Available online at www.sciencedirect.com
ScienceDirect
Physica A 379 (2007) 235–240
PHYSICA A
www.elsevier.com/locate/physa

A power law distribution in patients' lengths of stay in hospital

A. Hellervik*, G.J. Rodgers

Department of Mathematical Sciences, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK

Received 26 September 2006; received in revised form 10 December 2006
Available online 18 January 2007

Abstract

The distribution of patients' lengths of stay in English hospitals is measured by using routinely collected data from 11 years. It is found to be well approximated by a power law distribution spanning over more than three decades. To explain this observation, a theoretical resource allocation model is presented. It is based on iterative long-term scheduling of hospital beds, and its main assumption is that future beds are allocated preferentially. This represents a situation where different parts of the health care system compete for resources, with bargaining powers proportional to current resource levels.

Keywords: Health care; Power law distributions; Preferential attachment; Resource competition; Rate equation

1. Introduction

As a simple indicator of hospital activity, length of stay (LOS) has been used for many purposes on several levels of management. It can be used to monitor quality and appropriateness of resource allocations both internally in hospitals and at a national level [1]. LOS is clearly related to resource consumption and can thus be used for planning and efficient sorting of patients into groups [2].

However, it is well known that the distribution of LOS is very asymmetrical with a broad tail, which makes many statistical estimates less robust. In a small sample, it can be very difficult to track changes over time since a single outlier can significantly change the mean and variance.

A review of different commonly used distributions for LOS is presented in Ref. [1]. Lognormal, Weibull, and Gamma distributions are compared with empirical data from a number of countries. The results show that both Weibull and lognormal are useful, and which one to choose depends on the specific country. In contrast to the study presented in this paper, the authors considered specific diagnostic groups separately, here we are only concerned with the aggregation of all diagnoses.

In Ref. [3] it is pointed out that resources allocated can be excessive for a small subset of users and at the same time, others get few resources or even none at all. The skewness in the distribution of LOS reinforces this picture.

*Corresponding author.
E-mail addresses: Alexander.Hellervik@brunel.ac.uk (A. Hellervik), G.J.Rodgers@brunel.ac.uk (G.J. Rodgers).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2006.12.045

There have been some attempts to explain the underlying cause for the skewed distributions. One example is described in Ref. [4], where a model, employing the transitions between different states of severity in the illnesses of patients, is suggested.

The hypothesis brought forward in this paper is that LOS is mainly determined at the resource allocation stage. This would mean that the competition for resources between different groups of patients is the significant underlying cause of the observed distributions. The competition takes place on all levels, between hospitals, between departments and between different types of patients within the same department.

We find that the distribution of LOS is well described by a power law, which is a common observation in both natural and artificial systems and no specific underlying generative model can be inferred from the observation [5–7]. However, the association between LOS and resource allocation, makes it possible that a competitive multiplicative process is creating the fat tailed distribution. Such processes are well known and historically trace back to the Simon model [8]. More generally, it has been shown that exponential growth combined with random survival times is sufficient to produce power law distributions [9]. This broad class of models includes the one proposed in this paper.

This paper is organised as follows: In Section 2 the empirical findings are presented, in Section 3 the model is formulated, analysed and compared to the empirical data, and in the final section the results are discussed.

2. Empirical findings

The Hospital Episode Statistics (HES) database contains routinely collected data on patients in the English National Health Service (NHS). The NHS is managed by the Department of Health and is funded by taxpayers and provides care mostly free of charge to the individual patients. For the analysis presented in this paper, an extract of all patient records from 1989–1990 to 2002–2003 is considered. (The NHS years do not follow calendar years, instead they span between April 1 and March 31). Some patients admitted before 1989 are present in the data set, but no discharge data exist for patients before 1991–1992. LOS is calculated by taking, for every episode, the difference between the discharge date (which can also mean the date of transfer or death) and the admission date. The data set contains 130 920 010 *valid* values for LOS. This is about 95% of the recorded episodes.

Here we consider all patients in the data set with a valid entry for LOS, which means that data from all diagnostic groups are aggregated. In Fig. 1 a double logarithmic histogram for LOS is shown. It is obvious that the distribution of LOS is very fat tailed and that it can be described fairly well as a power law over several decades. This means that $P[LOS = x] \sim x^{-\gamma}$. One of the most reliable ways to find a value for the exponent γ is to use the maximum likelihood estimate [5],

$$\gamma = 1 + \frac{n}{\sum_i^n \ln(x_i/x_{\min})}, \tag{1}$$

where x_i are the empirically measured LOS, $x_{\min}$ is a lower cutoff for the power law, and n is the sample size. LOS-values less than $x_{\min}$ are not used in the estimation of the exponent. We choose $x_{\min} = 5$ (giving $n = 33\,956\,629$) and obtain $\gamma = 2.12$.

3. Model

Our model is formulated from the viewpoint of a scheduler, which means that only future hospital episodes are taken into account. Its essential ingredient is that planning is taking place iteratively at the same time as parts of the plan are being realised. We have b beds and a planning window of length T days. This gives a total number of available time slots $S = bT$, where every time slot corresponds to one day of hospital stay for one patient. From now on we do not consider the actual beds, just the fact that we have a pool of S time slots that can be used for planning future hospital episodes. These time slots are allocated between J planned episodes. We assume an arbitrary, initial allocation, where s_j denotes the number of time slots allocated to episode j with $j = 1, 2, \ldots, J$. The initial condition is constrained by the total number of available slots, $\sum_{j=1}^J s_j = S$. As time passes in the real world, the planning window is shifted forward which adds new time slots that have to be