of the size and nature of the sampling error and allow comparisons between groups to be based upon familiar statistics such as the mean difference, while making allowance for the non-normal distribution of hospitalization measures in the calculation of confidence intervals and P-values.6 Estimates of the effects of risk factors are provided on the original, not a transformed, scale (thus aiding interpretation) and the complexity of the models that can be considered is not restricted as it is for conventional non-parametric statistics. One study in this review used these techniques for measures of hospitalization.4 Bootstrapping is available in commonly used statistical software such as S-Plus and Stata.

A graphic of the accumulating number of admissions, or days in hospital, over the follow-up period can be constructed, this approach having been used by Stewart and colleagues.7 Such a graphic can give some indication of the period over which a risk factor or treatment is having the greatest effect on hospitalization. Note however, that the rate of accumulation will decrease over the follow-up period, at least in part due to individuals dying or being lost to observation. Figure 1 presents an example using data from the Bromley Heart Failure Study.8 Those patients assessed as having illness of greater severity at baseline are observed to undergo a higher rate of hospitalization (and hence a more rapid increase in the number of admissions), but this...
only starts to become apparent after about 6 months. In this case the groups being compared contain new-equal numbers of patients. Were this not the case comparison would be facilitated by plotting the mean number of admissions per patient for each group.

**Death and hospitalization**

There are advantages in using dichotomous outcomes based on the first admission only. Firstly, in clinical trials, it may be that treatment changes following an admission will violate a study protocol. Use of only the first admission in an analysis avoids the difficulties of interpreting the meaning of admissions that occur after such a violation. Secondly, a dichotomous outcome allows the use of the log rank test or proportional hazards regression, both of which incorporate varying follow-up periods across patients.

The binary distinction between any and no admissions is often elaborated to distinguish patients who were admitted or who died, vs those who survived and avoided admission until the end of follow-up. One motivation for this is that the log rank test and proportional hazards regression will give biased results if individual lengths of follow-up are associated with the likelihood of an individual being admitted. This is likely to be the case; severe disease will make an individual more susceptible to admission, but more likely to have their follow-up terminated by death prior to admission being observed. Consequently, examining admission without taking account of mortality will likely give a biased estimate of the effect of a risk factor on admission. Combining death and hospitalization into a single outcome is one method of avoiding this bias, but at the expense of obtaining an outcome measure with a very different meaning to hospitalization alone. Unless caution is exercised, this composite outcome also has the potential to mislead.

**Incorporating variable lengths of follow-up**

The high mortality rate among patients with heart failure results in a variable length of follow-up of patients, even if a standard length was intended. Assuming a constant rate of admission, those individuals observed for longest will be subject to a greater number of hospitalizations. A small number of studies attempted to obtain a comparable outcome from each patient by calculating individual rates of hospitalization, admissions per year for example. Windeler and Lange point out that since the rate of admission is, in fact, likely to vary over the disease course, such annual rates will remain dependent upon the period of time over which a patient was observed.

We exemplify this using data from the Bromley Heart Failure Study. Figure 2 shows the hospitalizations undergone by 297 new cases of heart failure subsequent to their first episode. Each bar depicts the proportion of patients still under observation in a given month, who spend at least 1 day in hospital during that month. Applying the simple approach described above with these data, an individual observed for 6 months would have an estimated yearly rate based upon the typical high rates of admission in the first 6 months following the first episode. In contrast, if we had followed that individual for 12 months, a lower annual rate would be calculated due to the lower prevalence of admission in the second half of that first year. In this way, simplistic adjustments for variable follow-up can be extremely misleading.

In fact, in most cases modest variation across individuals in the length of follow-up, with a more or less equal distribution of follow-up times between the groups considered, will have little effect on the observed differences in mean admissions and mean hospital days. In this situation there is no need to take individual lengths of follow-up into account in the analysis. Where the distribution of follow-up times differs between groups, perhaps due to differing mortality rates, caution must be exercised. As discussed above, the hospitalization rate is likely to vary as the follow-up period progresses, this preventing any straightforward ‘adjustment’ of the individual rates to achieve comparability. To allow consideration of the effects of mortality on the observed differences
in hospitalization measures, the latter should be accompanied by, but not combined with, the estimated ratio of the mortality rates. This will give an informal indication of those situations where the comparison of hospitalization measures is being coloured by a difference in the mortality rate. Multi-state models are a more formal approach to learning about the relationship between hospitalization and mortality, but they are of a higher order of complexity and require substantial statistical and computational expertise.

**Conclusion**

A greater understanding of the effect of treatment and risk factors on hospitalization, and consequently upon the disease burden for patients and health services, requires that more than just the first occurrence of hospitalization be considered. The widespread reliance upon such restricted analyses leaves us ignorant of the recurrence and duration of hospitalizations. In only a small number of the studies reviewed was information on the number of patients undergoing any hospitalization complemented with information on the mean number of admissions or the mean number of days in hospital. In addition, the use of Kaplan–Meier curves to depict mortality, or time to first admission or mortality, was seldom accompanied by graphics depicting the rate of accumulation of admissions by patient groups over the follow-up period. These are simple methods yet lead to succinct and relatively comprehensive summaries of hospitalization in different patient groups.

**Acknowledgements**

Chris Metcalfe is funded by a Medical Research Council/National Health Service Eastern Region Special Training Fellowship in Health Services Research and a Raymond and Beverly Sackler Studentship. We are grateful to Angela Wood, Research Associate at the MRC Biostatistics Unit, for her help in assessing the reliability of this review.

**References**