Volatile inpatient costs: implications to CCG financial stability

Dr Rod Jones (ACMA, CGMA)
Statistical Advisor
Healthcare Analysis & Forecasting, Surrey, UK

hcaf_rod@yahoo.co.uk

For further articles in this series please go to: http://www.hcaf.biz
The published version of this article can be obtained from: http://www.bjhcm.co.uk

Key Points

- The year-to-year volatility in costs related to occupied beds increases exponentially as size decreases and the majority of CCGs will experience greater than ± 9% volatility due to their small size
- Volatility is highly location specific depending on the prevailing environment (weather, air quality, infectious outbreaks)
- Even large risk pools containing 1 to 3 million head experience anywhere between 2% to 6% average volatility depending on which locations are in the risk pool
- The most equitable risk pools would contain a mix of disparate locations such that the average volatility would be around 3.3%, i.e. the volatility observed at national level
- Financial planning (including retaining of surpluses) and risk instruments of considerable complexity will be required to enable CCGs to remain financially stable over long periods of time
- Indeed it is questionable if any commissioning organisation of less than 1 million head can be financially viable in the long term

Abstract

The introduction of GP commissioning almost assumes that there is a weak link between financial risk and size and that risk is independent of location-specific environmental factors, i.e. costs are entirely predicted by a capitation formula based on population demographic and socio-economic variables. This is the first of three articles seeking to investigate the volatility associated with different aspects of health care costs, namely, occupied beds as a proxy for inpatient costs, cancer care as an example of initial diagnosis triggering a cascade of costs and death as a proxy for end of life care costs. The observed volatility across these three domains can be compared to give an estimate of the real world volatility in the total costs of a commissioning budget. This study extends the previous analysis of year-to-year volatility at national level down to local level using English PCT populations mainly aligned to local authority boundaries. The difficulty of formulating equitable risk sharing instruments in the presence of significant environment- or location-specific volatility is discussed along with the implications to population- or person-based funding formula.
Introduction

In England, health service reforms have placed GPs in charge of the majority of NHS resources to commission healthcare for their patients. Whilst some may argue that GPs are well placed to do this, after all, they see more than 1 million people per day; this may vastly underestimate the level of expertise or sophistication needed especially in the area of maintaining financial stability. Under the “localism” agenda, GPs have been encouraged to form commissioning groups, called clinical commissioning groups (CCGs) based on small clusters of GP practices. Whilst the theory of “localism” insists that small is ‘effective’, there may well be a hidden flaw in the argument.

The flaw arises due to the dual effects of size and the environment (weather, air quality, infectious outbreaks) on volatility in costs (Jones 2011e, 2012a,c). Smaller organisations experience high volatility due to the fact that the standard deviation associated with health care events is equal to the square root of the average. Hence the 95% confidence interval associated with national volumes of, say, 1 million admissions is merely ± 0.3% but with 100 admissions is ± 30% (Jones 2009a). Additional volatility arising from the environment (weather, air quality, infectious outbreaks) can be 3-times higher than from statistical variation (Jones 2004) and can show systematic trends (cycles) over time depending on the environment (Jones 2011c,e). The impact on financial management should be obvious – but is usually ignored, especially where year on year break-even is an assumed sign of ‘good’ management. The consensus of the literature is that a minimum population of 100,000 head is required to deliver reasonable financial stability (Jones 2009a) although these studies have not included any allowance for the additional contribution of the environment.

Analysis of total PCT expenditure suggests that a CCG with a budget of £100 million could expect average year-to-year volatility in costs of around ± 6% (Jones 2012b). This may seem high and hence this paper will seek to cross-check this against the natural volatility associated with total occupied bed days. In 2010/11 acute inpatient costs accounted for somewhere between 19% (Camden) and 34% (Dorset) of total PCT expenditure and year-to-year volatility in these costs would therefore have the potential to create a significant source of financial risk with knock-on instability in other budgets. This potential appears to be confirmed by recent research into the natural levels of volatility associated with both admissions and bed occupancy seen over very long time periods (Jones 2010b, 2011e).

This study uses a nine year time series (2002/03 to 2010/11) in total occupied bed days (elective plus emergency admission types) for the smaller pre-2004/05 English PCT populations to demonstrate that bed day related costs at local level are highly volatile and show distinct patterns over time. The implications to financial stability and management in CCGs are explored along with the necessary risk pooling arrangements needed to mitigate these risks.

Occupied Bed Days

The year-to-year variation in occupied beds has been chosen to demonstrate the potential volatility associated with acute inpatient costs for several reasons. The first, which is unique to England, lies in the introduction of the four hour target for a stay in the emergency department (ED). It has been argued that this target was never fully achievable and led to ED attendances being re-badged as emergency ‘admissions’. The introduction of the target also coincided with the opening of emergency assessment units and such assessment activities were classified as ‘admitted’ care in order to fall outside of the scope of the target (Jones 2011b). Occupied beds are counted at midnight and avoid these issues since all same day or zero day stay admissions are excluded. Bed days also tend to weight the admission by severity, in that a less complex patient has a shorter stay and cost (Davenport et al 2005). The use of total bed days also avoids issues around ambiguity in clinical coding and since the number of occupied beds (distinct from number of available beds) in England has been fairly constant for many years (Jones 2009b) this avoids major issues around correcting the...
trends for growth (the equivalent to both growth and inflation adjustment of costs). Issues around
the shift of overnight to day case have been mitigated by assigning a nominal one day stay to all day
case admissions and adding this to the total for overnight stays. From a purchaser perspective in a
Payment by Results (PbR) transaction environment bed days represent an excellent approximation
to total cost, especially when percentage differences between years are used. While the provider
perspective is not the main thrust of this article the issues discussed will have direct application to
capacity planning, average occupancy and the knock-on effects to hospital acquired infection (Jones
2011a,c). Finally, it has recently been demonstrated that admissions, bed days and costs for England
all follow the same long term cycle (Jones 2012a). Given the above reasons changes in occupied bed
days is an excellent approximation to changes in total inpatient costs and this allows us to study the
volatility associated with a time series at local rather than national level. Having excluded particular
anomalies which could arise from PbR the estimate of financial risk should therefore be
conservative. Based on 2010/11 data there are around 1.1 occupied bed days per head of population
or 3 occupied beds per 1,000 head.

**Small Organisations**

Figure 1 explores the average year-to-year volatility in bed demand (occupied beds) and hence costs
for English PCTs over a nine year period. The largest PCT on the right is around 1.3 million head while
the smallest is around 100,000 head. As can be seen the average volatility in bed days (costs) rises
rapidly as size reduces. The line labelled ‘1.3 standard deviation’ has been added to give an overall
context to both the average and the slope of the relationship. One standard deviation can be
approximated as the square root of the number of occupied beds (Jones 2011e). The average is high
due to the fact that both statistical and environmental (weather, air quality, infectious outbreaks)
factors contribute to the overall level of volatility (Jones 2011e, 2012a). However, depending on a
host of environmental (location); and with possibly some contribution from structural factors
(perhaps related to deprivation, number of GPs per 1000 head, population density, etc), some health
systems experience higher intrinsic volatility than others.

The key point to note is that the smallest PCTs having around 200 occupied bed equivalents of
population (around 100,000 head) are near to the size of most CCGs and will therefore experience
an average year-to-year volatility of around 9%. The full range in volatility is 4% to 20% depending on
location and by implication it is far harder to forecast future costs and achieve break-even in some
locations than others. While a few PCTs may have experienced only 4% volatility, either due to
chance or location the fact remains that this is more than likely to have had nothing to do with
management competence. This intrinsic difference between PCTs based on location will now be
explored.

This picture of financial volatility was compared to the volatility associated with PCT total admissions
(after adjusting for growth, counting changes and using the square root of the average length of stay
as an approximation to the additional volatility arising from the multiplication of activity by price)
and average (range in brackets) volatility was estimated at 10 000, 20 000 and 100 000 admissions to
be around ± 8% (4% - 22%), ± 6% (2.8% - 10%) and ± 3% (1.8% - 7%) respectively (data not shown).
This confirms the analysis using bed days, i.e. beds days are as good a basis as any to investigate
volatility in costs over time. The implied high volatility in bed related/inpatient costs therefore
confirms the high volatility seen in total PCT expenditure (Jones 2012b).

According to the Department of Health ([http://healthandcare.dh.gov.uk/context/consortia/](http://healthandcare.dh.gov.uk/context/consortia/)) CCGs
range in size from 15,000 to 1.2 million head (average 203,000 and median 167,000) with around
24% smaller than 100,000 head. This gives a range in bed equivalent size from 45 to 3,600 occupied
beds (average 610, median 500) and 300 occupied beds for the suggested minimum population of
100,000. Figure 1 implies that the average volatility in inpatient costs for most of these CCGs is
Location-specific Effects

The difference between each location and the 1.3 standard deviation line was calculated and locations ranked according to this difference. This is presented in Table 1 where the best and worst 25 locations are listed for the average volatility. The best 25 locations appear to be characterised by relatively rural or low population density locations with lower than average deprivation while the worst 25 locations are mainly situated in London or higher population density urban locations. Since there is no mechanism by which the health service can regulate volatility such differences are likely to be unavoidable and Figure 2 explores the relative trend in occupied bed days for such high and low volatility locations as well as a cluster of ‘average’ locations all having around 10,000 occupied beds.

How do we explain the observed environmental sensitivity (and hence volatility) implied by these findings? It has been known for many years that numerous common illnesses and conditions show seasonal behaviour (Fleming et al 1991). The expression of seasonality is variable from one year to the next and this provides the basis for year-to-year volatility (HESonline 2010, Jones 2010b). For example, a 14 year time series of Canadian hospital admissions has demonstrated that 33 out of the 52 highest volume diagnoses are strongly (i.e. bronchiolitis, urinary tract infection, uterine fibroids, etc) or moderately (i.e. diabetes, tonsillitis, pancreatitis, etc) seasonal (Upshur et al 2005). Hence the total costs of inpatient care are strongly related to the environment for 60% of the highest volume admissions. Another study based on a 16 year time series of emergency admissions in New Zealand reached a similar conclusion with 12% of all diagnoses (accounting for 40% of total emergency admissions) showing high sensitivity to the environment. In the New Zealand study it was only the environment sensitive diagnoses which were responsible for the sustained rise in emergency admissions over the 16 year time period along with evidence for long-term cyclic behaviour (Jones 2012c).

It has been proposed that a large part of the additional environmental volatility may be due to two infectious outbreaks which occurred around September in 2002 and 2007 (Jones 2010a, 2012a-c). Figure 2 is consistent with this hypothesis in that infectious outbreaks are known to show spatial granularity (Ruiz-Moreno et al 2010). Hence the low volatility locations are characterised by a relatively modest response to the 2002 and 2007 events, especially so for the 2007 event while the high volatility locations show a consistently large response to both these events. The response in the ‘average’ locations is somewhere between the two extremes. Whatever the fundamental cause the volatility shows a consistent time pattern which requires explanation. These cycles then establish corresponding cycles in the average occupancy experienced in hospital bed pools in England and elsewhere around the world (Jones 2011c,e).

At this point it is vital to note that cost savings merely reduce the average cost over time and do not address the issue of volatility since this is an unavoidable source of variation determined by the laws of mathematics (variation around an average due to size) and the variability in the environment and how this impacts on the expression of poor health (Jones 2012a-c). Irrespective of whether the time pattern is a fundamental property of population health or not we need to explore how best to shield individual locations from undue financial risk.

Risk Pools

In an ideal world with a homogeneous population (only approximated in large population groups) and the absence of location-specific environmental effects the financial risk can be estimated using simple actuarial principles (Jones 2009a). However given the presence of location-specific effects
and the generally very small size of CCGs we need to explore how to mitigate financial risk by formulating equitable larger risk pools (Jones 2008). If we follow the average line in Figure 1, a risk pool with around 1.3 million head (3,000 occupied beds) would experience roughly 2.5% average (unavoidable) volatility in costs while a single CCG with 200 occupied beds (equivalent population) would experience 9.3% volatility and up to 15% for one with around 100 occupied bed equivalents (approx 45,000 head), etc. Extrapolating the trend line to 10,000 occupied beds suggests that around 4 million head are required to reduce the average year-to-year volatility down to ± 1.3%, i.e. around 15 Health Authority sized organisations should be involved in the ‘head office’ support (executive, statutory, governance, data processing, information, commissioning) and risk management functions. This is close to the minimum size for US health insurance companies (Woolhandler et al 2003).

However, in practice, constructing random groups of PCTs with a range around 10,000 occupied beds (see box at far right of Fig 1) leads to average volatility of between 2% and 6% (over the nine year period). This range is due to the location-specific nature of volatility identified above. The plateau between 7,000 to 10,000 occupied beds is due to two competing forces. Firstly the trend to lower statistical risk as size increases and secondly to the mix of low and high volatility locations which act to constrain the expressed volatility in England. This is explored further in Figure 3 for risk pools consisting of populations greater than 10,000 equivalent occupied beds. As size increases beyond 10,000 occupied beds the maximum possible volatility declines due to the effect of size while the minimum possible rises due to the fact that there are only a limited number of very low risk locations. The English average of 3.3% is the ultimate balance between these forces.

The key point that Figure 3 re-enforces is that regional risk pools (while perfectly appropriate in a homogeneous world) are a totally inappropriate way to minimise risk. For example, a risk pool for London would contain predominantly high volatility locations (as per Table 1) and would lie on the maximum possible risk line in Figure 3. The only equitable way to share risk is to construct the risk pool as a mixture of locations (Jones 2012a) whose net effect is close to national average. This is totally in the opposite direction to current policy around CCGs.

Indeed the whole issue of location-specific volatility points to significant flaws in the equitability of the current version of capitation formulae and suggest that successful financial management may be more difficult than first appears to be the case, especially if you are in the ‘wrong’ location.

Efficiency & Volatility

Every health care system (primary plus acute care) reaches a point of equilibrium which is unique to that system and this includes admission rates for particular diagnoses and procedures (Jones 2006a,b). High levels of variation in apparent admissions rates between different health care systems are well documented, e.g. NHS Atlas of Variation (http://www.rightcare.nhs.uk/index.php/nhs-atlas/).

The point to consider is whether inter-location variation (i.e. higher intervention rates in some areas than others) is contributing to the volatility over time observed for each location in this study. In this respect the Torbay Care Trust is well recognised for its 30% lower consumpption of acute bed days relative to other PCTs (Jones 2011d). Torbay ranks 83 out of 303 PCTs (rank of 1 for lowest volatility relative to the 1.3 standard deviation line) and has a volatility of 5.2% compared to 6.8% for average behaviour at the same size. Hence if 'efficiency' (i.e. low rates of admission) were the issue then Torbay should occur in the 'Best 25' list in Table 1 rather than somewhere in the second quartile. It could well be argued that less efficient (less integrated) health care systems may be more sensitive to environmental challenges (Jones 2012c) and if this were so then, based on Torbay, a reduction in average volatility of 1.6% (6.8% - 5.2% = 1.6%) may be possible. This possibility needs to be investigated.
However if we consider the fact that different Western health care systems (with vastly different models of funding and management) all appear to follow a similar time pattern to that demonstrated in Figure 2 (Jones 2012d,e) then the possibility of a common cause needs to be considered. Up to the present the calculation of differences in admission rates has assumed that they are due to population demographic and socio-economic factors (i.e. the capitation formula). Environmental differences are usually ignored due to the difficulty of constructing national models for their effects. This does not mean they are insignificant and one study has demonstrated that even background noise levels are sufficient to account for a 10% difference in all cause emergency admissions (Tobias et al 2001). It would seem that far more attention is required to this neglected area of research into the environment-induced volatility associated with costs.

Three interpretations of the events seen in 2002 and 2007 are possible. The first would be some form of random congruence in environmental conditions leading to unexpected very high volatility about once every five years. The alternate is a recurring step-like event which has increased bed demand by over 6.5 million beds days, i.e. in the absence of these two step increases there would now be 18,000 fewer occupied beds and the NHS would have avoided around £3 billion of cumulative costs (assuming a marginal cost of only £75 per extra bed day). A similar series of unexplained step increases in occupied beds have been shown to occur in Australia (Jones 2011e). The third would involve some form of coordinated change in the threshold to admission which would contradict research conducted in the USA which demonstrated that the hospital admission threshold remains constant even in the face of fluctuating demand (Sharma et al 2008). An accompanying article in this edition investigates the behaviour of GP referrals and outpatient attendances and concludes that biological rather than organisational behaviour seems more probable (Jones 2012c). Far greater research is required to fully understand these issues within the wider context of the external environment (weather, air quality and infectious outbreaks) and its effect upon health care costs.

Conclusions

From the perspective of cost saving and reducing inter-location variation in admission rates the application of “localism” to small CCGs is entirely appropriate and will undoubtedly achieve these aims. However from the perspective of financial stability we can say that this is an inappropriate application of “localism” since it loses economies of scale and exposes each CCG to unacceptably high levels of financial risk which arise from the inherent volatility due to size, the environment and the fundamental cycles in health care costs (Jones 2012a-e). This will inevitably lead GPs (in some locations more so than others) into an ethical dilemma with decisions dictated by the financial pressures arising from the local volatility in costs. In the world of business these issues would be partially resolved by having local branches and a single head office – simultaneously delivering “localism” and economy of scale. Indeed is this ‘reform’ an attempt to solve the wrong problem? Could the correct solution be to understand why the diagnoses showing high environmental sensitivity are showing high growth in admissions and costs (Jones 2012a-e) and solve the cause rather than seeking to apply structural and organisational solutions to the symptoms?

References

http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=972
An edited version of this article has been published as: Jones R (2012) Volatile inpatients costs: CCG financial stability. British Journal of Healthcare Management 18(5): 251-258. Please use this to cite.


Figure 1: Volatility in bed demand for English Primary Care Trusts

Footnote: Data covering total occupied bed days for elective and emergency admission types to acute, mental health and maternity over the period 2002/03 to 2010/11 is for PCT responsible patients and was obtained from the HES Online website (http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=213). The absolute difference in occupied bed days from one year to the next was calculated as a percentage of the first value in each paired set and then averaged over the time series. Bed days were converted to occupied beds by division by 365 days per year. Occupied beds were calculated as a nine year average. The locations covered by PCTs were for the pre-2005/06 NHS changes, however, some larger post-2005/06 configurations have been calculated to extend the range above 1,000 occupied beds.
An edited version of this article has been published as: Jones R (2012) Volatile inpatients costs: CCG financial stability. British Journal of Healthcare Management 18(5): 251-258. Please use this to cite.

Figure 2: Trend in occupied beds for various combinations of localities

Footnote: Clusters of between 20 to 25 PCTs were selected with close to 10,000 occupied beds. All combinations have been adjusted to an average of 10,000 occupied beds over the nine year period.
An edited version of this article has been published as: Jones R (2012) Volatile inpatients costs: CCG financial stability. British Journal of Healthcare Management 18(5): 251-258. Please use this to cite.

**Figure 3: Maximum and minimum volatility in different sized risk pools**

Footnote: Average volatility was calculated for random groups of between 20 to 300 localities (using pre-2005/06 PCTs as a proxy for localities). National average is 3.3% for the nine year period.
An edited version of this article has been published as: Jones R (2012) Volatile inpatients costs: CCG financial stability. British Journal of Healthcare Management 18(5): 251-258. Please use this to cite.

Table 1: Best and worst PCTs for financial stability

<table>
<thead>
<tr>
<th>Best 25</th>
<th>Gap</th>
<th>Worst 25</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>5DT North East Oxfordshire</td>
<td>-6.2%</td>
<td>5NC Waltham Forest</td>
<td>6.2%</td>
</tr>
<tr>
<td>5GN Uttlesford</td>
<td>-5.7%</td>
<td>5LW Northampton</td>
<td>5.9%</td>
</tr>
<tr>
<td>5DX South East Oxfordshire</td>
<td>-5.4%</td>
<td>5NA Redbridge</td>
<td>5.8%</td>
</tr>
<tr>
<td>TAG Witham, Braintree &amp; Halstead</td>
<td>-5.2%</td>
<td>5LD Lambeth</td>
<td>5.7%</td>
</tr>
<tr>
<td>5GL Maldon &amp; South Chelmsford</td>
<td>-4.8%</td>
<td>5HY Hounslow</td>
<td>5.7%</td>
</tr>
<tr>
<td>5AG South Peterborough</td>
<td>-4.6%</td>
<td>5A5 Kingston</td>
<td>5.6%</td>
</tr>
<tr>
<td>5E5 Eastern Hull</td>
<td>-4.5%</td>
<td>5K7 Camden</td>
<td>5.4%</td>
</tr>
<tr>
<td>5DN Wokingham</td>
<td>-4.2%</td>
<td>5MD Coventry Teaching</td>
<td>5.2%</td>
</tr>
<tr>
<td>5E3 East Yorkshire</td>
<td>-4.4%</td>
<td>5CN Herefordshire</td>
<td>5.1%</td>
</tr>
<tr>
<td>5H6 Ellesmere Port &amp; Neston</td>
<td>-3.9%</td>
<td>5H1 Hammersmith &amp; Fulham</td>
<td>5.1%</td>
</tr>
<tr>
<td>5AN North East Lincolnshire</td>
<td>-3.9%</td>
<td>5C1 Enfield</td>
<td>4.8%</td>
</tr>
<tr>
<td>5JJ South Cambridgeshire</td>
<td>-3.6%</td>
<td>5MQ South Warwickshire</td>
<td>4.8%</td>
</tr>
<tr>
<td>5FX Mendip</td>
<td>-3.7%</td>
<td>5MA Crawley</td>
<td>4.7%</td>
</tr>
<tr>
<td>5EC Gedling</td>
<td>-3.3%</td>
<td>5HX Ealing</td>
<td>4.7%</td>
</tr>
<tr>
<td>5D4 Carlisle &amp; District</td>
<td>-3.5%</td>
<td>5L East Southwark</td>
<td>4.4%</td>
</tr>
<tr>
<td>5DC Harlow</td>
<td>-3.3%</td>
<td>5L4 Swale</td>
<td>4.3%</td>
</tr>
<tr>
<td>5H7 Derbyshire Dales &amp; South Derbyshire</td>
<td>-3.2%</td>
<td>5HW Newcastle-Under-Lyme</td>
<td>4.3%</td>
</tr>
<tr>
<td>5HH Leeds West</td>
<td>-3.3%</td>
<td>5C3 City &amp; Hackney Teaching</td>
<td>4.3%</td>
</tr>
<tr>
<td>5HC South Liverpool</td>
<td>-3.3%</td>
<td>5KH Hambleton &amp; Richmondshire</td>
<td>4.3%</td>
</tr>
<tr>
<td>5FY Teignbridge</td>
<td>-3.2%</td>
<td>5L Westminister</td>
<td>4.1%</td>
</tr>
<tr>
<td>5GK Royston, Buntingford &amp; Bishop's Stortford</td>
<td>-2.9%</td>
<td>5M6 Richmond &amp; Twickenham</td>
<td>4.1%</td>
</tr>
<tr>
<td>5EE North Sheffield</td>
<td>-3.3%</td>
<td>5LX Fareham &amp; Gosport</td>
<td>4.0%</td>
</tr>
<tr>
<td>5F4 Heywood &amp; Middleton</td>
<td>-2.9%</td>
<td>5L East Kent Coastal</td>
<td>3.9%</td>
</tr>
<tr>
<td>5ER Erewash</td>
<td>-2.9%</td>
<td>5K Durham &amp; Chester-Le-Street</td>
<td>3.8%</td>
</tr>
<tr>
<td>5EG North Eastern Derbyshire</td>
<td>-3.1%</td>
<td>5L Adur, Arun &amp; Worthing</td>
<td>3.8%</td>
</tr>
</tbody>
</table>

Footnote: The gap is calculated relative to the 1.3 standard deviation line