Planning beds, bed occupancy and risk

NHS England and NHS Improvement

Planning beds, bed occupancy and risk

ECIST guidance

Version number: 1.0

First published: May 2019

Author: Dr Rob Findlay, Director, Gooroo Limited

Rob Findlay would like to thank: the Emergency Care Intensive Support Team (ECIST) at NHS England and NHS Improvement for their helpful suggestions during drafting and for assistance with publication; Piotr Fryzlewicz, Professor of Statistics at the London School of Economics, for reviewing and advising on the statistical methods used; and those acute hospitals in England and Scotland who helped to pilot the software behind these methods.

This information can be made available in electronic format, which is scalable for large print, upon request. Please contact the author on rob.findlay@gooroo.co.uk

Contents

Planning beds, bed occupancy and risk 1

Contents 2

1 Introduction 4

1.1 85 per cent bed occupancy? 4
1.2 The right bed occupancy is different everywhere 4
1.3 Smaller bed pools 5
1.4 Bed occupancy and risk 5

2 Calculating bed occupancy vs. risk 7

2.1 Possible approaches 7
2.2 Predictable and unpredictable variation 7
2.3 Week by week variation 8
2.4 Hour by hour variation 9
2.5 The right bed occupancy 11

3 Some strategies for tackling excessive bed occupancy 12

3.1 High occupancy has consequences for clinical safety 12
3.2 Bed pools merge as occupancy rises 12
3.3 Do morning and weekend discharges help? 13
3.4 The potential of reducing lengths of stay 14

4 Morning discharges and emergency department waits 16

4.1 More consequences of excessive bed occupancy 16
4.2 Durations with insufficient beds 16
4.3 Better to lower bed occupancy and risk 18

5 Different beds, different bed occupancies 19

5.1 59 different bed pools 19
5.2 Smaller bed pools need lower bed occupancy 19
5.3 Different patterns of variation in different bed pools 20
5.4 Every bed pool is different 21

6 Separating electives from non-electives 22

6.1 Elective bed occupancy 22
6.2 The qualitative case for hot and cold sites 22
6.3 The new quantitative case 23
6.4 A benign general hospital scenario 23
6.4.1 Moving orthopaedics and daycases to a cold site 24
6.4.2 Moving all elective care to a cold site 24
6.5 Scenario of a general hospital with high bed occupancy 25
6.5.1 Moving orthopaedics and daycases to a cold site 25
6.5.2 Extra capacity on the hot site 26
6.6 Conclusion 26

1 Introduction

1.1 85 per cent bed occupancy?

Conversations about bed capacity sometimes go along these lines:

> *"Right, we've got the activity and length of stay figures. What bed occupancy do you want to assume?"*
>
> "Let's use 85 per cent."
>
> *(crunches numbers) "Well in that case, you'll need this many beds."*
>
> "Oh. We don't have that many. Can we try 90 per cent?"
>
> *(more crunching) "There you go."*
>
> "Ah. How about 95?"

This is not a particularly scientific approach. Neither party knew what 85 (or indeed 95) per cent bed occupancy represented in terms of risk. So numbers were used quite casually, and in the end the capacity plan was simply tailored to fit the capacity available.

The 85 per cent figure, which the conversation started with, is widely cited by the Royal College of Surgeons,[1] the Royal College of Emergency Medicine,[2] the Royal College of Nursing[3] and the National Audit Office[4] as a widely-applicable safe limit to bed occupancy.

But all of those references trace back to the same source: a 1990s spreadsheet simulation of 200 hypothetical acute beds,[5] whose conclusion that "Risks are discernible when average bed occupancy rates exceed about 85%" has become a totemic number for all hospitals to aspire to.

1.2 The right bed occupancy is different everywhere

More recently NICE[6] looked for real-world evidence that risks such as hospital acquired infections, mortality, length of stay, emergency department waiting time, and readmission rates might be worse at high bed occupancy, but unfortunately the quality of evidence was graded "very low".

They concluded that "Healthcare providers should: … Plan capacity to minimise the risks associated with occupancy rates exceeding 90%" while noting that "optimum occupancy levels may vary with the size and type of the hospital (small versus large hospitals or tertiary versus general hospitals), case mix, the degree of predictability of bed availability from different wards and seasonal effects (winter period with more

1 www.rcseng.ac.uk/news-and-events/media-centre/press-releases/nhs-bed-occupancy-rates/
2 www.rcem.ac.uk/docs/Policy/Making%20the%20Case%20for%20the%20Four%20Hour%20Standard.pdf
3 www.rcn.org.uk/news-and-events/news/winter-crisis-exacerbated-by-under-investment-in-nursing-staff
4 www.nao.org.uk/wp-content/uploads/2018/02/Reducing-emergency-admissions.pdf
5 www.ncbi.nlm.nih.gov/pmc/articles/PMC28163/
6 www.nice.org.uk/guidance/ng94/evidence/39.bed-occupancy-pdf-172397464704

infections)”, which takes the crucial step of recognising that no single bed occupancy figure is going to be right everywhere.

There is also some evidence that high bed occupancy is associated with a higher risk of long waits in the emergency department, with NHS Improvement[7] putting the threshold at 92 per cent occupancy, and NHS Providers[8] settling on 88 per cent as the level when long waits tend to rise.

1.3 Smaller bed pools

So far we have discussed adult general and acute beds, which tend to form the largest bed pool in an acute hospital. The recommended bed occupancies are even lower for smaller, higher-risk bed pools where variation is harder to absorb.

The Intensive Care Society[9] have said “The Welsh government has indicated that units should run at an average occupancy of around 65-70%. Occupancies higher than this are known to lead to cancelled operations, non-clinical transfers and delayed admissions, each of which have their own impact on outcomes for patients”, and in a rare example of a circular reference the Welsh Government[10] have returned the favour by saying “The Intensive Care Society (ICS) states that critical care units should run at occupancy of 65-70%.”

In neonatal units the British Association of Perinatal Medicine[11] have previously suggested 70 per cent as a sensible cot occupancy, whereas the Department of Health[12] recommended in 2009 that “Planned capacity should not exceed an average occupancy of 80%, as the increase in mortality becomes statistically significantly worse above this level”.

1.4 Bed occupancy and risk

Whichever bed pool we look at, recent bed occupancies are often much higher than these recommendations.[13] That is clearly undesirable, but it begs several important questions:

- What level of risk do today’s excessive bed occupancies mean?
- What level of risk would be acceptable, considering the impact of running out of beds on clinical safety and the wider hospital?
- What can we do to avoid or mitigate these risks?

---

7 improvement.nhs.uk/documents/3201/NHS_review_of_winter_2017.18.pdf
8 nhsproviders.org/the-nhs-funding-settlement-recovering-lost-ground/performance
9 www.ics.ac.uk/AsiCommon/Controls/BSA/Downloader.aspx?iDocumentStorageKey=dcbb4ee2-8ad6-4ff4-9d32-451460f675a4&iFileTypeCode=PDF&iFileName=Guidelines%20for%20the%20Provision%20of%20Intensive%20Care%20Services
10 gov.wales/topics/health/nhswales/plans/delivery-plan/?lang=en
11 www.london.gov.uk/sites/default/files/gla_migrate_files_destination/archives/assembly-reports-health-counting-cots.pdf
12 webarchive.nationalarchives.gov.uk/20130123200735/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/@ps/@sta/@perf/documents/digitalasset/dh_108435.pdf
13 www.nuffieldtrust.org.uk/resource/hospital-bed-occupancy

If a hospital is running at 95 per cent bed occupancy in its general and acute beds (a level exceeded by 23 trusts - including some of the largest - in quarter 3 of 2018-19),[14] then what are the implications? Should they expect to run out of beds 10 per cent of the time, or 30 per cent? It matters: you might struggle through the former by cancelling elective surgery and letting queues build up in the emergency department from time to time, but the latter is a different level of pressure altogether.

And what about the consequences? If running out of general and acute beds means cancelling routine elective surgery, then clearly that is not desirable, and the delay may increase clinical risk to the patient but it would not usually be immediately life-threatening. A hospital might accept this happening as often as several hours per week – a risk of several per cent.

But running out of maternity, paediatric, or critical care beds is altogether more serious for patients. The acceptable level of risk would be much lower – even a few hours per year, or around 0.1 per cent, might have serious safety implications.

So risk is a much better basis for informed discussion than bed occupancy. Once we can identify the consequences of running out of beds, and agree the acceptable risks of that happening, it should then be a technical exercise to calculate the corresponding levels of bed occupancy – as the next chapter will show.

14 www.england.nhs.uk/statistics/statistical-work-areas/bed-availability-and-occupancy/bed-data-overnight/

2 Calculating bed occupancy vs. risk

2.1 Possible approaches

In the previous chapter we saw that running out of beds has consequences for patients and the hospital. The acceptable risk of that happening, rather than bed occupancy, should be the basis for informed discussions of bed capacity. So when the acceptable level of risk has been agreed, we then need a way of converting it into bed occupancy for capacity planning purposes.

There are several ways of doing this.

We could run a computer simulation, which was how the well-known 85 per cent bed occupancy figure[15] was obtained, taking care to make sure that the distribution of admissions and discharges is realistic.

Or we could use the queueing theory techniques developed by Erlang.[16]

Or we could analyse the variations in non-elective admissions and discharges over different timescales, using real data from the specific beds concerned. This is the method we are going to use here.

2.2 Predictable and unpredictable variation

Before going into the numbers, we need to consider what bed occupancy is for. It isn't simply about absorbing variation – it's about absorbing that variation which we cannot predict and manage by other means. So we know that respiratory admissions rise in winter, and trauma in summer, but we can manage that by adjusting the number of staffed beds and reprofiling elective admissions. Bed occupancy only needs to be low enough to absorb the other variation that we cannot predict and manage.

We therefore need a way of forecasting non-elective admissions, and then we need to test it against some real data to see how well it does. The variation it successfully predicts – with enough notice that we can take action to deal with it – is something we can manage by other means. Whereas any error in those predictions is something we cannot predict and manage, and that is what bed occupancy needs to absorb.

Therefore if we can improve our forecasting, so that less of the variation is unpredictable, then we can operate at higher bed occupancy with fewer beds and lower costs – a nice example of the benefits of good analysis.

How shall we do our forecast? Forecasting is a huge subject, but for current purposes a straightforward approach will suffice. Operational managers often forecast non-elective demand by calculating the average over the last six weeks

15 www.ncbi.nlm.nih.gov/pmc/articles/PMC28163/

16 www.hcaf.biz/Hospital%20Beds/Microsoft%20Word%20-%20Bed%20planning%20HMC.pdf

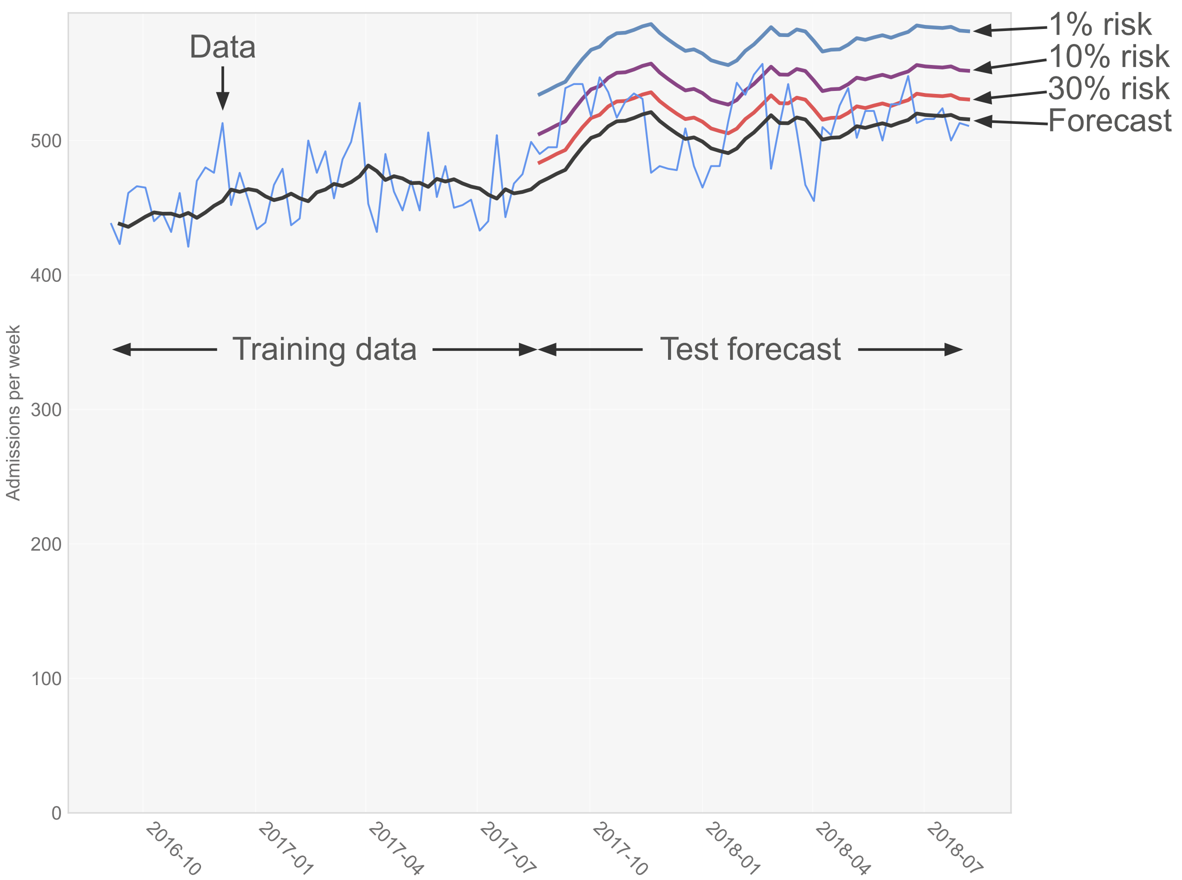
(considering each weekday separately), and this method performs reasonably well.[17] A slightly more advanced technique is simple exponential smoothing[18] which can be tuned to fit the variability of the data and is fairly easy to do in a spreadsheet; this is the method we will use here because (as we will see later) we need to be sensitive to the different patterns of variation in different bed pools.

2.3 Week by week variation

Let's apply exponential smoothing to some real data.

The jagged blue line in the chart below shows the actual number of non-elective admissions to adult general and acute beds, in a particular hospital, every week for two years. The first year of data was used to tune the forecasting model, and then the model was allowed to run on through the second year to see how well it performed. Each forecast was just for the following week, and the forecasts are shown by the dark grey line running through the middle.

The other lines show how well the forecast performed in the second year. For instance, there is a 1 per cent risk of admissions breaching the top (blue) line in any given week, and a 10 per cent risk of breaching the second (purple) line down.



17 www.southampton.ac.uk/~sks/research/papers/admissionmetsim.pdf
18 en.wikipedia.org/wiki/Exponential_smoothing

Now we are ready to calculate bed occupancy.

If we make enough beds available to cope with the top line, then we can expect to run out of beds only 1 per cent of the time. The forecast line running through the middle shows the average number of beds that will be occupied.

Now recall that, when planning bed capacity, a good definition of bed occupancy is the average proportion of available beds that are occupied over time. Therefore the bed occupancy that is consistent with a 1 per cent risk of running out of beds, is the ratio between the forecast line and the top line.

We can use the same method to calculate bed occupancy for any level of risk in this bed pool – which gives us the results we wanted.

2.4 Hour by hour variation

So far our calculation of bed occupancy has taken account of the week-by-week variations in demand. But that isn't the whole picture, because non-elective bed usage also varies by the time of day, and day of the week, which we will refer to as 'intra-week variation'. We need to take account of this variation too.

The chart below shows the intra-week variation in occupied non-elective beds (i.e. admissions minus discharges) in the same adult general and acute bed pool.

The heavy black line in the middle shows what happens on average, and it is a strong and predictable cycle.

Every day, the beds fill up and then empty out. This is because admissions are fairly steady throughout the day (except in the small hours of the morning), but most discharges from this hospital happen in the late afternoon with a peak around 6pm. So beds are fullest with non-elective patients in the middle of the day.

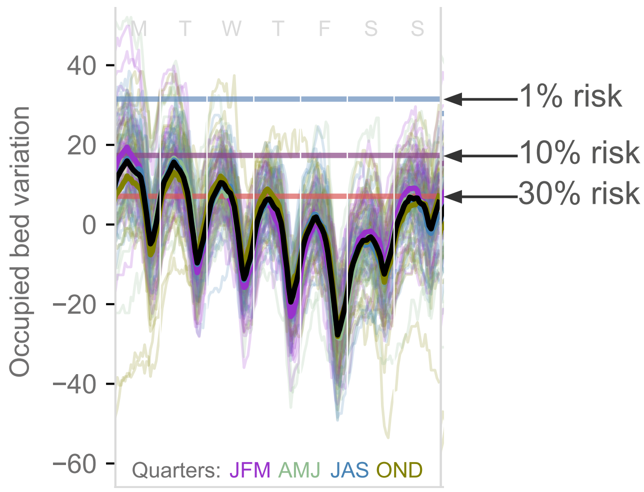
We can also see that beds tend to empty out during the working week and fill up at the weekend. This is because, although the rate of admissions is lower at weekends, the rate of discharges is even lower, so the net effect is for beds to fill up with non-elective patients on Saturdays and Sundays. The emptying out during the working week is the converse of this effect.

If the heavy black line shows the average pattern, what about the variation?

Each individual week (over a two year period) is shown by a different coloured line in the chart below, and each of those lines has been zero-based because we have already taken week-by-week variation into account.

Each week's line has been coloured according to the calendar quarter in which it began. There are also heavy lines showing the average for that quarter; the fact that they are hard to see shows that intra-week variation does not vary much from season to season.

There are a few extreme lines going from top left to bottom right (the weeks before each Christmas), and from bottom left to top right (the weeks after each Christmas). But even if we ignore those, we can see that one week is very different from another – even though we have already zero-based them all.



So we have a strong, regular cycle, as well as random variation – in other words, we have both predictable and unpredictable variation. Which of them should we take into account when calculating bed occupancy?

Again, it depends on whether we can manage the predictable variation by other means. On the assumption that we are not literally going to open and close beds hour by hour and day by day, we need to maintain a low enough bed occupancy to absorb the predictable as well as the unpredictable variation. (And if that assumption were disputed, it turns out to make surprisingly little difference anyway, as we will see in the next chapter.)

The risk lines have therefore been drawn straight across the chart, and we are now ready to look at the relationship between risk and bed occupancy for this intra-week variation.

If we make enough extra beds available to satisfy the top horizontal line, then there will be a 1 per cent risk of running out of beds based on the intra-week variation. The same method generalises to different levels of risk, and next two horizontal lines show the extra beds associated with 10 per cent and 30 per cent risk.

Finally, we need to combine these intra-week risks with the week-by-week variations examined in the previous section (on the basis that the variations are not correlated, which they aren't).

2.5 The right bed occupancy

The end result is the overall relationship between the risk of running out of beds, and bed occupancy, for this particular pool of adult general and acute beds. It is summarised in the table below.

Risk of running out of beds	Bed occupancy
30%	97%
20%	95%
10%	92%
5%	90%
1%	87%
0.5%	85%
0.1%	83%

If the acceptable risk of running out of these beds is a few per cent, then it is apparent that the popular figure of 85 per cent bed occupancy would be unnecessarily low and achieving it would be an unnecessary use of resources. On the other hand the more recently cited figure of 92 per cent would be too high and present an unacceptable risk of running out of beds.

In the NHS today, however, real bed occupancies are often much higher than either figure. The table shows that if bed occupancy were 95 per cent in this example then there would be insufficient non-elective beds about 20 per cent of the time, with the familiar consequences of frequently cancelled elective surgery, widespread outliers in other parts of the hospital, and increased clinical risks.

Such high risks are clearly unacceptable. But it may not be possible to open more beds. So in the next chapter we will look at what else might be done.

3 Some strategies for tackling excessive bed occupancy

3.1 High occupancy has consequences for clinical safety

Rising bed occupancy has consequences for clinical safety: the boundaries between different bed pools break down until eventually there is little distinction between surgical and medical beds, the infection control cordon around orthopaedics is breached, female surgery spills out into gynaecology, older teenagers are bedded out in paediatrics, and patients end up sleeping overnight in the emergency department and the day surgery unit.

When patients end up in the wrong beds, they have longer lengths of stay (which makes the bed occupancy problem worse), which in turn is associated with poorer care and higher cost[19] – a vicious cycle. There is also evidence of increased mortality and higher emergency readmission rates.[20]

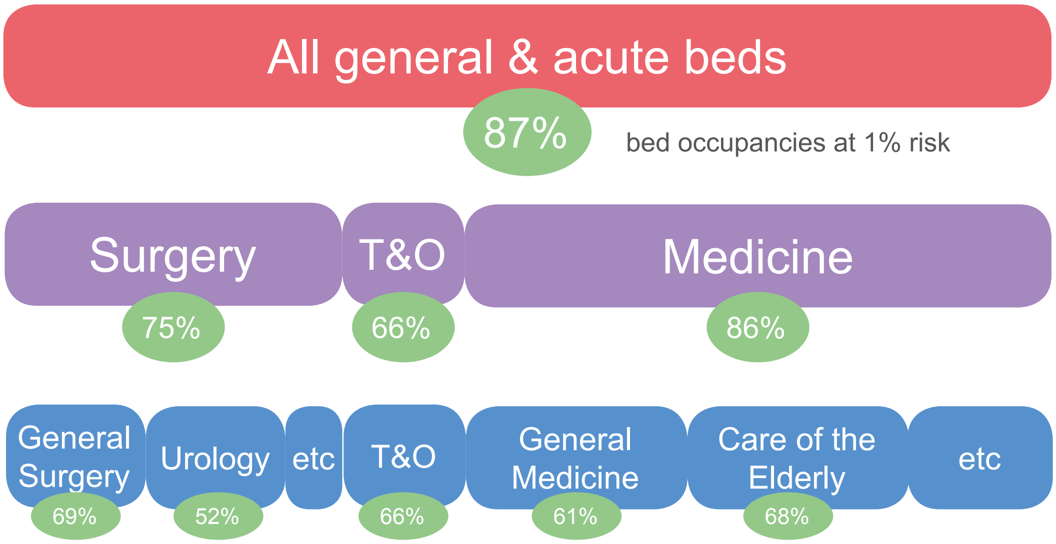
3.2 Bed pools merge as occupancy rises

Acute hospitals take a pragmatic approach to deciding which beds should be used for each specialty. The financial planning process allocates budgets in some detail, and this translates into staff and therefore staffed beds. But nobody realistically expects every specialty to keep within its own detailed bed allocation all the time. That would require a very low bed occupancy which is simply unaffordable – see for instance the example occupancies in the bottom row of the figure below.

So specialty beds are usually combined into larger bed pools – such as medicine, surgery, orthopaedics, paediatrics, and maternity – and some adult bed pools are shown in the middle layer below. By working flexibly within each bed pool, services can run at higher bed occupancy for the same overall risk of running out of beds. Smaller services such as maternity and critical care may form close relationships with neighbouring units and lower their risk that way.

19 bmjopen.bmj.com/content/7/5/e015676

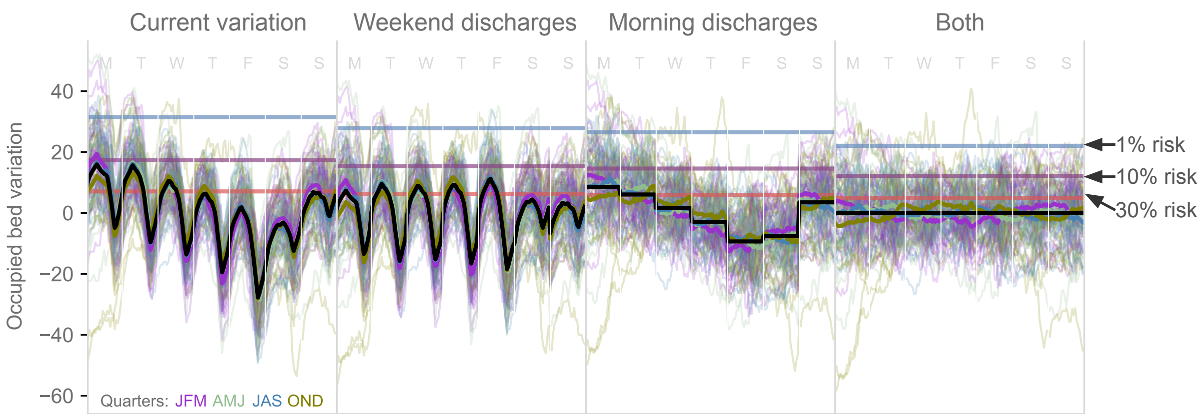
20 www.acutemedicine.org.uk/wp-content/uploads/2014/11/Plenary-5-1030-Wrong-Place-Anytime-Why-Boarding-is-Bad-for-Patients-Hospitals-and-Healthcare-Systems.pdf



But if bed occupancy is too high even for these arrangements, then the remaining boundaries break down and the hospital essentially becomes one giant bed pool with all the risks outlined above. This is shown by the top layer in the diagram above. What can be done to avoid this, if opening extra beds is not an option?

3.3 Do morning and weekend discharges help?

We saw in the previous chapter how bed occupancy can be calculated for a chosen risk of running out of beds, and we can use the same method to work out whether morning and weekend discharges are likely to help.



The four charts above show, from left to right:

1) the actual variation in bed usage during the week – we saw this chart in the previous chapter.

Then the next three charts show the potential to reduce that variation (and hence raise the bed occupancy consistent with any given level of risk) by:

2) improving weekend discharges, so that the average discharge rate matches the average admission rate every day of the week;

3) improving morning discharges, so that the average discharge rate matches the average admission rate every hour of the day (but leaving the days different);

4) doing both, so that the average discharge rate matches the average admission rate every hour of every day of the week.

We can tell whether these interventions make much difference by looking at the horizontal lines marking (from the top down) the 1 per cent, 10 per cent, and 30 per cent risks of running out of beds. They narrow as discharges improve, but not dramatically.

And when that intra-week risk is combined with the week-by-week variation to calculate the overall bed occupancy, the benefits of better discharge patterns turn out to be negligible, as shown in the table below.

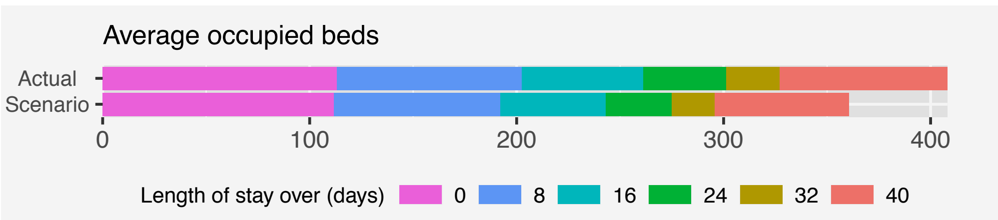
	Current variation	**Weekend discharges**	**Morning discharges**	**Both**
Risk	**Bed occ**	**Bed occ**	**Bed occ**	**Bed occ**
30%	96.5%	96.7%	96.7%	96.7%
20%	94.7%	94.7%	94.9%	95.0%
10%	92.0%	92.2%	92.4%	92.5%
5%	90.0%	90.4%	90.5%	90.7%
1%	86.5%	86.9%	87.0%	87.4%
0.5%	85.4%	85.7%	85.7%	86.2%
0.1%	83.0%	83.2%	83.5%	83.9%

But surely morning discharges do help with 4 hour waits in the emergency department? Indeed they do, sometimes. We will see why, and when, in the next chapter.

3.4 The potential of reducing lengths of stay

If better discharge patterns produce disappointing results, what about reducing lengths of stay (LoS)[21] and avoiding admissions[22]?

The chart below shows a scenario in which any patient with a LoS over 21 days has their LoS somehow reduced by 20 per cent, or 10 per cent if it is over 7 days. This turns out to reduce the number of occupied beds by as much as 11.5 per cent.



21 improvement.nhs.uk/resources/reviewing-stranded-patients-hospital-what-are-patients-waiting/
22 www.longtermplan.nhs.uk/online-version/chapter-1-a-new-service-model-for-the-21st-century/2-the-nhs-will-reduce-pressure-on-emergency-hospital-services/

It is admittedly an arbitrary scenario. But the point is that we aren't scratching around for tenths of a percentage point any more – this is a huge difference. And it arises because the majority of these beds are occupied by those patients who have the longest lengths of stay.

Idle time between patients[23] is significant too, and comes from delays in communicating that the bed is empty and in cleaning and portering. The potential of both these approaches together is very large indeed, and in many hospitals may be enough to reduce bed occupancy and risk to acceptable levels at moderate cost.

In this chapter we have only looked at one example, but similar conclusions have been reached with all the general and acute bed pools that we have so far looked at: reducing the longest lengths of stay, as opposed to increasing morning and weekend discharges, is where efforts should be focused.

---

23 blogs.bmj.com/bmj/2019/01/21/richard-smith-bed-management-in-hospitals-horrible-and-badly-in-need-of-reform/

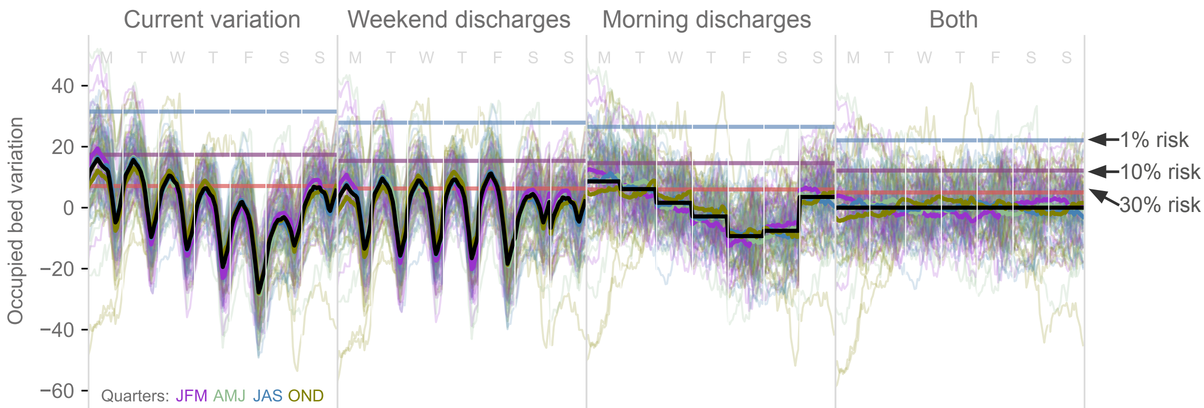
4 Morning discharges and emergency department waits

4.1 More consequences of excessive bed occupancy

In the previous chapter we saw that an increase in morning discharges makes only a negligible difference to the risk of running out of beds, and hence to non-elective bed occupancy. But this surprising conclusion is at odds with the experience of operational managers, who find that morning discharges do help to avoid 4 hour waits in the emergency department.

How can we reconcile these apparently contradictory positions? The answer lies in how long these periods with insufficient beds last, and whether bed occupancy is already very high.

We have already seen the figure below, in the previous chapter. Look at the 30 per cent risk line (the third (red) horizontal line down), which represents a bed occupancy of nearly 97 per cent in this example.



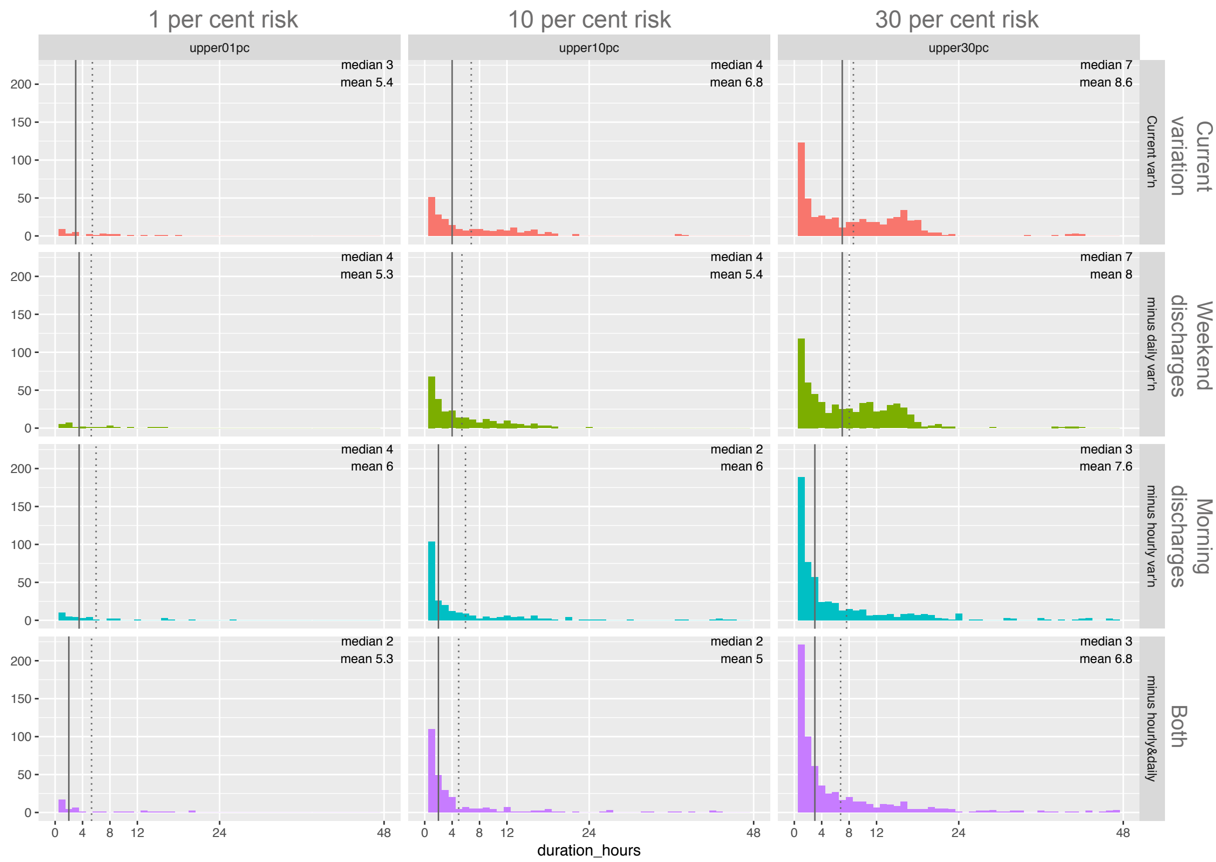
In the left hand ("current variation") chart the red line slices straight through Monday and Tuesday. This means that if bed occupancy is so high that there is a 30 per cent risk of not having enough beds on average, then most of Monday and Tuesday will probably be spent with insufficient beds because they are the worst days.

That means very long delays in accessing beds, twice a week, most weeks – a frequent and significant risk to clinical safety.

4.2 Durations with insufficient beds

The array of charts below shows how long all the periods with insufficient beds last, under the various scenarios and risk levels. Let's start by looking at the top right hand chart, which represents this very high bed occupancy scenario at the current patterns of variation.

This chart is a histogram. The horizontal axis shows the duration (in hours) of each period with insufficient beds. The vertical axis shows how many times those periods occurred in the two years of data (taking just the intra-week variations into account).



The spike towards the left of the chart shows that the most common period with insufficient beds was just one hour long. But the smaller, flatter peak further to the right shows that there were also a lot of periods lasting more than 12 hours – like all the bed-starved Mondays and Tuesdays.

Looking down the right hand column of charts, we can see how the distribution of these durations changes as we test the various scenarios: from the current situation (in red, at the top), to increasing weekend discharges (green), to increasing morning discharges (blue), and doing both (purple, at the bottom).

A big change happens as soon as morning discharges are introduced – the peak in over-12-hour periods with insufficient beds disappears, and they are replaced with a larger number of short periods.

This change is also reflected in the statistics marked on each chart above. The solid vertical black line is the median duration of periods with insufficient beds, so half of the periods are shorter than the median and half longer. Morning discharges have the potential here to reduce the median from 7 hours to just 3 hours.

This matters because 3 hours is less than the 4 hour emergency department waiting time standard that is currently in force. If most periods with insufficient beds are less than 4 hours long, then these surges can be buffered in the emergency department without triggering large-scale breaches of the standard or the consequent clinical risks to patients. However if many of those periods last longer than 12 hours, as happens at present, then that is not possible and those very long waits come with significant clinical risk.

So far we have only looked at the right hand column of charts, which represents a 30 per cent risk of running out of beds (at nearly 97 per cent bed occupancy in this example).

4.3 Better to lower bed occupancy and risk

What happens at lower risks? The middle column shows a 10 per cent risk (92 per cent occupancy in this example), and the left hand column shows a 1 per cent risk (87 per cent occupancy).

At these lower bed occupancies the median duration of periods with insufficient beds is not above 4 hours even at current patterns of variation. Morning discharges make less and less difference as bed occupancy and risk are reduced.

So morning discharges are a useful ‘fire-fighting’ tactic to help reduce 4 hour (and longer) waits in the emergency department if bed occupancy is already very high. But they are not a strategic route towards a sustainably lower risk of running out of beds, nor do they make much difference if those risks are already lower.

5 Different beds, different bed occupancies

5.1 59 different bed pools

So far we have looked at a particular pool of adult general and acute beds as an example, but every bed pool is different. The following table applies the same methods to 59 different bed pools, picking out the figures for a 1 per cent risk of running out of beds.

1% risk	all Gen & Acute	Medicine	Surgery	Elderly	Trauma	Cardiology	Critical care	Gynaecology	Maternity	Paediatrics	Neonatology
Hospital 1	89%	86%	80%		68%			45%		36%	
Hospital 2	88%	86%	79%	70%	75%	68%	58%	52%	71%	65%	56%
Hospital 3	87%	82%	72%	68%	67%	44%	37%	49%	70%	67%	65%
Hospital 4	87%	82%	75%		65%	52%		48%		67%	
Hospital 5	83%	83%	73%		69%	42%		60%	75%	72%	
Hospital 6	83%	81%	73%		61%	39%		35%	68%	71%	
Hospital 7	83%	83%	52%			29%					
Hospital 8	54%		54%								
Hospital 9	38%	38%									

Even within the larger units in the general and acute bed category, these hospitals range from 83 per cent to 89 per cent bed occupancies to have the same risk of running out of beds. Why the variation? Mainly because different hospitals have different sizes of bed pool, and in general a larger bed pool can absorb variation more easily.

5.2 Smaller bed pools need lower bed occupancy

To see why, imagine that you are about to toss a number of coins. What are the odds that 70 per cent or more will be 'heads'?

The answer depends on the number of coins. If there are 10 coins, then the odds are better than 1 in 6. But if there are 100 coins, the odds are worse than 1 in 25,000.

Similarly in larger bed pools, a surge in demand in one area is more likely to be averaged out by a lull in another.

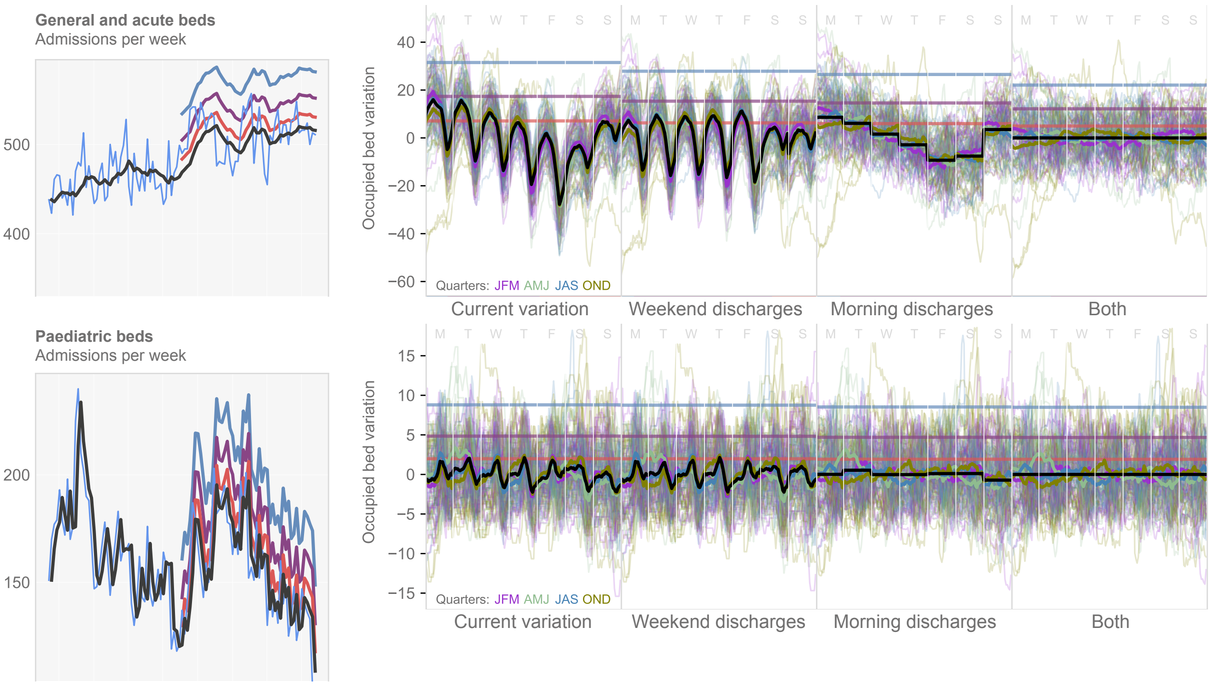
The differences in bed occupancy are even larger when we look at paediatrics and neonatology – even assuming a constant 1 per cent risk in all cases. But it is a bit artificial to hold the risk level steady. In real life the differences would be bigger because general and acute beds could probably accept a higher risk of several per cent, whereas in the other bed pools the consequences for clinical safety would be worse and the risk would need to be much lower.

In Hospital 3, for instance, a 0.1 per cent risk of running out of beds would imply a bed occupancy of just 64 per cent in maternity, 60 per cent in paediatrics, and a cot occupancy of 58 per cent in neonatology. Those occupancies are far lower than the recommendations of official and professional bodies that we reviewed in the first chapter, let alone the higher occupancies that are prevalent today.

5.3 Different patterns of variation in different bed pools

Looking into the detail behind these numbers, the pattern of variation is also different between these bed pools, and again this has been a theme across the hospitals we have looked at so far.

The charts below show the week by week and intra-week variations in some adult general and acute beds (top charts), and paediatric beds (bottom charts).



The first difference is in the week by week variation (left hand charts). The forecast, which is shown by the central grey line, uses an exponential smoothing model which is ‘tuned’ using the first half of the data. This tuning process works out whether the forecast needs to respond quickly to the most recent changes, or whether it is better to ride out the fluctuations and produce a smoother forecast.

In the top left chart showing general and acute beds, the tuning process has concluded that riding out the fluctuations with a smooth forecast will produce the best results. In practical terms this means that if you want to forecast next week, then you should average many recent weeks for a reasonable estimate.

The bottom left chart for paediatrics is completely different. Here, the tuning process has concluded that only the recent past matters, and a few weeks ago is already ancient history. So if you want to know what next week is going to look like, then it’s best to look at this week. The result is a much more variable forecast, pointing to the need for much greater flexibility in the way that paediatric bed capacity is used (or alternatively to even lower bed occupancy to maintain acceptable levels of risk).

The second difference is in the intra-week variation. The paediatric charts do not show the characteristic regular daily and weekly cycles of the general and acute beds, and instead random variation dominates. As a result, there is almost no difference in risk when morning and weekend average discharges are adjusted to match average admission rates.

This is just one comparison, but similar patterns were seen across most of the hospitals looked at here.

5.4 Every bed pool is different

The conclusion then is that every bed pool is different, and different clinical areas differ in their patterns of variation. Different types of bed pool need to operate at different levels of risk, because running out of beds has different consequences for clinical safety. Different sizes of bed pool need different bed occupancies. Not only that, but different bed pools show different patterns of variation, and the past is a different guide to the future.

In the first chapter we looked at the variety of bed occupancy levels being recommended by different official and professional bodies, and now we can see that none of them can be applied universally (though there may be themes). If they had instead recommended an acceptable risk of running out of beds, they would have been closer to the mark. Because, as NICE[24] concluded, “optimum occupancy levels may vary”.

This concludes the analyses of non-elective bed occupancy. In the final chapter we will look at the consequences for elective beds, and how a risk-based approach can help with the long-running debate about separating ‘hot’ and ‘cold’ facilities.

24 www.nice.org.uk/guidance/ng94/evidence/39.bed-occupancy-pdf-172397464704

6 Separating electives from non-electives

6.1 Elective bed occupancy

Elective activity is scheduled in advance, so it is predictable, and therefore bed occupancy can be very high – perhaps 100 per cent minus the Did Not Attend (DNA) rate and bed idle time, or even higher if there is deliberate overbooking in anticipation of DNAs.

This is a much simpler formula than for non-elective care, because we don't need to account for variations in demand – routine patients can be booked around the cancer and other urgent elective patients to smooth out the peaks and troughs.[25] (The main exception is when a service is purely cancer, and then the week-by-week variation does need to be taken into account.)

The difficulty starts when elective and non-elective services share the same bed pool, because surges in non-elective demand spill over into elective beds, and elective patients end up being cancelled. Careful planning is needed to adjust the number of non-elective beds, or to profile elective inpatient admissions around the peaks in non-elective demand, or both.

Alternatively this problem can be avoided by separating non-elective and elective care onto 'hot' and 'cold' sites. But which approach is better?

6.2 The qualitative case for hot and cold sites

The most common NHS model is to mix hot and cold on a general hospital site. But separate cold sites have also been around for years, including Independent Sector Treatment Centres in England[26] and the Golden Jubilee[27] hospital in Scotland, and there is a new wave of Scottish NHS treatment centres[28] on the way. Further south, Plymouth[29] and Cornwall[30] are moving orthopaedics out to cold sites, and cold sites are also being encouraged more widely in the NHS Long Term Plan.[31]

The clear benefits are that cold sites can run more smoothly and efficiently away from any disruptive surges in non-elective demand. It is also a relatively straightforward way to create much-needed new capacity.

But there are repercussions for services left behind on the 'hot' site. In a mixed elective and non-elective hospital, you have the option of cancelling some routine surgery if you run out of non-elective beds – but if elective services have moved to a cold site then that safety valve goes with them.

---

25 blog.gooroo.co.uk/2018/06/demand-capacity-and-the-law-of-averages/
26 www.kingsfund.org.uk/sites/files/kf/Briefing-Independent-sector-treatment-centres-ISTC-Chris-Naylor-Sarah-Gregory-Kings-Fund-October-2009.pdf
27 www.nhsgoldenjubilee.co.uk/our-services/
28 news.gov.scot/news/major-investment-in-elective-treatment-centres
29 www.hsj.co.uk/integration/why-plymouth-hospital-is-insourcing-an-elective-unit/7023684.article
30 www.hsj.co.uk/service-design/troubled-trust-moves-major-service-in-bid-to-improve-flow/7024109.article
31 www.longtermplan.nhs.uk/online-version/chapter-3-further-progress-on-care-quality-and-outcomes/better-care-for-major-health-conditions/short-waits-for-planned-care/

6.3 The new quantitative case

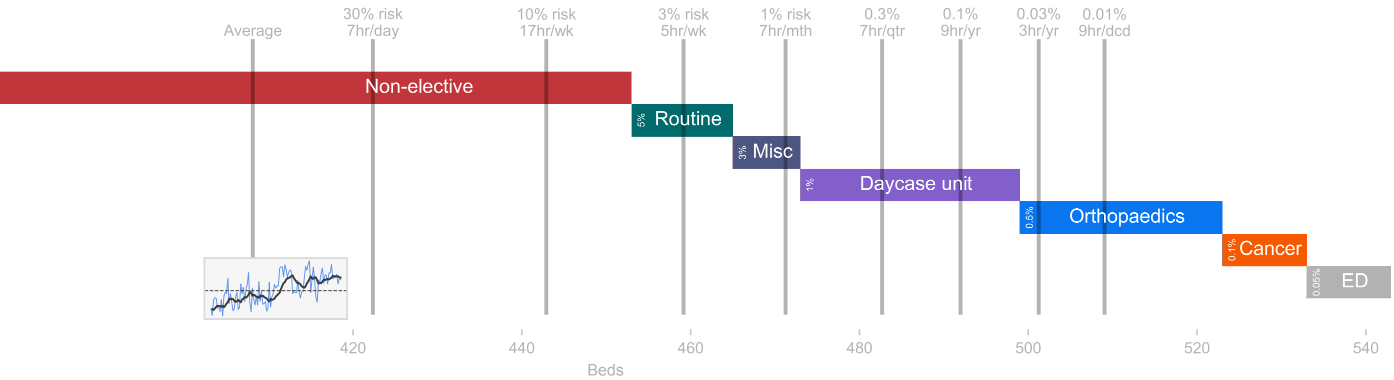
In the past the discussion may have rested on those qualitative arguments. But now we can calculate bed occupancy from the acceptable risk of running out of beds, so it is possible to put numbers around this.

Let's start with a few (illustrative) assumptions that the acceptable risk of running out of beds in a general hospital – with all the consequences for clinical safety – is:

- 5 per cent (about 8 hours per week) if the consequence is rearranging routine elective patients in non-ring-fenced services like general surgery;
- 3 per cent (5 hours per week) if that means using the little miscellaneous areas around the hospital where 3 or 4 patients could go at a push;
- 1 per cent (2 hours per week) if non-elective patients use the day surgery unit;
- 0.5 per cent (4 hours per month) if they use the ring-fenced orthopaedic ward;
- 0.1 per cent (9 hours per year) if the consequence is rearranging cancer patients;
- 0.05 per cent (4 hours per year) if the consequence is keeping patients overnight in the emergency department (ED), or in corridors, or if ambulances are diverted.

6.4 A benign general hospital scenario

We'll start with the scenario of a general hospital with reasonably benign non-elective bed occupancy. It might look something like the diagram below.



In a typical NHS acute general hospital, most of the beds are non-elective, so the red bar actually extends a long way off the left of the diagram – as the bed scale along the bottom indicates.

Then the steps to the right of the red bar show various kinds of elective bed – in practice some of the categories are mixed together on the wards, but here they are shown separately to reflect the different acceptable risks of cancelling patients.

In this discussion we will want to keep two concepts separated in our minds. One is the intended purpose of the available beds (shown by the bars), and the other is how they are actually used at different times (shown by the vertical risk lines). For the purposes of this chapter we will assume that the number of available beds has been forecast and managed so that the risk lines represent only unpredictable fluctuations (the thumbnail chart of week-by-week variation at the lower left corner of the diagram serves to remind us that such predictable variation is being managed in other ways).

As the demand for non-elective care fluctuates unpredictably, the number of beds occupied by non-elective patients extends in and out from the left. Most of the time, non-elective demand can be absorbed within the available non-elective beds (i.e. within the red bar).

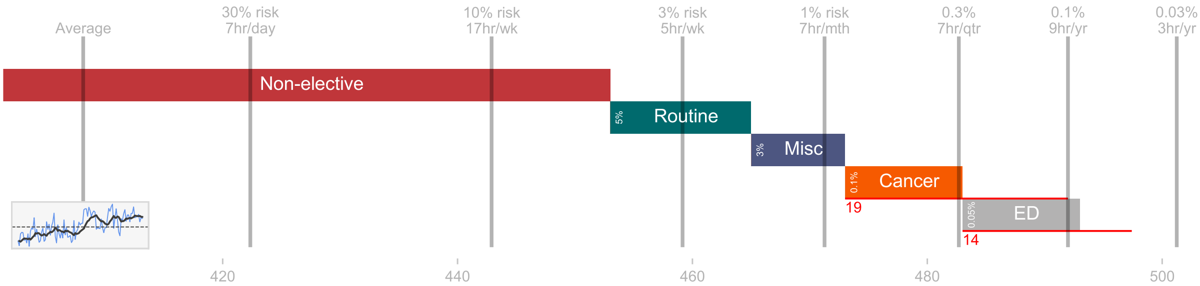
But about 5 per cent of the time, non-elective patients overflow into routine elective beds. The vertical grey lines indicate how likely it is that different numbers of beds will be occupied by non-elective patients.

In this benign scenario, the risk of non-elective patients overflowing into the various elective beds is roughly in line with the acceptable risks that we outlined above.

6.4.1 Moving orthopaedics and daycases to a cold site

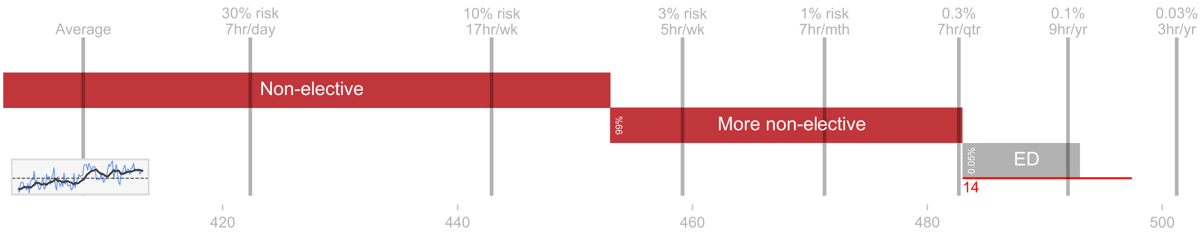
What happens if we move orthopaedics and day surgery to a cold site elsewhere? If we closed those beds back on the hot site, then other beds become the safety valve. The risk of rearranging cancer surgery, or stacking non-elective patients up in ED or corridors, would be much higher than we intended – as the diagram below shows.

So the level of clinical risk has become unacceptable at the hot site, and there will be pressure to re-open some of those beds. The number of extra beds needed to achieve the acceptable risk levels is shown in red below the cancer and ED bars.



6.4.2 Moving all elective care to a cold site

What happens if we then go further, and move all elective care to a cold site?

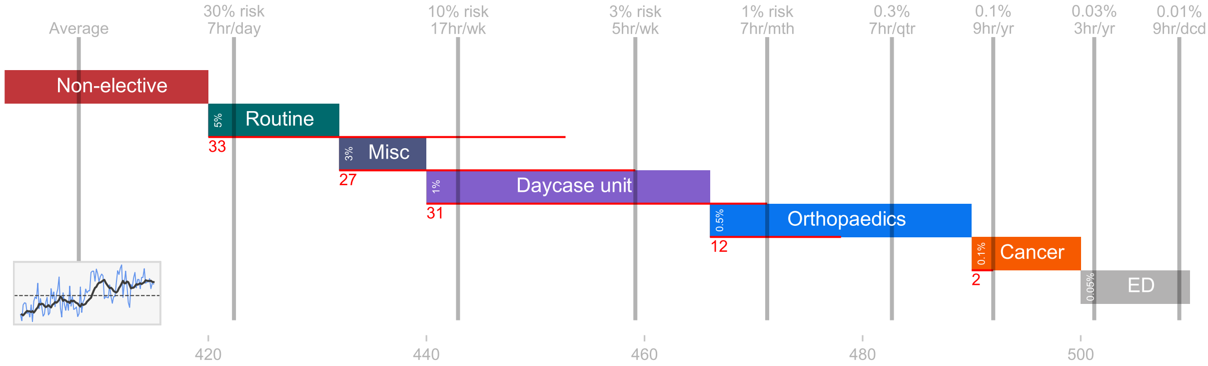


We cannot accept an even a higher risk of stacking patients up in ED or corridors, so we cannot close the newly freed-up beds – in this scenario we have ended up duplicating them on both the hot and cold sites. So although we have gained the benefit of undisrupted elective care on the cold site, we have paid handsomely for the privilege by providing the same beds twice.

6.5 Scenario of a general hospital with high bed occupancy

All that was for a hospital with a benign bed occupancy. But what happens if bed occupancy is already too high – the common position of many hospitals today? Then we might want to move services to a cold site, not so that we could close those beds on the hot site, but to convert them into non-elective beds.

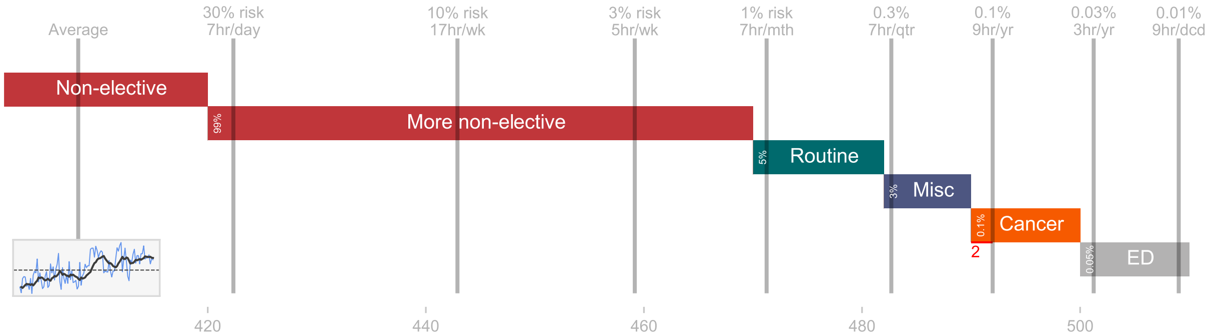
The diagram below shows a general hospital with very high bed occupancy:



We can tell it is high because the average number of non-elective beds occupied is close to the number available. Now the risks of non-elective patients occupying the various kinds of elective bed are unacceptably high. The red numbers indicate how many beds we are short of delivering the acceptable risks, and the largest red number (33 beds, in this case) is the number of extra beds that would achieve acceptable risks throughout.

6.5.1 Moving orthopaedics and daycases to a cold site

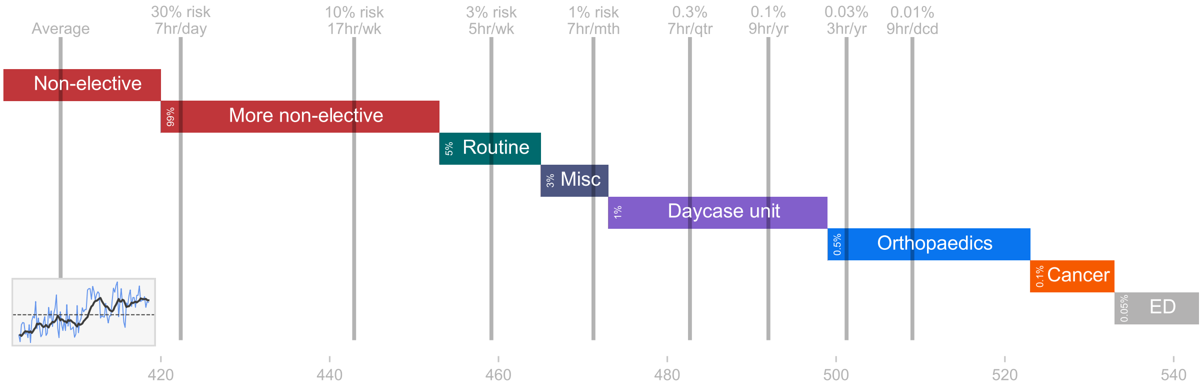
If we try to solve these bed pressures by moving orthopaedics and daycases to new capacity on a cold site, and converting those beds to non-elective on the hot site, then the picture might look like this:



The risks are now acceptable for the routine and miscellaneous beds, and borderline for cancers. And this has come at a high cost, because we have reprovided all the orthopaedic and daycase beds on the cold site, but not closed any beds on the hot site.

6.5.2 Extra capacity on the hot site

If instead we had just created extra capacity on the general site, and left orthopaedics and daycases where they were, we could have achieved acceptable levels of risk like this:



This is effectively identical to the very first scenario we looked at: a general hospital with benign bed occupancy. It achieves an acceptable level of risk everywhere, but with fewer extra beds than the previous 'cold site' scenario.

But it also reveals that we could still move orthopaedics out to a cold site and safely close the beds back on the hot site. So orthopaedics does look like a promising candidate for a cold site, which may explain why so many successful cold sites involve orthopaedics.

6.6 Conclusion

Standing back from the detail, these scenarios show that the qualitative arguments for separating hot and cold facilities didn't resolve the pros and cons clearly enough, because the benefits for the cold site were clear but the drawbacks for the hot site were not. The consequences for the hot site only became clear when we calculated the risks.

We can also see how the balance of risk will vary from one hospital site to another – the numbers of beds will vary and so will the risks. This particular example came to one conclusion, but another example might come to a different one.

In summary, a proper, risk-based, quantitative analysis can help build a well-founded business case on how many beds are needed, and whether a hot/cold split is better than a general hospital model.

Get in touch:

With the author:	rob.findlay@gooroo.co.uk gooroo.co.uk
With NHS England & NHS Improvement:	nhs.ecist@nhs.net