Presented at 'Re-thinking Bed Management – Opportunities & Challenges'. Harrogate Management Centre Conference, London, Thursday 27th February, 2003

Bed Management

(Tools to aid the correct allocation of Beds)

Dr Rod Jones (ACMA)
Healthcare Analysis & Forecasting
Camberley, Surrey, UK
hcaf_rod@yahoo.co.uk

Executive Summary

1. The mathematics of queuing developed by the Danish mathematician A.K. Erlang in the early 1900's has been used widely throughout industry & commerce for many years. It gives amazing insight into the correct sizing of hospital bed pools[1].

2. The methods used to forecast bed requirements within the health industry are shown to be seriously flawed and have led, in part, to the current bed crisis.

3. Population demographics coupled with an analysis of bed-days (by age band) rather that consultant episodes (FCE) are shown to give far more reliable estimates of future growth. FCE-based forecasts tend to underestimate true growth – another major flaw in the current method of forecasting.

4. A bed pool is defined as a group of beds that meet the needs of a similar group of patients, e.g. Maternity, Surgery & Urology, T &O, Head & Neck, etc.

5. The size of the constituent bed pools rather than the total size of the hospital determine apparent hospital efficiency.

6. These economies of scale within a bed pool are most dramatic up to 100 beds.

7. Major benefits in terms of throughput per bed can be obtained through consolidating two smaller bed pools into one larger one.

8. The correct sizing of bed pools can also eliminate one source of the delays leading to long A&E trolley waits, cancelled operations and the other hidden queues such as GP stacking systems.

9. Sufficient beds are required to cope with the winter peak(s) in medical bed demand.

---

[1] Jones, R.P., 2001, Health Service Journal, 111(5752), 28-32

Introduction

Here we aim to explain in simple terms the overall method used to forecast the required number of beds in an acute hospital.

In many ways the process is similar to everyday events where there is the potential to form a queue. Hence by arriving at a petrol station with the expectation that we will not queue for any great length of time we understand that this is because someone has worked out how many pumps to install to keep the queue to an acceptable length.

Surprisingly it is only recently that this approach has been applied to the provision of hospital beds. The mathematics behind queuing (using an equation called the Erlang equation) is extremely well known and used extensively and with great confidence in all manner of industries. So why has it not been used in healthcare? The rather blunt answer is partly ignorance and partly that there already was a method that most people thought gave the correct answer!

The currently accepted method used in healthcare is to take the forecast number of admissions and to multiply this by the forecast average length of stay in hospital and to then make some allowance for the fact that you cannot operate with the beds 100% full all the time – this is called the bed occupancy. Using the analogy of a petrol station our common sense would tell us that if all the pumps were occupied 100% of the time then there would most likely be a massive queue and we would go elsewhere.

Hence if in a single year we expect 100 admissions with an average stay of 3.65 days then we will need 365 bed-days over 365 days giving a requirement for 1 bed. The recent national bed inquiry in the UK recommended an average occupancy of 82% hence we then divide by 0.82 to give the supposed available beds, which in this case is 1.2 beds (which would be equivalent to 1 bed available 7 days of the week and a further bed only made available for 1 week in five or 1 day in five, etc).

Any layman will quickly recognize the serious flaw in this approach, namely, if 2 or 3 of our 100 expected admissions happen to arrive on the same day we may not have enough beds to cope and 1 or more may have to wait for a bed to become free. The resulting wait could be many days particularly if the length of stay (LOS) of the patient already in the bed is longer than the average of 3.65 days. The Erlang equation has the advantage of incorporating this randomness in the arrivals and in the LOS into the calculation of required beds. The method currently used within the NHS is completely unable to account for this vital component of real bed demand and as a consequence will tend to underestimate true bed demand.

In summary, the currently used method for calculating beds is very similar to the Erlang equation in that both use an arrival rate (admissions per year or per day) and both also use a length of time at the point of service (i.e. the average length of stay). The only difference is that the Erlang equation gives us powerful insight into both the length of the queue and the time spent in the queue arising from our chosen or actual average occupancy.

Hence using the Erlang equation we can understand why the UK government's promise to reduce the waiting list and the waiting time has not been met. Most hospitals (especially in the S.E. of England) are operating at an average occupancy that is so high that the only outcome possible is a long (and growing) queue. In this case politicians have made promises which are mathematically difficult to deliver!

There is a further more subtle factor which the Erlang equation is equipped to deal with, namely, the difference between expected average and actual numbers due to randomness.

For example, if we expect one car per minute (on average) to arrive at our petrol station how many can arrive in a one minute period. Common sense tells us that the answer must be more than one. Encapsulated within the Erlang equation is what is called Poisson statistics. This branch of statistics tells us about arrival events such as cars to a petrol station, GP referrals, emergency admissions, arrivals at A&E, etc.

Hence Poisson statistics tells us that for an average arrival rate of 1 per minute we can get up to an amazing 9 arrivals in a particular minute and in around 37% of minutes we can expect no arrivals. It is this perfectly natural erratic behavior around the average that leads to the formation of queues. It is this queuing that is one cause of the unacceptable A&E trolley waits currently experienced within many hospitals.

In actual fact the Erlang equation is the only method which can be accurately used to determine the size of the bed pool required for Intensive Care, High Dependency and Special Care Baby units as well as Maternity and Paediatric units where patients cannot queue for entry. The currently accepted NHS method only gives answers which are massively too small. This may be acceptable for containing costs (by withholding treatment) but is hopeless for delivering a modern-day service.

Turning now to a review of hospital services we can use the Erlang equation to predict the following very important outcomes:

- The larger the bed pool the shorter the queue, i.e. one larger hospital delivers a far shorter total waiting list than two smaller hospitals.
- The larger the bed pool the higher the bed occupancy and hence the higher the productivity, i.e. one larger hospital delivers more operations per bed and hence costs less to run than two smaller hospitals.
- The larger the bed pool the shorter the stay in the queue, i.e. one larger hospital delivers shorter A&E trolley waits, fewer medical patients in surgical beds, etc than does two smaller hospitals.
- By consolidating beds from two sites into one we can avoid the expense of having to provide & maintain additional beds at each individual site.
- Separating elective and emergency surgery onto two sites is likewise not as beneficial as may at first appear since it loses the benefits of size derived from the larger combined bed pool. The perceived benefits gained on the elective site are in fact an illusion due to a poor understanding of the nature of demand and the resulting efficiency of the emergency site will be even worse than when the two were combined!

- Within a single hospital the best performance will always be obtained if smaller bed pool boundaries are merged into a larger overall pool – with appropriate bed management to support such a move.

So if we combine the bed pool from two smaller hospitals (or specialties) into one we will obtain a host of positive benefits, namely, shorter waiting lists, shorter waiting times and far fewer poor quality outcomes such as long A&E trolley waits and patients placed into a bed in the wrong specialty. We also achieve slightly lower costs per patient due to the higher occupancy that can be sustained by a larger bed pool.

Having established the usefulness of the Erlang equation in correctly sizing a hospital bed pool we now need to look at how we estimate the future admissions and the future length of stay (LOS). Once again there are accepted methods used within the NHS and sadly these also tend to underestimate the true future demand.

At this point you may be asking why all NHS methods seem to underestimate the true future demand. The answer is quite simple; they were developed over 20 years ago when the main thrust was cost containment. It was also a period of rapid developments in medical science leading to reductions in length of stay; hence, any tendency to underestimate was partly compensated for and partly allowed to overspill into a growing national waiting list. Both politicians and public now find this an unacceptable outcome

At first glance the Erlang equation uses both arrival rate (admissions) and average length of stay (LOS) to calculate required resources and queue length. However a deeper study shows that the total beddays can be used directly to calculate the required bed pool size.

This opens the way for extremely rapid calculation of bed pool size and consequent 'what-if' calculations for alternative bed arrangements. This method is the proprietary knowledge of Healthcare Analysis & Forecasting.

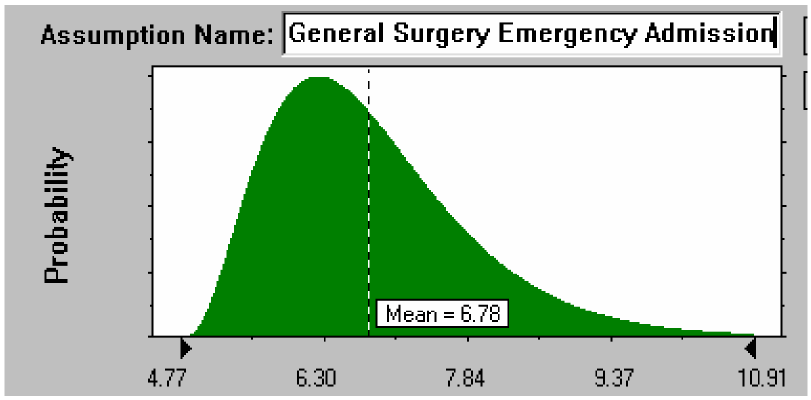
Forecasting Demand Based on Bed-days

There is much to recommend a departure from the accepted route of using admissions to forecast future demand via changes in population demography. The reasons are as follows

1. A period of care with one consultant is counted as a Finished Consultant Episode (FCE). FCE inflation due to the increasing specialisation within medicine and the transfer from on-take emergency teams is well documented in the NHS and results in higher apparent growth. However this same inflation acts to simultaneously reduce the apparent LOS since a fixed number of bed-days is divided by a higher number of FCE to give an apparent lower LOS. Some of the recent trend to supposed lower LOS is driven simply by FCE inflation.

2. LOS has already been demonstrated to be highly age related. In reality we are simply stating that bed-days increases with age. Hence an aging population will bring increasing pressure on apparent LOS.

3. My own research shows that the apparent casemix for emergency admissions is not the result of simple randomness. This area is poorly understood although the UK MET Office Health Forecasting Unit is beginning to reveal the basis for many of these changes.



As a result both numbers of admissions and individual LOS do not follow simple trends and are not as tightly linked to clinical practice as many have assumed. This is partly the reason for the complex statistical distributions behind average emergency LOS. The figure above shows one such distribution of the monthly average LOS for General Surgery emergency admission in one Hertfordshire hospital. Interestingly the Gamma distribution (used to describe the LOS data) is often used to describe metrological processes – do we really understand the basis for emergency admission so well that we can apply a simplistic average LOS to future bed planning? The trends in bed-days are more robust and far easier to follow since they are the end result of a number of complex processes.

4. The calculation of LOS is skewed by the inclusion of zero LOS admissions. Such admissions do occur and lead to a false reduction in calculated average LOS.

5. Finally, the forecast annual growth using bed-days tend to give far more realistic numbers.

Having chosen bed-days as the basis for forecasting future bed requirements how do we reconcile this with our use of the Erlang equation with its requirement for an arrival rate and a length of stay?

Fortunately this linkage is easily made via the choice of an appropriate level of turn-away. A direct linkage between bed-days, occupancy and beds is thereby established.

In conclusion, an approach based upon bed-days has been used to forecast both the future growth for a hospital or region and to determine the current baseline bed demand for the individual hospital sites.

Effect of Bed Pool Size of % Occupancy and Queuing

What is a bed pool?

While this may seem a trivial question we do need to be clear about the definition of a bed pool in order to understand the implications of the use of the Erlang Equation.

A bed pool is a common set of beds receiving similar patients.

Hence examples of a bed pool would include:

General Surgery + Urology; Head & Neck Specialties (ENT + Ophthalmology + Oral + Plastic Surgery); Obstetrics (mothers & babies only); Gynaecology (bed pool is defined by gender); Paediatrics (bed pool is defined by age); Medical specialties, etc.

The entire hospital does NOT class as a bed pool. Hence the large economies of scale forecast by the Erlang equation result principally from the combination of two General Surgery pools into one, or two Head & Neck pools into one, etc.

What does turn-away mean?

Turn-away is an overall description of the deleterious consequences of insufficient resources (e.g. beds). Arriving patients are turned-away from their desired destination (i.e. the correct bed type in the correct specialty bed pool) either to another hospital, to a trolley wait, by cancellation of their operation if the admission is from the waiting list, or to some other hidden queue such as a GP stacking system. A 5% turn-away implies that 5 in 100 patients suffer some deleterious consequence.[2]

The figure of 3% turn-away used in this work is one derived from experience as a balance between affordability and extent of queuing. Remember that this nominal 3% is the average over 24 hours (higher during the day and lower at night).

What is average occupancy?

Average occupancy is simply the number of patients occupying a bed divided by the number of available beds. Hence in the Erlang equation the figure for average occupancy assumes a 7-day, 24-hour average, i.e. a true average.

In the NHS the so-called 'average' occupancy is measured and reported at midnight, i.e. the point of lowest occupancy in the 24-hour cycle[3]. Hence while many hospitals may be quoting an average acute bed occupancy of 94% their true 24-hour average will be closer to 97% since most hospitals run at or near 100% occupancy during the daytime hours.

[2] Judging from the universally high midnight occupancy reported throughout the NHS the current incidence of real turn-away is likely to be far higher than 5%. Attempts to measure this are usually confounded by 'generous' interpretations of the various definitions.

[3] In reality the average occupancy quoted within the NHS excludes periods of time when a bed may be closed due to staff shortage or planned closure over a weekend. There is also a distinct weekly cycle to occupancy with highest occupancy at the middle of the week and lowest occupancy over weekends.

From this simple observation we can therefore understand why so many patients experience a real delay before a correct bed can be located for them to occupy and why true A&E trolley waits are so long. Imagine turning up at a petrol station where the petrol pumps are on average 97% occupied – you would expect a very long queue.

Since most hospitals do not record admissions and discharges (and hence bed occupancy) in real time it is difficult to get a true average. The simplest way to convert is to average midnight occupancy and daytime occupancy.

If we get more efficient can we do better than Erlang?

Efficiency expresses itself in many forms. Hence at a basic level if there are more arrivals that resource available to meet those arrivals then the queue will be long and growing. What the Erlang equations tells us is how many resources have to be (made) available to meet the demand and avoid an unacceptable queue. Efficiency in this instance would imply the use of the Erlang equation as a prerequisite to knowing how to allocate resources in the first place!

Within the NHS the cause of the queue may not be immediately apparent. For instance, long trolley waits in A&E may be due to poor 'efficiency' or resource allocation within A&E rather than bed availability per se. Hence the principles behind the Erlang equation would apply to the A&E waiting time as a calculation separate from that around bed requirements.

Another aspect of efficiency is the throughput per bed. The Erlang equation tells us that this is fixed partly by the bed pool size and partly by the LOS. The only way to increase throughput per bed in a fixed bed pool is to increase occupancy (and hence turn-away) or to reduce the average LOS.

Given the pressure to meet waiting list targets most hospitals have increased the apparent efficiency by going down the increased occupancy route. Hence to go from 90% to 95% occupancy yields a 5.5% increase in throughput. Hospitals achieve this increased 'efficiency' by resorting to tactics such as progressing with an operation in the hope that a bed will be found later in the day. Obviously a bed does get found but only at the expense of a long trolley wait for an emergency admission, i.e. the consequences predicted by the Erlang equation are shifted to another area.

The other route to increased throughput is via a reduction in average LOS. The old style of bed planning used within the NHS appears to indicate that a 10% reduction in LOS translates directly into a 10% increase in throughput. The Erlang equation shows us that this is not true and hence starting with 100 beds and changing average LOS from 5 days to 4.5 days (10% reduction) only reduces bed demand by 9%.

The distribution of LOS and the non-relevance of average LOS is important because the shape of the distribution tells us that it is the longer LOS stays (i.e. adverse outcomes, hospital acquired infections, drug reactions, lower frequency complicated procedures) that actually drive the average.

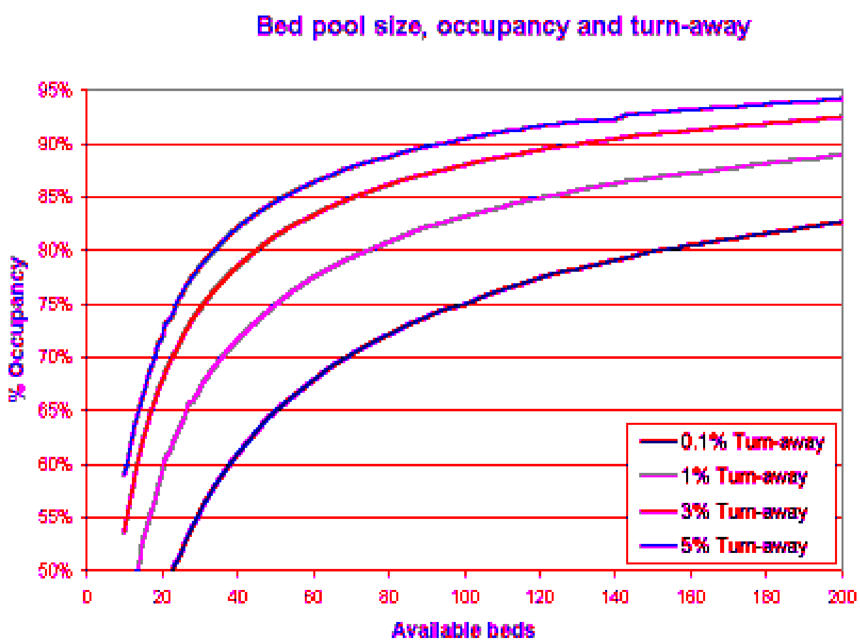
In this respect it is of interest to note that the process of placing medical patients in non-medical beds (due to bed shortages) itself leads to a lengthening of LOS since the patient is no longer in the right place to receive optimum care.

In the southeast of England the so-called bed blocking problem is also another powerful contributor to increased LOS in the Medical and Geriatric specialties. Obviously this later problem is a whole system issue where so-called hospital efficiency is driven by forces outside their control.

In conclusion, efficiency does not circumvent the applicability of the Erlang equation. A lack of efficiency will only make matters worse than that predicted by the equation.

How important is size?

The following graph shows the dramatic effect of the size of a bed pool on its associated occupancy and turn-away[4].



By turn-away we mean turned away from their destination (i.e. a bed in the appropriate specialty at the point of need) to go elsewhere or wait in a queue for admission. Even elective admissions face this possibility since at the point of admission they can face cancellation due to the lack of a bed.

From this chart it is immediately apparent why consolidating two medium sized hospitals into a single larger one has such a dramatic impact. To go from 20 to 40 beds increases throughput per bed by a massive 30% while maintaining the same level of turn-away. To go from 40 to 80 beds has less of an impact but still achieves a 20% increase in throughput per bed for the same level of turn-away.

[4] N.B. Do NOT look at the bed numbers and think in terms of a total hospital – this only works when you apply the definition of a bed pool, namely, a group of beds admitting a common set of patients.

A specialty with only 20 beds and 72% average occupancy has 5% of admissions queuing to get a bed. The same specialty with 40 beds and 72% average occupancy has only 1% of admissions joining a queue. A remarkable 5-times reduction in the number forced to wait in a queue. No amount of management intervention can change this behavior, it is dictated purely by bed pool size.

The Erlang equation also clearly demonstrates the exponential decline in quality of service (e.g. increased queuing and delays) as the occupancy increases. Hence for a pool of 100 beds we get the following:

Occupancy and turn-away in a pool of 100 beds

Average occupancy	Turn-away
75%	0.1%
83%	1%
88%	3%
91%	5%
97%	20%
99%	50%

As soon as we realize that the bulk of NHS hospitals currently operate at 94% midnight occupancy we see that if the bed pool size was 100 then at midnight over 5% of daily arriving admissions (emergency + overnight elective) will have queued to gain entry to a bed. It is clearly important to avoid too few beds.

We now need to ask ourselves what level of turn-away is acceptable to the NHS. Equivalent turn-away in the USA is around 1%[5] and this is probably beyond the NHS. At 5% turn-away the incidence of unacceptable outcomes, e.g. long trolley waits in A&E, high levels of cancelled operations, etc is probably too high hence we arrive at a rough upper limit of around 3% for a typical NHS application. This is a balance between access and affordability.

It is of interest to note that the recommended 82% average occupancy arising from the National Beds Inquiry delivers approximately 1% turn-away for bed pools of size 100.

Before progressing further it must be stressed that for elective work the occupancy figure is for the 5 working days during a week. Hence the raw number of bed-days needs to be adjusted to account for lower occupancy over the weekend and public holidays.

An example will now be given to show how this works in practice. If we had a combined General Surgery & Urology bed pool with a baseline 1,000 emergency bed-days and 2,000 elective bed-days we would apply adjusting factors giving, say, 1,010 emergency bed-days and 2,160 elective bed-days. Hence, a combined 3,170 bed-days or the equivalent to 8.68 beds at 100% occupancy. We now convert from bed-days to beds via the Erlang equation to see that we need 14 available beds to deliver a

[5] In 1995 the average US hospital had 6,100 admissions per annum and a bed occupancy of 65% (Gaynor et al 1995)

weekday midnight occupancy of 61% with around 3% of patients having to wait (but not for long) for the correct bed.

If we were to increase this requirement 10-times from 3,170 to 31,700 bed-days our requirement would be for 100 beds at 88% average weekday midnight occupancy but still maintaining a 3% turn-away. Over the whole week this translates into an 82% average occupancy – in this example the same as the stated aim for NHS bed occupancy. This implies that any surgical bed pool of size less than 100 beds will struggle to meet government targets for occupancy and thus avoid the deleterious consequences of turn-away. The implications to the current small to medium sized hospitals in the UK should be obvious.

A minimum limit for the surgical elective bed pool

Implicit in most peoples thinking is the assumption that the elective bed pool will continue to shrink, i.e. they have assumed that the relationship between beds, occupancy and turn-away is linear. The fact that this relationship is non-linear as the bed pool size reduces below 100 indicates that there is a point of diminishing return.

The next question we need to ask is – how big is the elective bed pool? The surprising answer is that it is already a small number. For example, a large NHS Trust (within the top 10% by size) will complete some 8,000 elective overnight operations per annum with a typical average elective LOS of 3.3 days. This is the equivalent of 100 elective surgical beds. An increase in the emergency bed demand of just 10 is therefore sufficient to remove 10% of the elective capacity.

While it may be mathematically possible to reduce the elective bed pool there are issues around exposure to risk. The uncertain nature of emergency demand has already been demonstrated for T&O and is even more so for the medical group of specialties. The fact is that all emergency demand even in the surgical specialties is slightly more uncertain than may be predicted by the Erlang equation. Ophthalmology emergency admissions respond to large changes in barometric pressure (unpublished research), urology emergency bed demand peaks in the summer, etc.

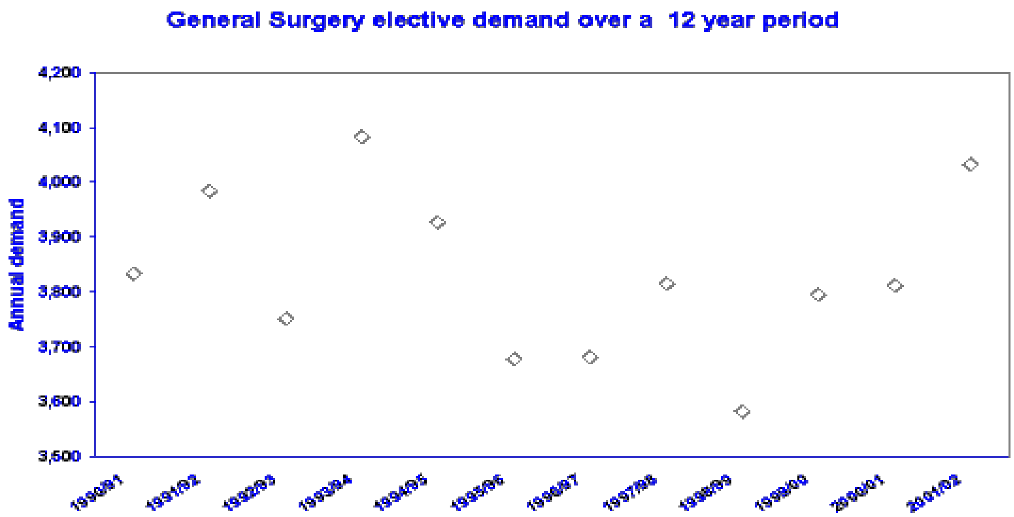
Some would use this as an argument to separate the two bed pools onto different sites. This however results in two smaller bed pools and makes the emergency bed pool even more susceptible to short term peaks.

Use of the Erlang equation for elective admissions

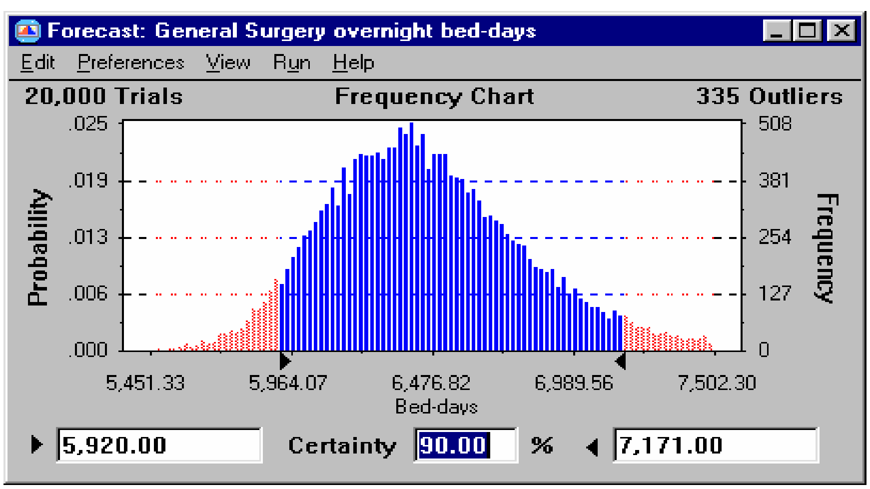
The next argument that could be raised is that elective admissions are planned and are therefore not random. Unfortunately this ignores the fact that elective demand is indeed driven by Poisson randomness and is also coupled with a variable average LOS. The following example comes from Berkshire and shows elective (overnight + daycase) demand. The scatter is characteristic of Poisson randomness plus some additional variation due to metrological & possibly viral causes.

Hence both conditions eminently satisfy the use of the Erlang equation and indeed justify the combination of emergency and elective streams into a common bed pool.

It is of interest to note that the UK national booked admissions program is based on flawed assumptions, namely, the designers have not understood the controlling fact of randomness. The mathematics clearly shows that any booked appointment/admission scheme can only be successful when there is an excess of resources[6,7].



Hence in our example the average General Surgery elective demand is 3,810 per annum although the demand in any year has only been close to this average in 4 years out of 12. If no allowance is made for randomness how do you cope with the year when demand hits 4,100? Indeed without a long run average how do you know the average with any degree of certainty? The planning of elective capacity is not as straight forward as has been assumed.



The above simulation shows the bed-days required in meeting General Surgery elective overnight demand given randomness in both demand (average of 1,690 FCE)

[6] Jones,R.P. 2001 Health Service Journal, 111(5778), 20-23.
[7] See BMJ (2000) 324, 280-282 where a similar conclusion has been reached for cardiac surgery.

and annual average LOS (average 3.73 days). While the average is 6,300 bed-days the 100% range is between 5,450 and 7,500 bed-days, i.e. 22 to 29 beds! No wonder hospitals find it hard to deliver against others and sometimes their own (simplistic) expectations.

Are there benefits to separating emergency and elective streams?

The idea that emergency and elective care should be separated onto different sites has gained increased interest. There are many apparent benefits, namely,

- higher occupancy due to the planned nature of the elective work
- no medical patients in surgical beds
- no cancelled operations due to medical or surgical emergencies

There is however a hidden assumption – both sites are of adequate size to meet demand.

The preceding section has already demonstrated that elective demand is subject to the same randomness as emergency demand. Hence any supposed booked admissions approach has the implicit assumption that there are excess resources (beds, theatres, consultants) relative to the average demand in order to cope with the natural fluctuations in demand due to randomness. The mathematical consequences of this implicit assumption have been demonstrated elsewhere[8]. This represents a fatal flaw in any booked admissions type program.

The second point to be addressed lies around the use of booked admission units within an acute site. For example, Aintree Hospital (Liverpool UK) operates one such unit in conjunction with a patient hotel. The bed pool is ring-fenced. At this point we need to recognise that the patients streamed through such units are those with a predictable length of stay. Hence the less predictable are sent via the general hospital bed pool. In other words the problem is sent somewhere else. The statistical distribution in elective LOS demonstrated above can not be circumvented. If you concentrate one type (i.e. the more predictable elective procedures) in one area then of necessity you increase the unpredictability in another area.

Much of the argument to separate emergency from elective comes from the very fact that current hospital bed pools are too small in the first place. This generates a perceived problem, and hence a perceived solution, if only we could get rid of all those emergency admissions then we can get on with the work. This problem is exacerbated by our lack of understanding of the true need regarding Medical beds.

The simple fact, which is elegantly demonstrated by the Erlang equation, is that it is overall bed pool size that brings the benefits.

Hence, to go down this route would imply concentrating, say, all Orthopaedic emergency work on one site instead of say five sites. Using the Erlang equation we

[8] Jones, R.P. 2001. Health Service Journal, 111(5778), 20-23; Jones,R.P. 2001. Health Service Journal, 111 (5740), 28-31; Jones, R.P. 2000, Health Service Journal, 110 (5732), 28-31.

can rationally investigate the options but we must be clear that the overriding factor is to maximise the bed pool size.

Special Cases

There are a number of special cases where a 3% turn-away would not be appropriate. These will now be discussed.

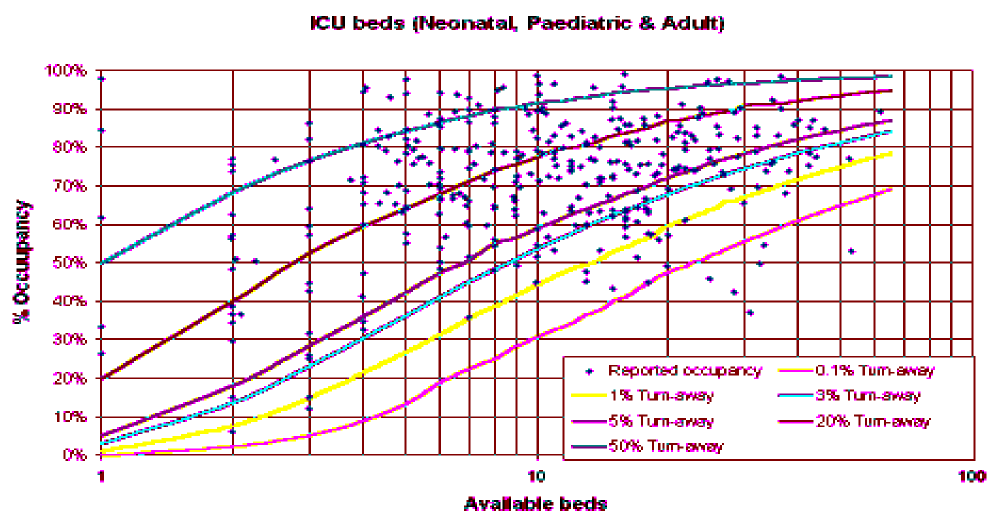
Intensive Care (Adult, Neonatal & High Dependency Units)

The validity of the Erlang equation to accurately predict the size of intensive care bed pools has been established[9]. In order to accommodate arriving patients a turn-away of around 0.1% needs to be established.

Due to the very small size of most of these bed pools they benefit enormously from an increase in size. Hence the combination of two 10 bed ICU's into a single 20 bed unit results in a 58% increase in the effective workload which can be carried before patients are turned-away. Obviously, grouping of ICU's into Regional pools will gain the benefit of the larger pool size; however, journey time can be critical to the survival of a patient.

Data for intensive care units in England (2000/01) is presented in the following figure.

As can be clearly seen all suffer from their relatively small size with unacceptable levels of turn-away, i.e. the majority run at higher than 5% turn-away with some even above 50% turn-away.



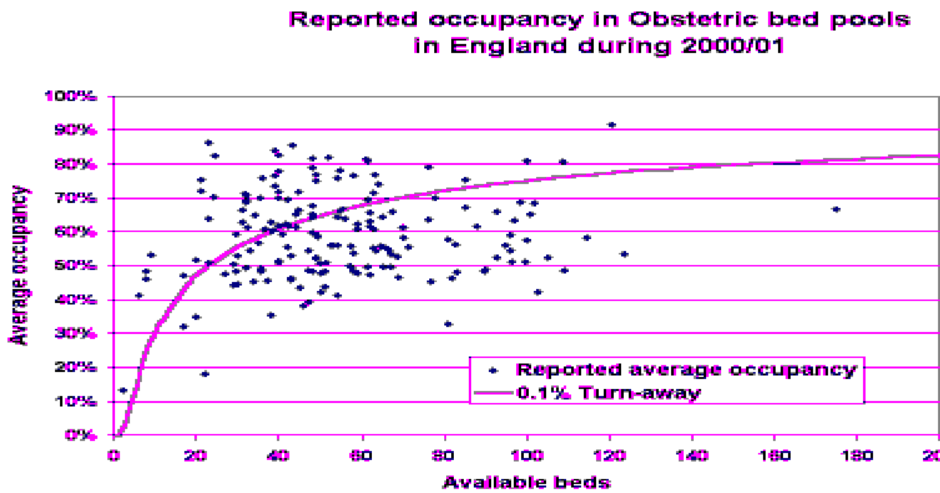
The consolidation of two sites into one (wherever possible) would therefore deliver enormous benefits to both intensive care but also to elective care requiring ICU recovery.

[9] Lamiell,J.M. 1995, Military Medicine, 160, 227-232.

The Obstetric Bed Pool

For obvious reasons it is not acceptable to turn pregnant women away at the point of delivery. For this reason the turn-away appropriate to the sizing of Maternity bed pools is around 0.1%. A figure of 0.1% implies that 1 admission in 1,000 would have to wait (for a very short period). This theoretical figure is acceptable because in reality a bed would be found from among those about to be vacated. This explains the typical low average occupancy figures for maternity units.

Since maternity units function 7 days per week and with modest seasonality there does not need to be any adjustment to go from raw bed-days to available beds. Actual data for Obstetric bed pools is given in the following chart.



Several points can be made from this data. Firstly, in general, occupancy increases with bed pool size as predicted by Erlang. Secondly, some Obstetric units are operating with more beds than that required to achieve a 0.1% turn-away. Lastly, some 50 out of 180 units appear to operate at occupancy higher than that consistent with 0.1% turn-away. There are three potential reasons:

- These units have too few beds but are able to operate in the range 0.1 to 1% turn-away (30 out of the 50 units)
- The higher reported occupancy is due to the recording of 'available beds'. The definition of an 'available' bed gives scope to exclude beds which are not staffed, hence, some maternity units open and close bays as required and do not count the full pool of available beds. This practice leads to a higher apparent occupancy than the 'true' figure.
- Most of the larger bed pools (i.e. >80 beds) are actually split over two or more sites, hence, these data points should be shifted to the left.

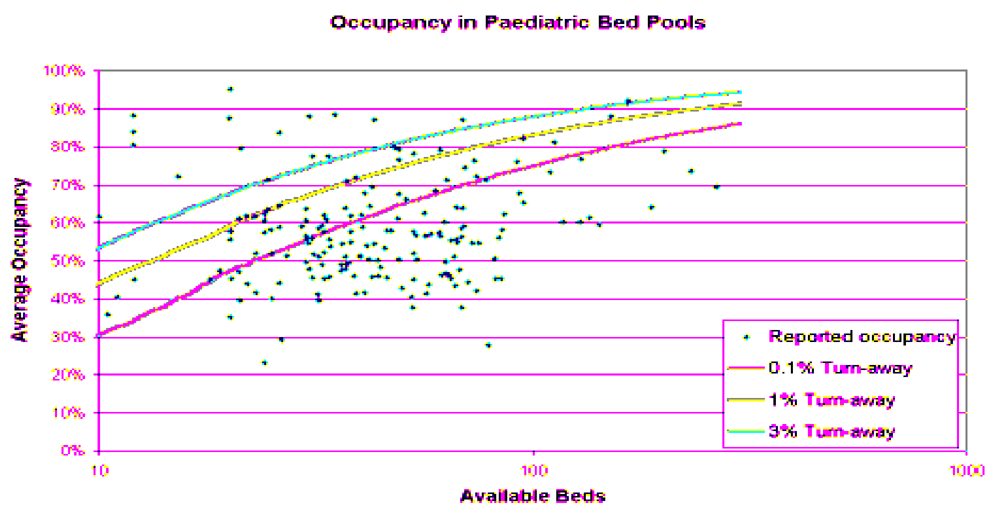
Hence the data (with allowance for differences in reporting of beds and aggregation of beds over multiple hospital sites) supports the use of the Erlang equation. The combination of two 50 bed maternity units into a single unit would decrease the beds required by over 10% and still maintain the same high standards of access.

Paediatric Units

Paediatric units are generally run on a no turn-away basis and hence the Erlang equation is generally applicable. The situation with Paediatric emergency care is however complicated by the highly seasonal nature of the incidence of the group of respiratory ailments that form the bulk of bed demand within this specialty[10]. In many ways Paediatrics is very similar to medicine in that the peak in bed demand in both specialties is related to respiratory conditions, although the exact respiratory conditions are different and peak in slightly different points in the year.

However the general principle of size still applies since the underlying randomness becomes smaller in proportional terms as size increases, e.g. two smaller units combined into one always require fewer beds.

Data for Paediatric units in England (2000/01) is given in the following chart.



As can be seen most do function below 0.1% turn-away. The winter peak in bed demand is the main reason that so many units appear to have such a low turn-away, i.e. they are resourced to deal with the winter peak rather than an annual average. The best method for estimating the required number of beds is to calculate the number based on the annual average bed-days and then to add an allowance for the winter peak. These extra beds could potentially be closed over the summer months.

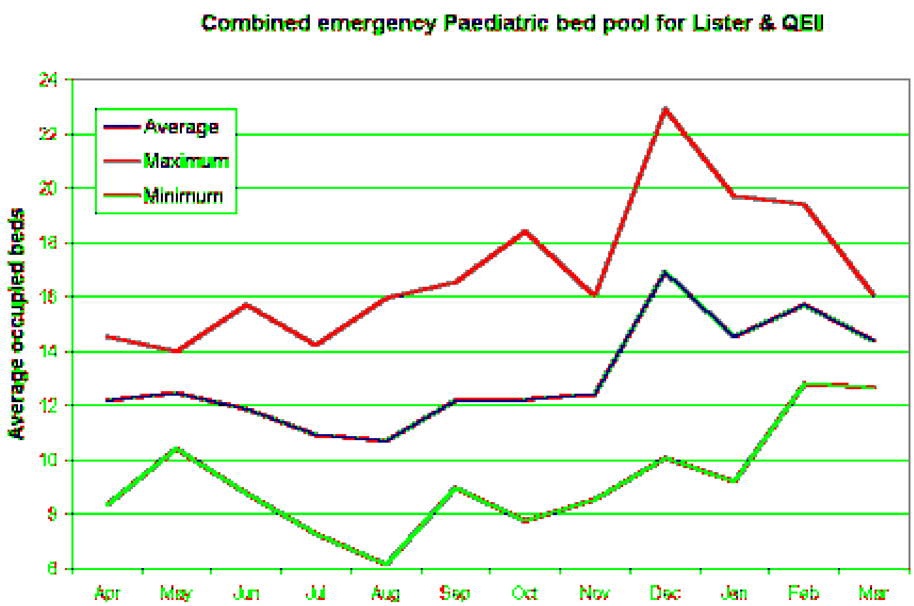
These factors will now be illustrated using data from the Lister & QEII hospitals (Hertfordshire UK). The following figure gives the results from the past 7 years. Over this time scale the underlying bed demand has remained relatively constant. The important point to note is that monthly bed demand is highly erratic hence for a December the highest bed demand was in 1995 while the lowest was in 2001 or for August the highest was in 2000 and the lowest in 1999, etc.

[10] Damiani,M and Dixon,J. 2001, Emergency pressures in London, 1997-2001, Kings Fund, London, UK.

The overall average is for 13 occupied beds, however, on average December requires 31% more than this while August requires 17% less. The chart also clearly shows the fact that a 'bad' summer can imply just as many occupied beds as an 'average' winter.

Using the overall average gives 25 beds as the suggested bed pool size. We need to ask if this would be sufficient to cope with the worst ever outcome of 711 bed-days demanded in December 1995 (next highest 651 bed-days in December 1996 and 611 in January 1996).

Given 25 beds this would represent 91.7% average occupancy. This would imply that at peak demand around 1 in 5 arrivals would experience some delay before finding a bed. Using Erlang-C we can show that the average delay (for the 1 in 5 experiencing a delay) would be 4 hours. This delay would be in the Paediatric assessment unit and hence patients would be receiving suitable care during this delay.



Given that this worst case scenario has only occurred once in 7 years the suggested bed pool of 25 is considered adequate for at least 6 out of 7 winters.

An inspection of the chart also shows that it would be possible to close 5 to 7 of the 25 beds over the months April to September and still be able to cope with the worst possible case over these months. This results in a highly efficient arrangement of resources that is not possible with the current operation on two sites.

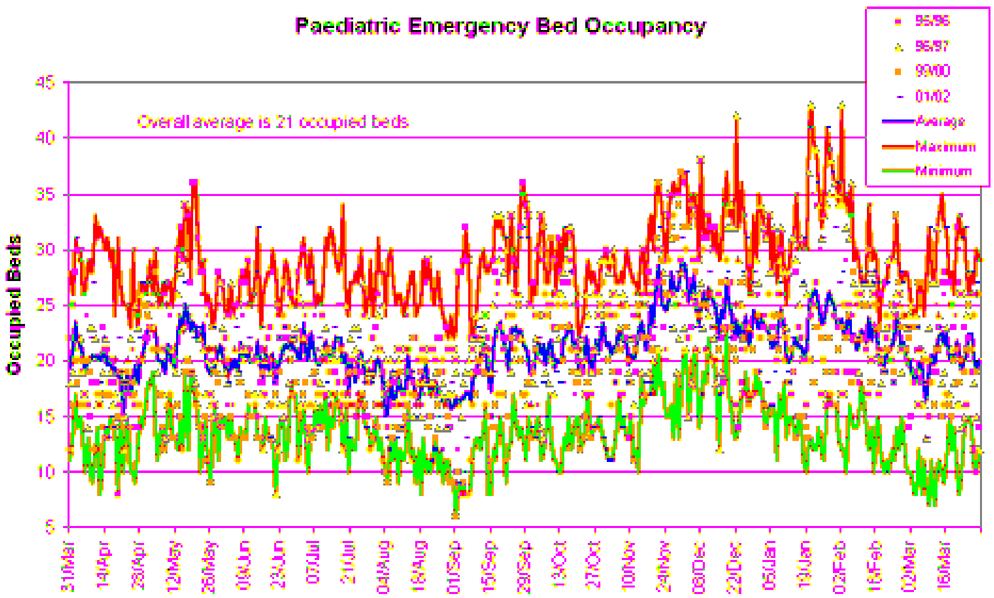
The above analysis is still open to the criticism that it is based on monthly averages rather than actual daily figures..

Daily occupancy data for a Paediatric unit in Berkshire is used to illustrate this technique. Data covers the eight-year period 1994/95 to 2001/02. High and low years are identical to those experienced in Hertfordshire suggesting a common cause[11]. In

[11] See BMJ (2002) 324, 763-766 for a link between viruses and allergens in adult asthma.

the chart data has been lined-up so that each year starts with the first Sunday of the year. This explains the saw tooth behaviour in the maximum, minimum and average where occupancy is highest on a Thursday (6.3% higher than the weekend).

Data from several years is displayed to illustrate the volatility in bed occupancy over short periods of time. Hence the summer of 1995 was typically high bed occupancy and the winter of 96/97 achieved highest daily occupancy although the summer of that year was average. The summer of 1999 was low bed occupancy with the winter of that year being average, etc.



The overall average is 21 occupied beds which translates into a recommended 36 beds. This was exceeded on only 16 days within the seven-year period and could have been accommodated by using the additional Paediatric beds usually set aside for elective admission.

In conclusion, the method based on 0.1% turn-away gives a realistic figure for bed requirements in medium to large Paediatric units. The small Paediatric units currently functioning in many locations would therefore benefit greatly from consolidation into a larger unit.

Healthcare Analysis & Forecasting uses a mixture of proprietary forecasting tools, simulation software and novel adaptations of the Erlang equations to solve resource allocation issues within healthcare, namely, how many beds does a specialty or hospital need, how many urgent, soon & routine outpatient appointment slots are required to guarantee waiting time targets, how much activity needs to be in a contract to guarantee achieving a target, what are the seasonal profiles behind activity and waiting lists, etc.

About the Author

Rod has a Ph.D. in Chemical Engineering from the University of Queensland where he was the first person to switch from science to engineering to do a doctorate. He is a Chartered Accountant in the Chartered Institute of Management Accountants.

His career includes 7 years in academia and 10 years in industry covering the biotechnology and food industries & as General Manager of an international laboratory proficiency testing organisation where he authored a handbook on statistical quality control for microbiological water testing. He has completed the Hewlett Packard intensive course in Total Quality Management for senior managers. In addition he has over 15 years experience in healthcare both within the NHS (Information, Contracting, Performance Management, Business Analysis & Decision Support) and as an independent consultant.

Healthcare Analysis & Forecasting was established in 1995 with clients including Trusts, Health Authorities, Prudential and Glaxo plc. A disease management study in gastrointestinal bleeding & ulcers won an international award within Glaxo plc.

His research has led to the development of many innovative methods for forecasting healthcare activity and demand.

1994 Statistical deficiencies in the methods used to formulate GPFH budgets
1996 Methods to forecast year-end activity
1996 Published a book on healthcare forecasting & waiting list management
1997 Socio-economic & demographic factors in DNA rates
1997 Role of viruses in determining the winter peak in medical bed demand
2000 Reasons for the September peak in outpatient waiting times
2000 Role of randomness in determining outpatient waiting times
2001 Relationship between bed occupancy & turn-away
2001 Resource implications of guaranteed waiting time targets
2001 Use of queuing theory in healthcare

In 1996 he completed a review of bed requirements for the Royal Berkshire & Battle Hospitals. His longer-term projections of bed requirements, which at the time were rejected by the Heath Authority, have been validated in practice and a second business case was required to get the correct number of beds!

In 1997 he ran a series of seminars for NHS managers teaching queuing theory as a method for determining bed pool size. At the time the methodology was ideologically unacceptable and it was not until the advent of the National Beds Inquiry that he was able to reintroduce the method back into the NHS with a publication in the Health Service Journal.

During 2001/02 he was involved in the DOH HOIP project investigating best practice in the use of hospital operational intelligence to match capacity with demand. He has a keen interest in healthcare finance and has developed methods to predict the risk associated with HRG based income and uses advanced methods to explore the sensitivity of business case assumptions.