New methods for forecasting bed requirements, admissions, GP referrals and associated growth

Dr Rod Jones (ACMA)
Statistical Advisor
Healthcare Analysis & Forecasting
Camberley

For further articles on the topic of bed occupancy and planning please refer to: www.hcaf.biz

hcaf_rod@yahoo.co.uk

Introduction

In the early 1990's the author was involved in the planning for a new hospital. An external 'expert' was consulted to derive demographic-based forecasts of demand for 10 years into the future. This forecast of admissions was exceeded within two years!

Thus began my search for reliable methods for estimating future healthcare demand and the required number of overnight and day surgery beds which enable a hospital to be both efficient and effective.

The forecasting of future healthcare demand has relied heavily on the assumption that demographic growth is the main force behind growth in admissions & GP referral and that length of stay efficiency is reducing bed demand. Both of these assumptions fail in the real world[1]. This paper summarises alternative approaches to address this limitation.

Why Demography Fails to Forecast Activity

Demographic forecasts (using present day activity split by age band, divided by the current population in that age band and then multiplied by the future population in the respective age bands) of future healthcare activity has been the unquestioned methodology of choice for over 30 years.

Rightly or wrongly, large hospitals are built based on this fundamental assumption. The author is not aware of any publication which has actually proved the validity of this approach; however, by default a precedent has been established. It would seem that all concerned have simply assumed that this is the case. It is the author's experience that attempts to suggest to healthcare managers that demography may give unreliable estimates of future demand are met with great scepticism. Such an ideological framework may therefore act to limit the exploration of alternative concepts[2].

Anyone who has attempted to follow trends in inpatient care over a long time period will be acutely aware that many diagnoses do follow unique trends which are unrelated to any underlying demographic change. These 'unexplained' trends have largely been ignored as data artefacts due to poor coding processes within hospitals or simple random variation. In other words, the expectation of gradual and roughly linear demographic growth leads to the pre-conceived idea of what is considered a 'normal' trend.

Recent research is however beginning to question the validity of this approach. A study in Scotland looking at trends in emergency admission for older people concluded that demographic growth may only explain as little as 10% of the actual long-term increase. Other studies have noted that in some specialties the long-term trends appear to follow cycles while growth in medical admissions appears to involve

1 Jones R (2009) Building smaller hospitals. British Journal of Healthcare Management 15(10), 511-512.

2 Refer to the following paper for the references applicable to this section: Jones R (2010) Can time-related patterns in diagnosis for hospital admission help identify common root causes for disease expression? Medical Hypotheses 75(2): 148-154.

step-changes. Admissions involving 'injury' have been shown to exhibit a unique growth pattern.

A recent study comparing trends in emergency admission for a wide range of diagnoses against an assumed linear growth model showed that over 58% of total activity was subject to a high degree of 'special cause' variation and that only 18% could be said to approximate straight line growth in the absence of 'special cause' variation. Special cause variation will include all weather and environmental factors (such as viruses and other infectious agents) influencing the development and expression of poor health and all other sources of non-linear growth.

The linkage between human health and the environment is widely appreciated. Long term cycles in human longevity, gender ratio at birth and susceptibility to disease have been demonstrated to be linked with the approximate 11 year cycle in solar flare intensity. Temperature is a fundamental regulator of human health and hence global warming will both increase and decrease the incidence of various diseases and conditions. Global warming itself follows a complex series of short, medium and long term cycles and hence hospital admissions for particular conditions should exhibit the same behaviour. Most infectious diseases show a periodic nature with the time between successive outbreaks being specific to each disease. Even common conditions such as appendicitis have shown long term trends which are unrelated to demography and may involve links to factors causing inflammation such as air pollution and viral infection.

Of even greater importance is the fact that the lifetime usage of hospital beds by individuals is concentrated in the last year of life (more specifically the last 6 months) and that the average bed usage is almost independent of age at decease. Hence for the majority of non-elective admission events it is progression to death which drives bed usage and not demography per se[3]. Hence the exact equation predicting the number of admissions has a demographic component which is largely applicable to elective admission and a component which largely depends on death (excluding instantaneous causes for death) for non-elective admissions. Child birth is a special case which reflects complex issues of ethnicity, socio-demographic status and societal trends regarding career and home ownership.

In addition to the above issues, recent research has identified that a particular 3 to 6 year cycle may regulate admissions to certain medical diagnoses. This research is still in the early stages, but if it is correct, then the issue of medical emergency admissions will require a completely different approach to the forecasting of future demand. Updates regarding this research can be found at www.hcaf.biz

What Occupancy Level Is Appropriate?

The issue of hospital occupancy is poorly understood. The occupancy level applicable to each hospital is unique and requires an understanding of the role of variation in demand[4]. Refer to a variety of articles available at www.hcaf.biz for further details.

3 Jones R (2010) Myths of ideal hospital size. Med J Australia 193(5): 298-300.

4 Bain C, Taylor P, McDonnell G, Georgiou A (2010) Myths of ideal hospital occupancy. Med J Australia, 192:42-43.

The Importance of Specialty of Care

In the UK a Finished Consultant Episode (FCE) records the length of time a patient stays under the care of a single consultant within a specialty. More complex patients can move between specialties, i.e. two or more FCE. When determining bed requirements it is vitally important to assign the length of stay to the correct specialty. Hence forecasting of bed requirements will use FCE (or an FCE equivalent in other countries).

Beds and Bed Equivalents

Bed requirements are best forecast using past trends in bed days (projected forward) rather than attempting to multiply FCE by average length of stay (ALOS) to estimate bed days.

Disadvantages of FCE x ALOS

- FCE are not a basic unit of healthcare resource demand since they do not measure time
- For emergency admissions FCE inflation (usually 0 LOS admissions) clouds the issue[5]
- Average specialty LOS is derived by dividing bed days by FCE, hence, multiplication back to give bed days only introduces errors and bias due to misspecification
- Average LOS has little real meaning when compared against the LOS distribution from which it is calculated
- Computer simulation shows that the calculated average LOS is highly dependant on statistical variation in the underlying age distribution of the arriving patients
- The calculated average LOS is likewise highly dependant on the relative proportions of 0 LOS and >7 days LOS patients
- Even HRG-adjusted LOS is subject to methodological bias and high uncertainty in particular HRG's where it is assumed that every hospital has patients at the national average age distribution and at the national average mix of conditions within that HRG[6]
- Attempts to forecast future average LOS are of dubious validity
- Forecasts based on FCE shown high statistical uncertainty due to the inherent statistical variability in healthcare demand. This leads to both an uncertain current and future average. Use of single year values therefore have the potential to give very high bias in future estimates

[5] Jones R (2009) Length of stay efficiency. British Journal of Healthcare Management 15(11), 563-564.
Jones R (2010) Benchmarking length of stay. British Journal of Healthcare Management 16(5), 248-250.
[6] Jones R (2008) Limitations of the HRG tariff excess bed days. British Journal Healthcare Management 14(8), 354-355.
Jones R (2008) A case of the emperor's new clothes? British Journal of Healthcare Management 14(10), 460-461.
Jones R (2008) Limitations of the HRG tariff the trim point. British Journal Healthcare Management 14(11), 510-513.
Jones R (2008) Costing orthopaedic interventions. British Journal of Healthcare Management 14(12), 539-547
Jones R (2009) Limitations of the HRG tariff efficiency. British Journal of Healthcare Management 15(1), 40-43.
Jones R (2009) Limitations of the HRG tariff the RCI. British Journal of Healthcare Management 15(2), 92-95.
Jones R (2009) Limitations of the HRG tariff local adjustments. British Journal Healthcare Management 15(3), 144-147

Advantages of Bed days

- Bed days are basic units of healthcare resource demand (i.e. they measure time) and can be diminished by bed day equivalents in other settings
- Reduced average LOS for specific groups of patients are best expressed as an overall saving in bed days rather than cumbersome attempts to adjust the overall specialty average LOS
- The same demand as measured by FCE when expressed in bed days shows lower levels of statistical variation
- 0 LOS admissions are treated as a special form of bed day in that they express additional daytime rather than midnight occupancy
- Shifts from overnight to day case are also best identified as bed days since the past trend in bed days associated with each procedure can be determined and the forecast average removed from the specialty total

Establishing trends in demand

Growth in demand is best forecast by comparing demographic-based estimates (for elective) and trends in total deaths (for non-elective) against existing trends in demand which may also be reflecting additional technological factors.

- Demographic growth is only one of a number of forces determining the ultimate expression of healthcare demand
- Trends in medical technology, i.e. joint replacement technology in the early 1990's, can have much greater effects than demography
- Medical admissions have been shown to exhibit a periodic step increase with an associated step increase in total bed days[7]
- The chosen growth rate should therefore be a balance between past trends and demographic change
- A method similar to traditional access rates but using bed days is best used to estimate growth in demand for overnight beds
- Growth in basic or raw outpatient demand expressed as GP referrals is a special case with up to seven different growth mechanisms.
- Growth in inpatient demand is not as variable since the raw demand has been sent through a filter mechanism, i.e. consultant review

The traditional method for estimating growth using access rates are flawed and gives answers which in practice are subject to unknown bias:

Access rates are based on FCE

- For elective admissions these are based on activity rather than demand – where demand is activity adjusted for the change in the waiting list
- For emergency activity the above mentioned FCE inflation varies from hospital to hospital due to consultant rotas, emergency assessment units

[7] Jones R (2009) Trends in emergency admissions. British Journal of Healthcare Management 15(4), 188-196.
Jones R (2009) Cycles in emergency admissions. British Journal of Healthcare Management 15(5), 239-246.
Jones R (2009) Emergency admissions and hospital beds. British Journal of Healthcare Management 15(6), 289-296.
Jones R (2009) Emergency admissions and financial risk. British Journal of Healthcare Management 15(7), 289-296.

and different interpretation of an inpatient admission – which in some cases may more correctly be interpreted as 'urgent' outpatient-type attendances

Access rates ignore the important contribution from private medicine

- Areas of high affluence will therefore tend to have a lower elective access rates which in practice is hard to estimate other than via the actual level of elective demand experienced at each hospital, i.e. the catchment population of a hospital is described by those patients who require treatment at that hospital
- For these reasons reported access rates differ significantly depending on the chosen time frame and location, hence, in practice they become an arbitrary choice

A modification of the traditional access rate calculation is therefore required

- Age banded information is extracted from the hospital data base rather than using a health authority boundary – this avoids the problem of specification of the catchment area
- Bed days rather than FCE are use for inpatient calculations – avoiding issues of FCE inflation
- Population growth is determined as a ratio of future to present for each age band – implied in the traditional access calculations but obscured in the method of application
- Overall growth is then expressed as a percent growth rate[8] – allowing greater flexibility in application
- This percent growth is then applied to a statistically averaged estimate of the current years demand – avoiding high misspecification arising from the use of a single year of activity data
- The hospital-based data is also used to determine the overall split of activity between locations. The ten year ONS population growth (by age band) for each district is then blended in proportion to overall activity to establish the overall growth by age band for the catchment area. The inherent uncertainty in this process is compensated for by rounding up/down the proportion for those areas with higher/lower population growth and using data for Milton Keynes (highest UK growth in the more mobile age bands) as the default for the group of 'other' purchasers. The resulting bias to slightly higher growth
- Allowance is made for the impact of increasing numbers of deaths (emergency admissions) and the impact of technology (elective admissions)
- A similar approach can be applied to forecast demographic-based outpatient growth – although in practice the actual growth is always higher than that due to demography alone (see comments in #1 above)

[8] While growth is expressed as a percentage in practice true percentage growth is rarely seen in healthcare trends. The most common form of growth is linear and hence the percentage value is best multiplied by the most recent activity to give a number value for growth. When added year by year this number value will give linear growth.

Resources required to deliver guaranteed waiting times

In practice it is the natural variation in healthcare demand rather than growth per se which determines the level of required resources[9]

- The variation in demand from one year to the next is orders of magnitude higher than underlying growth
- This variation can be expressed as a standard deviation around the expected average – in most healthcare settings this variation is in the range 3% to 25% of the annual total (larger variation is seen as the annual volume reduces)
- For a given guaranteed waiting time it is the level of variation associated with the demand in that time frame which determines the upper limit to the potentially available physical and workforce resources. Hence for a guaranteed inpatient wait of less than 3 months the variation is that associated with just 3 months worth of demand. This is typically twice the variation seen for the annual volume
- For inpatient beds this is reflected in the choice of average occupancy – the Erlang equation can be used to match the size of the bed pool with the appropriate occupancy – the one size fits all approach of 82% average employed by the NHS misses opportunities for economies of scale[10]
- For outpatient and day case situations Poisson statistics can be used to calculate the required average utilisation rate (the equivalent to occupancy)
- In general the required occupancy or utilisation rate implied by a 13 week or 3 month time frame is far less than current NHS practice
- The statistical basis for variation in demand further suggests that over-capacity in physical resources is required while current levels of permanent staffing will have to fall to be replaced by a much higher proportion of on-call staffing.
- By implication the use of industrial-style process control charts will become far more common as a tool for triggering the mobilisation of on-call staff.

Contracting to achieve waiting time guarantees

Calculations based on last years out-turn plus something extra to reduce the waiting list is not an acceptable basis for a contract.

- Any activity figure needs to be converted into demand by adding the change in the waiting list which occurred in that year
- They are based on a single year value thereby ignoring the statistical variation inherent in healthcare demand
- Even elective out-turn is part of a statistical distribution influenced by random events such as staff illness, unavoidable bed closures, case-mix, etc. The probability of last years value reoccurring is therefore low. A longer term average is a far better basis.

Delivery of waiting time guarantees imply some form of short-term over-contracting in order to avoid breaching the target.

[9] Jones R (2009) What next for 18 weeks? British Journal of Healthcare Management 15(8), 404-405.
Jones R (2009) How to maintain 18 weeks. British Journal of Healthcare Management 15(9), 456-457.

[10] Jones R (2009) Emergency admissions and hospital beds. British Journal of Healthcare Management 15(6), 289-296.

- It is far less likely to breach a target when you are well below the target rather than close up to it[11]
- Characterisation of the standard deviation associated with each type of demand is therefore extremely important[12]
- During the period over which the waiting time is reduced it is suggested that the historic elective activity is used as the basis to determine the standard deviation associated with 'elective' demand. One standard deviation should then be added to the Trust total with allocation pro-rata of this Trust total amount down to specialty level.
- Variation in emergency admissions should be handled via contract tolerances (expressed as standard deviations rather than % variances)
- Total elective contracted activity will therefore equal:
- Forecast average demand
  - + one standard deviation (the risk margin)
  - + reduction in the number on the inpatient waiting list
  - + additional inpatient work arising from extra outpatient volumes

Over-contracting is avoided in practice by the simultaneous use of waiting list control charts.

- The natural variation in the waiting list is used to calculate an upper control limit for the waiting list downward trajectory. This is then used for end of month waiting list review
- Only when the upper control limit is breached is extra activity sanctioned – the activity delivered is thereby kept to the minimum required.

Require Further Assistance?

Dr Rod Jones has been assisting hospitals to forecast demand and the required number of beds for over 15 years. Often the best that can be achieved (given financial and other pressures) is a compromise solution; however, it is far better to know the real situation than to pretend that 'efficiency' is going to save the day. Efficiency is far harder to achieve in the midst of the chaos which results from having too few beds!

For further details:

Website: www.hcaf.biz
Email: hcaf_rod@yahoo.co.uk
Telephone: +44 (0)1276 21061

[11] Calculations show that the target of a 13 week maximum outpatient wait implies an average wait of 9 weeks to avoid breaching the upper limit of 13 weeks.
[12] As a general rule surgical inpatient demand varies between one and two-times the variation expected by simple Poisson randomness. Medical inpatient demand varies between two- and three-times that expected of simple Poisson randomness. The higher variation is a reflection of the greater impact of the environment (mainly the weather – shifts in temperature, pressure & humidity) on the triggering of an acute episode of a pre-existing medical condition or weakness.