End of life and financial risk in GP commissioning

Dr Rod Jones (ACMA, CGMA)
Statistical Advisor
Healthcare Analysis & Forecasting, Camberley, Surrey
hacf_rod@yahoo.co.uk

Further articles in this series are available at: http://www.hcaf.biz
The published version is available at http://www.bjhcm.co.uk and those who have an Athens login can obtain copies via this route.

Key Points

- Health care costs rise toward the end of life especially in the last six months of life
- The year-to-year volatility in the number of deaths seen at local authority and regional level across the UK is both surprisingly high and locality dependant
- For areas with 1000 deaths per annum the average volatility ranges from 3% to 8% depending on location.
- A long-term cycle in deaths appears to mirror the cycle observed in medical admissions, GP referrals and emergency department attendance.
- A common infectious source is implicated with the 2002 and 2007 events leading to over 8,000 excess deaths in the 12 months following each event.
- Large cost savings in acute admissions for death in hospital may be an illusion created by the HRG tariff

Abstract

Rising health and social care costs toward the end of life mean that this represents an area of high financial risk. A study of the year-to-year volatility in deaths at local authority level across the whole of the UK shows that this volatility is location specific, i.e. some locations will experience higher volatility in end of life costs and hence knock-on effects upon other expenditure types within the envelope of a fixed total budget. Across the UK two events in 2002 and 2007 led to increases in deaths which were as high as that observed in major influenza epidemics seen in the 1980’s and 90’s, yet without the presence of unusual levels of influenza. These dates correspond to step-like increases in emergency medical admissions, A&E attendances and GP referrals. Potential reasons are discussed.
Introduction

In England total deaths have been declining since around 1993 and reach a minimum around 2012 and are then set to rise, although the extent of the rise is highly location specific (Jones 2011). In 2006 some 27% of the $330 billion (and growing) Medicare budget in the US was spent during the last year of life (Appleby 2006) with 52% spent in the last 2 months, and 40% in the last month. Inpatient expenses accounted for over 70% of the total costs and a considerable part of this was due to ineffective stays in intensive care (Luce & Rubenfeld 2002). In England some 58% of deaths occur in hospital (Department of Health 2008) and it has been estimated that the cost of care delivered to patients dying with cancer (27% of total deaths) was £1.8 billion and for those dying of organ failure £0.5 billion (Hatziantheou et al 2008).

Total inpatient and social care costs in the last year of life rise from around £7,000 per person dying at 45 up to £11,000 for death at 100 years. Costs are higher than this average for those with mental health problems, renal failure and cerebrovascular disease, and as in the USA, total costs vary by location (Bardsley et al 2010). The time period associated with end of life therefore represents a considerable potential cost risk to the budgets of health care organisations. Indeed, in the elderly, the occurrence of an emergency admission is highly significant and one study showed that for those aged over 70 subsequent to such an admission some 23% were dead within six months, 14% had been re-admitted and 5% had been moved to a nursing/residential home (Round et al 2004). The authors own research indicates that there are 170 diagnoses accounting for 8% of emergency admissions where 50% to 100% of persons are dead within 90 days of the admission date.

The last year of life is merely the end of a period of decline. Those who die prematurely usually have around three to four years of functional decline prior to death while those who die at an advanced age usually experience less than one year of functional decline (Mayhew 2001). While gauging the period to the end of a persons life may be difficult a pointer to the role of ‘functional decline’ comes from a study of healthy men aged over 70 which showed that those who walked faster than 1.8 mph were 1.23-times less likely to die, while those able to walk faster than 3 mph were highly unlikely to die within five years (Stanaway et al 2011). Given this time-dependant cascade in functional decline it should not be surprising that various types of health and social care costs follow different time trajectories (Mc Grail et al 2000). While hospital costs start to increase 15 years prior to death they are most prominent in the last year of life with a 10-times increase from five years to the last year. This increase overshadows a general 30% (1.3-times) increase from 65 to 85 years for decedents (Seshamani & Gray 2003).

Hence while death *per se* marks the cessation of health care expenditure, death signifies that, whether knowingly or unknowingly, that person has generally been in the highest cost period of their adult life and that these costs are committed to by virtue of the forthcoming death. In the context of the financial risk associated with end of life it is therefore useful to
study the volatility associated with annual deaths as the fundamental source of the risk attached to this expenditure. Factors influencing the volatility in deaths and end of life costs will now be discussed.

**Seasonal Factors**

To fully understand the volatility in costs implied by the approach of death it is important to understand the seasonal nature of death. To illustrate, Figure 1 presents a daily view of deaths in England & Wales over the period 1989 to 1999.

![Figure 1: Variation in daily deaths throughout the year (1989 to 1999)](image)

Footnote: Daily count of deaths in England & Wales kindly provided by the UK Office of National Statistics.

This period has been used because it was the only time over the past 50 years where total deaths in each year were relatively constant. The maximum and minimum values apply to each day of the year and can come from any of the years. In reality, the maximum or minimum is usually comprised as a string of days from a particular year when events such as influenza or other infectious outbreaks (for the maximum – see line for 1993 as an example) or extended mild weather in the absence of infectious outbreaks (for the minimum – see line for 1994 as an example) dominated the national scene. The curve is U-shaped with the maximum in winter and the minimum in the period toward the end of August. During the
period 1989 to 1999 influenza epidemics were the dominant force affecting the maximum line (especially in winter). Hence it is largely the winter period which determines the absolute number of deaths in any 12 month period. Deaths and emergency hospital admissions (especially medical) usually rise and fall in parallel. At these national totals the variation in daily deaths due to Poisson randomness has a standard deviation of ± 36 deaths (± 2.8%), i.e. the charts are dominated by environmental rather than statistical factors.

The minimum line in Figure 1 is probably due to the effect of temperature on the death rate where minimum mortality occurs at a (constant) daily average of 17 °C (occurring around July & August in England) but has increased by 35% at a daily average of 24 °C or -2 °C, although complex time lags apply for short periods of severe cold or heat (Ekamper et al 2009). In this respect the minimum line in Fig 1 is 31% and 44% higher in December and January respectively which reflects the generally lower temperatures in January. Hence temperature alone is only able to explain the minimum line in Figure 1 and the remainder of excess mortality is largely due to infectious outbreaks, of which influenza is but one example. We will return to this vitally important issue when discussing Figure 4.

Location Specificity

Earlier articles in this series have demonstrated that many health care cost categories display location specific volatility (Jones 2012a-e). To investigate further, annual deaths between 1993 and 2009 for over 400 UK-wide local, county and regional authority boundaries were adjusted for the general decline in deaths over the time period and the adjusted year-to-year volatility was then calculated and averaged. Figure 2 presents the results of this analysis along with lines for the volatility expected from Poisson statistical variation in deaths, i.e. the minimum case scenario which would occur at constant 17 °C in the absence of infectious outbreaks or other environmental effects. The lines for Minimum and Maximum Chance are the 99.99% confidence intervals which would arise from Poisson sampling with 18 years data. As can be seen the vast majority of locations fall between the average and the maximum indicating that all locations experience additional environmental increments in mortality. Some areas even fall above the 99.99% confidence interval indicating very strong environmental contributions to the volatility in death over time. At this point recall that we are measuring volatility in death (and thus associated costs) over time not the absolute death rate per se.

In terms of recent government policy around CCGs, Figure 1 tells us that environment-based volatility exists while Figure 2 demonstrates that at local (smaller) level. Hence below 10000 deaths per annum, statistical randomness acts to obscure reality, i.e. the statistical volatility begins to outweigh the environmental volatility. Hence while cost saving initiatives can be implemented locally it is not possible to ‘reward’ locally since the outcome is now dominated by the size-based statistical randomness. Rewards can only be allocated based on collective performance over larger localities, i.e. regions having greater than 10000 deaths per annum. It would seem that the recently enacted Health and Social Services bill
An edited version of this article has been published as: Jones R (2012) End of life and financial risk in GP commissioning. British Journal of Healthcare Management 18(7): 374-381. Please use this to cite.

has some very serious ‘real world’ flaws, in that it is not possible to do in the real world that which the bill purports to achieve.

**Figure 2: Average year-to-year volatility in deaths**

Footnote: Data for over 400 local authority, county and regional areas across England, Scotland, Northern Ireland and Wales were kindly provided by the UK Office for National Statistics. The year-to-year volatility was calculated as the absolute percentage difference between successive pairs of values less the value of assumed linear reduction in deaths over time (see Jones 2012b,c). For some locations the reduction in deaths was calculated as two linear segments. Maximum, minimum and average chance was calculated by Poisson Monte Carlo simulation (Oracle Crystal Ball software) using 17 data points (equivalent to 17 years data). The average volatility was calculated as per the deaths data. Above an average of 1,000 a Normal approximation to a Poisson distribution was used in the simulation which was repeated 10,000 times to give the equivalent to 99.99% confidence intervals.

Also of relevance is the fact that larger regions all experience far higher than expected volatility due to the fact that, to a greater or lesser extent, each larger region experiences a similar set of environmental conditions, i.e. the location specific nature of volatility is simply aggregated. Figure 3 explores this issue in greater detail by presenting the lines of maximum and minimum volatility for various hypothetical and actual regional configurations. In this figure the risk is purely around death per se and if the cost of dying were incorporated the whole figure would be shifted up to higher percentage values due to the fact that costs per death are variable and are the equivalent to sampling from a distribution of all possible death related costs.
An edited version of this article has been published as: Jones R (2012) End of life and financial risk in GP commissioning. British Journal of Healthcare Management 18(7): 374-381. Please use this to cite.

As has been stated earlier this figure merely confirms the fact that regional risk pools are themselves subject to considerable location-specific volatility and should share risk between themselves rather than with their own constituent locations (CCGs), which would only act to reinforce inequality, i.e. the post code lottery (Jones 2012a-d).

**Figure 3: Maximum and minimum volatility for larger risk pools**

Footnote: Locations were ranked from lowest to highest (or highest to lowest) volatility and one location was then added incrementally and the volatility re-calculated for such growing pools of locations. The actual data points represent larger counties or regions. The three points at the right hand end represent England, England & Wales and the UK respectively. Points for Wales, Scotland and Northern Ireland lie amid the other regional data points.

Assuming that gradients in health differentially interact with changes in the environment the proposal that location has a marked effect on the volatility in deaths could be supported by results from the US where people who report fair to poor health are higher in rural areas (15.7%) and rural towns (13.3%) compared to the urban fringe (7.9%) surrounding large cities (Ingram & Franco 2012). High and low mortality tends to be persistent over long time periods (Cossman et al 2002). Up to the present there are no studies to show if volatility is intrinsically higher in areas of high mortality, i.e. are less healthy populations more sensitive to environmental fluctuation and infectious outbreaks? It would appear that further research is required to resolve these issues.
Prior to December 1999 influenza epidemics and more local winter influenza outbreaks were very common. However between January 2000 and November 2009 the levels of influenza fell to a 100 year minimum. Such events are exceeding rare and the last extended period of the very low incidence of influenza occurred between 1879 and 1889 (Thacker 1986). Hence the period 2000 to 2009 offers a unique insight into the volatility in death in the near absence of influenza. In this respect Figure 4 shows the trend in deaths between 1993 and 2009 after correcting for the ongoing decline in deaths over this period. Influenza epidemics in 1993, 1995 and 1999 are highlighted; however, this cannot explain the very large ‘influenza-like’ peaks with maximum deaths in 2003 and 2008.

**Figure 4: Deaths in England (actual versus expected)**

Footnote: Actual deaths were compared to expected, where: Expected deaths = 531,891 + 974.2 x X – 319.45 x X^2 + 5.3404 x X^3 (R^2 = 0.9), where 1989 starts the series as year 1 through to year 21 at 2009. On this occasion the four years prior to 1993 were added to ensure that the polynomial captured the plateau prior to 1993.

Toward the middle of 2002 and 2007 outbreaks of what has been called an infectious immune disease (or more correctly immune impairment) resulted in the geographic spread of increasing emergency medical admissions, GP referrals, A&E attendances, ambulance journeys, etc, across the entire UK (Jones 2010a-d, 2012f). Full spread across the UK extended through to late 2008 (Jones 2010d, 2012f). The causative agent has been tentatively proposed to be the common herpes virus cytomegalovirus, which causes a cascade of immune impairments especially in the elderly (Jones 2012f). Prior to 2000 the
regular occurrence of influenza was probably masking outbreaks of this virus and judging from the size of the 2003 and 2008 peaks it looks highly likely that particular outbreaks may have been acting to amplify the magnitude of some of the earlier influenza epidemics. For example, the proposed outbreak in early 1993 led to a very high number of deaths in this year – see line for 1993 in Figure 1, especially after 12th February where deaths move upward to shadow the maximum line for most of the year.

The proposed late 1996 winter outbreak requires some comment. The winter of 1996/97 saw a typical period of seasonal influenza which peaked in January 1997; however, while influenza activity was only moderate there were a disproportionate number of deaths (Dedman et al 1997). Hence we have the possibility that dual outbreaks of both influenza (typical spike increase in admissions) and the proposed infectious agent (step change in admissions) acted to cause the observed higher than expected deaths with the slightly later arrival of influenza accounting for the tail of higher deaths stretching from mid-January to mid-February (Jones 2010a). It would appear that we may need to re-evaluate the exact contribution due to influenza alone in the absence of synergistic effects between it and the proposed infectious agent.

The very large peaks due to the outbreaks commencing in 2002 and 2007 probably arise as a consequence of the absence of influenza, i.e. deaths which would have been otherwise initiated by the dual action of influenza and the proposed virus accrued and were expressed via the sole action of the proposed virus. It has been previously observed that approximately 2½ years after the onset of each outbreak that emergency medical admissions suddenly decline back to baseline levels (Jones 2010b), an observation consistent with the known ability of cytomegalovirus to revert to a dormant state. It would appear that increased deaths last only for around 12 months (the most vulnerable die first) while the persistent or active form of the infection endures for a further 18 months (emergency medical admissions remain high) and at this point the immune response in the remaining survivors leads to roughly synchronous transition to viral dormancy. Considerable research is urgently required to verify this behaviour.

The possibility that different mechanisms of infectious spread were operating up to 2000 (with influenza) as opposed to from 2000 to 2009 (without influenza) is explored in Figure 5 where the average volatility (as the average year-to-year difference expressed as ‘standard deviation’ equivalent, i.e. size adjusted) is shown for each location in the before and after scenarios. Volatility should be higher in the ‘before’ period since deaths are being influenced by two major outbreak-types along with the general higher frequency of influenza epidemics and this is generally the case. The second observation (results not shown) is that locations typically showing low volatility (less than 1.5 standard deviation average volatility) in the ‘before’ case tend to show slightly higher volatility (+3%) in the ‘after’ case, i.e. the modus operandi or method of infectious spread and subsequent effect has been altered.

**Excess Winter Mortality**
Excess winter mortality (EWM) is defined by the Office for National Statistics as the difference between the number of deaths during the four winter months (December to March) and the average number of deaths during the preceding autumn (August to November) and the following summer (April to July). The excess winter mortality index is calculated as the percentage excess winter deaths divided by the average non-winter deaths. The very high deaths following the 2002 and 2007 outbreaks were apparently overlooked due to a focus on EWM alone rather than a broader focus on full year deaths. Since each outbreak appears to increase deaths throughout the whole year and not just the winter, this acts to give an underestimation of the true value of EWM.

**Figure 5: Volatility before (with influenza) and after December 1999 (without influenza)**

Footnote: Volatility was re-calculated as standard deviation difference (assuming Poisson variation) between each pair of years rather than as a percentage difference. This acts to partly adjust for relative size. Volatility for the UK, England & Wales and England are all higher after 1999 (without influenza).

This impact of the proposed outbreaks on the calculation of EWM can be discerned in Figure 1 for the 1993 outbreak where the background level of deaths was shifted upward from the end of February onward resulting in very high total deaths for the entire year. Hence EWM in the winter of 1993/94 is subsequently low due to the higher baseline and the absence of a substantial influenza outbreak in that year.
An edited version of this article has been published as: Jones R (2012) End of life and financial risk in GP commissioning. British Journal of Healthcare Management 18(7): 374-381. Please use this to cite.

However despite these limitations the very fact that the immune impairment acts to increase susceptibility to infection in general (Jones 2010c, 2012f) does lead to an overall increase in EWM following each outbreak which is presumably enhanced by co-infection with a range of other common winter viruses (Fleming et al 2005). This increase can be used to track the spread of the new disease. For example, following the 2007 outbreak EWM for those aged 85+ (the proposed immune impairment increases with age) increased first in the North East of England with moderately high EWM for the two winters 2007/08 and 2008/09, all other region show a very high peak in the winter of 2008/09 except for Wales with a late outbreak and hence an extended period of EWM in both 2008/09 and 2009/10 (see data at http://www.ons.gov.uk/ons/search/index.html?newquery=excess+winter+mortality). A broadly similar pattern also applies after the 2002 outbreak except that the maximum EWM is delayed until the 2005/06 winter in Wales suggesting slower spread in that outbreak. This pattern of spread is consistent for the known timing of parallel ‘outbreaks’ of increased GP referral seen first in Scotland then presumably via the North East, rest of England and then Wales (Jones 2010a-c, 2012a-e).

End-of-life Costs

As with inpatient costs where the volatility in admissions is amplified by the cost per admission so with end of life; where the volatility in deaths is amplified by the cost distribution associated with the cost trajectory up to death. Hence while the volatility in deaths is an unavoidable component of the financial risk the cost distribution can be modified to reduce the total risk.

For example, the cost of care in the last week of life for those with cancer was $1,040 lower for those who had an end of life discussion and patients with higher costs (no discussion) had a worse quality of death score (Zhang et al 2009). Another US study on Advance Directives (Living Wills) demonstrated $5,600 lower costs in the last six months of life and a higher proportion of death outside of hospital (Nicholas et al 2011). Likewise the proportion of people who die in hospital shows a wide range between countries from 34% in the Netherlands to 63% in Wales (Cohen et al 2008) as does primary and secondary care usage with one study showing an average of 24 bed days and 76 GP visits in the last six months of life in New York but only 11 bed days and 27 GP visits in San Francisco (Wennberg et al 2004).

Up to the present research on end of life costs has concentrated on estimating the extent of the cost and no studies exist to demonstrate the extent of year-to-year volatility. Figure 4 demonstrates that a very high year is (of necessity) followed by a very low year due to the fact that those who are the most vulnerable succumb first. It is unknown if this hastening of death acts to reduce or increase total costs, i.e. hospital admission is probably brought forward but other health and social care costs are probably curtailed, etc. Whatever the case the real world volatility cannot be lower than that demonstrated here and is unacceptably high especially for small health care purchasing organisations (Jones 2012a-d).
An edited version of this article has been published as: Jones R (2012) End of life and financial risk in GP commissioning. British Journal of Healthcare Management 18(7): 374-381. Please use this to cite.

Cost Savings

The high proportion of people dying in hospital has led to claims that less expensive alternatives need to be found. My own (unpublished) research indicates that the cost of a death in hospital is made grossly high due to how the HRG tariff is calculated. Length of stay for those who die in hospital is usually far lower than the corresponding cohort of patients in each HRG who do not die, but the same price is charged for all. The HRG tariff is not a statement of truth (Jones 2012g) but of illusionary averages and extreme caution is required least genuinely higher cost alternatives get implemented for something which has a lower real cost than its (current tariff) price. This should not be taken to mean that death in hospital is the preferred option, only that the real cost savings may be less than they appear and as alternatives to death in hospital are implemented the residual ‘average’ tariff price will rise to compensate.

Conclusions

Death is the end of an inevitable cascade in health and social care costs as functional capacity declines at the end of life. In the absence of any studies investigating the volatility in these costs death has been used as a proxy to investigate both the magnitude and spatio-temporal variability in costs. Infectious outbreaks are seen to be the main cause of volatility and outbreaks of a new type of infectious immune impairment were apparently masked by influenza in the period up to 2000 and an unusual nine year period of very low background influenza has allowed the huge potential impact of these outbreaks to be revealed. New infectious diseases aside it is noted that to reduce the volatility in costs associated with end-of-life below 3% implies that CCGs cover a population base equivalent to 10,000 deaths or over one million head. It would appear that the tantalising possibility of financially stable small CCGs was never an option (Jones 2012a-d).

References

An edited version of this article has been published as: Jones R (2012) End of life and financial risk in GP commissioning. British Journal of Healthcare Management 18(7): 374-381. Please use this to cite.

http://www.hpa.org.uk/Publications/InfectiousDiseases/RespiratoryDiseases/0501WintersTale/
Jones R (2010a) Additional studies on the three to six year pattern in medical emergency admissions. Healthcare Analysis & Forecasting, Camberley, UK.
http://www.hcaf.biz/Recent/Additional_Studies.pdf
An edited version of this article has been published as: Jones R (2012) End of life and financial risk in GP commissioning. British Journal of Healthcare Management 18(7): 374-381. Please use this to cite.


