The Relationship between Volume and Health Outcomes

Gordon D Murray, Professor of Medical Statistics, University of Edinburgh

1. Introduction

In January 1997 the NHS Centre for Reviews & Dissemination at the University of York published an authoritative systematic review of research evidence on the volume/outcome relationship in healthcare (Sowden *et al*, 1997). In essence the conclusion was that the bulk of research evidence was methodologically flawed and of little value in informing decisions concerning the planning of the delivery of health services. In the summer of 2004 I was asked to re-review the research evidence in the light of papers published since 1997, to inform the National Framework for Service Change in the NHS in Scotland. In view of the time constraints it was requested that my review be a conventional narrative review rather than a formal systematic review.

2. Background

The York Review (Sowden *et al*, 1997) reached three main conclusions on volume/outcome relationships:

(i) **Case-mix**: "*Most of the existing research, because it does not sufficiently take account of differences in case-mix, probably overestimates the impact of volume on the quality of care*." (Summary Report, Page 10).

(ii) **Causation**: "*… because none of the research indicates that increasing activity over time leads to improvements in clinical outcome, it is difficult to infer from results of cross-sectional studies which show better outcomes in higher-volume units that similar differences in outcomes can be expected by the expansion of an existing unit*." (Summary Report, Page 10).

(iii) **Thresholds**: "*The most that the research evidence can support is a conclusion that if there are significant quality gains from increased volume, these*

gains appear to be exhausted at relatively low threshold levels. Volumes of activity above these thresholds should be achievable without significant structural changes, but may require a more sharply defined internal division of labour across consultant staff (which may be consistent with increased sub-specialisation within disciplines)." (Summary Report, Page 11).

3. Methods

I undertook a general literature search on volume/outcome relationships together with more focused literature searches on methodological aspects of volume/outcome relationships, on studies relating to volume/outcome relationships which evaluated the impact of an intervention to increase regionalisation, and methodological aspects of assessing clinical 'learning curves'. These searches concentrated on publications from 1997 onwards, although some key references predate this.

These literature searches showed that more relevant papers have been published between 1997 and 2004 than were published up to and including 1996, and so it would indeed be a major undertaking to perform a formal systematic review of the more recent literature. My review was therefore highly selective, based on reading abstracts and obtaining full papers when appropriate and when they were relatively easily accessed. As most papers in this area are observational and cross-sectional by design, and are often very specific to a local area (e.g. a single US state), they tend to be published in low impact journals which are not held by the local University libraries. Thus most of the full papers reviewed are either from high impact journals or from journals whose contents can be accessed electronically.

4. Results

Over 500 abstracts were scanned and 50 full papers reviewed in detail. The vast majority of the papers relate to surgical procedures with outcomes assessed in terms of short-term (in hospital or 30 day) mortality. There is also an increasing number of papers relating to surgical oncology with outcomes assessed in terms of long-term (5 years plus) survival.

Many common procedures have low surgical mortality rates, so that even if a volume/outcome relationship does exist for mortality it would be difficult to demonstrate and might be of limited clinical relevance. So studies focussed on morbidity outcomes are also becoming more common. For example, avoidance of a stoma in bowel surgery has a major impact on quality of life and is a very relevant patient-centred outcome measure. Similar issues apply in prostate cancer (Begg at al, 2002) and in many other areas.

Two particularly useful systematic reviews were identified. Halm *et al* (2002) reviewed studies published between January 1980 and December 2000 and Gandjour *et al* (2003) reviewed studies published between January 1990 and December 2000. Halm *et al* was a conventional systematic review whereas Gandjour *et al* took the unusual approach of identifying the single most reliable study (based on criteria such as the quality of case-mix adjustment) in each of 34 distinct clinical areas.

4.1 Methodological Quality

The York Review (Sowden *et al*, 1997) was rightly critical of the methodological quality of volume/outcome studies published prior to 1997. There is clear evidence that case-mix adjustment using clinical data on individual patients leads to the most reliable results and that case-mix adjustment using only administrative data tends to overestimate the magnitude of volume/outcome relationships. Studies which use no case-mix adjustment at all are likely to overestimate such effects even more so. This is now widely recognised and methodological standards are higher in recent studies. For example, in the Gandjour review (2003), 16 of the 'best' 33 hospital volume/outcome studies published between 1990 and 2000 were published in 1999 or 2000.

Increasingly (and appropriately!) sophisticated statistical approaches are now being used to try to disentangle the complex issue of whether it is surgeon volume or hospital volume which drives the observed volume/outcome relationships (Birkmeyer *et al* 2003, Panageas *et al* 2003).

Methodologically flawed studies are still widely published, but there is now a strong core of methodologically sound papers which use high quality data and appropriate statistical methods to explore volume/outcome relationships. These papers are based on either series of patients with data extracted from administrative systems (giving very large sample sizes but incomplete case-mix adjustment) or on series of patients with data extracted from clinical databases (giving smaller sample sizes but good case-mix adjustment). Even when one restricts attention to these higher quality studies there is still very strong evidence of an association between volume and outcome in the direction that high volume surgeons and high volume hospitals tend to have superior outcomes compared to low volume surgeons and hospitals. The magnitude of this effect, and how it depends on the clinical area, is discussed in Section 4.4.

4.2 Causation

In spite of there being very strong evidence of an association between increased volume and better outcomes there are remarkably few studies which try to assess whether this association is causal. Indeed, it was stated in the York review that there was no evidence that increasing the volume of activity in a given unit would lead to an improvement in outcomes. This reflected a lack of evidence rather than evidence of a lack of effect.

The evidence in this area is still extremely limited, but a number of studies are beginning to appear which evaluate the impact of interventions designed to concentrate activity.

Trauma systems is an area with a long history of regionalisation, and where different approaches adopted by different countries constitutes a 'natural experiment' on the organisation of trauma care. Nathens *et al* (2004) review the history of trauma management in the US and in France, and demonstrate how outcomes of trauma victims improved in the US following the introduction of regionalisation. However, there was a substantial lag period between the introduction of regionalisation and an observed improvement in outcome.

The UK Neonatal Staffing Study Group (2002) review the evidence for regionalisation of neonatal intensive care units. The situation is complex, but evidence of volume/outcome relationships from older studies is not seen currently, and this is ascribed to lower volume units adopting developments in treatment which were initially used only in high volume units. Training and staffing levels appear to be more important than volume *per se*. The study also raises the caveats that high volume units with a large number of consultant staff had higher levels of nosocomial bacteraemia, and that units running close to capacity have worse outcomes then when there is 'slack' in the system. These findings on staffing levels are consistent with the analysis of English hospital death rates published by Jarman *et al* (1999).

Nobilio *et al* (2004) report on the impact of regionalisation of cardiac surgery in an Italian region. They looked at patient outcomes, accessibility for patients and the efficiency of referral systems following the adoption of a "hub & spoke" model. The study does provide evidence of benefit, and the authors conclude that their findings suggest "*that policies aimed at increasing cooperation rather than competition among health service providers have a positive impact on quality of care*."

This latter finding is consistent with data from the Lothian Surgical Audit which was presented at the recent Consensus Conference of the Association of Surgeons of Great Britain and Ireland on "Modernising Medical Careers and General Surgery". (The consensus statement can be downloaded from www.asgbi.org.uk). In the Lothian's experience restructuring of emergency surgical care focussed on appropriate sub-specialisation has led to improved outcomes.

4.3 Thresholds

Ramsay *et al* (2001) undertook a systematic review of methods used to analyse learning curves in healthcare, and more recently, Cook *et al* (2004) proposed methods for adjusting for learning curve effects in randomised trials of surgical interventions.

I could not identify any consensus on appropriate ways to analyse learning effects, with one of the major problems being that as experience is gained in a new technique it tends to be deployed for higher risk patients, meaning that outcomes can deteriorate

as experience is gained. This requires careful case-mix adjustment to interpret correctly, but almost by definition there is insufficient data for such analyses early in the learning experience.

Great import is placed on volume thresholds by the Leapfrog Group, a large US-based consortium of healthcare purchasers (Birkmeyer *et al* 2004). However, my impression from the literature is that definitions of 'low volume' and 'high volume' relate more to potential volumes than to any objective evidence on the level of activity which is required to achieve and/or maintain competence. For example, a unit performing 100 carotid endarterectomies per year could be classified 'high volume' whereas a unit performing 400 coronary artery bypass graft procedures per year could be classified as 'low volume' (Gandjour *et al* 2003, Birkmeyer *et al* 2004).

Studies which present outcome data for a range of activity volumes, as opposed to a simple low volume/high volume dichotomy, do often report poor outcomes at low activity levels and then a levelling off with outcomes in moderate volume units being comparable to outcomes in high volume units. This is partly the result of a statistical artefact, with greater variation being observed in the small samples which derive from low volume units. However when this excess variation is taken into account there is still evidence of discrepant outcomes being observed in very low activity units (see, for example, the review by Shahian and Normand, 2003).

4.4 Examples of the Magnitude of Volume/Outcome Relationships

Halm *et al* (2002) summarise volume/outcome effects in terms of absolute differences in mortality between high and low volume hospitals and Gandjour *et al* (2003) present mortality rates for high volume relative to low volume hospitals. Absolute differences in mortality rates of the order of 10% are reported when high volume units are compared to low volume units in a number of complex high risk surgical procedures including paediatric cardiac surgery, surgery to repair ruptured abdominal aortic aneurysms, pancreatic cancer surgery and oesophageal cancer surgery. Relative differences in mortality rates of at least 10% are reported in a range of common lower risk procedures including percutaneous transluminal coronary angioplasty, carotid endarterectomy, knee replacement and surgery for hip fracture.

5. Discussion

The interpretation of the results of volume/outcome studies is highly complex. At the time of the York Review, methodological deficiencies in the evidence base meant that the studies had little if any relevance to health service planning. Recent improvements in the methodological rigour of at least a proportion of published volume/outcome studies mean that there is a substantial body of credible evidence which shows substantial effects in a limited range of complex high risk surgical procedures and modest but clinically relevant effects in a wide range of more general procedures.

The major current problem in interpreting these studies is the dearth of evidence supporting the hypothesis that the volume/outcome association is a causal association, whereby manipulating volume might have a beneficial impact on outcome. The problem here is a lack of evidence rather than clear evidence of a lack of a causal effect. There is some limited evidence accumulating to support the association as being causal, but a great deal more research is needed in this area.

Given this lack of evidence, it becomes a value judgement to explain the observed volume/outcome relationships. The two widely cited explanations are 'practice makes perfect' and 'selective referral' (i.e. patients are selectively referred to clinicians or hospitals that have historically achieved good outcomes). The former would suggest that volume/outcome associations are causal but the latter would imply that the observed associations are artefactual rather than causal. There are also issues such as aspects of process and/or structure which are associated with high volume (e.g. a large well staffed intensive care unit) and which might lead indirectly to better outcomes.

Another point to consider is whether it is the volume of activity for an individual surgeon or physician which is important or the volume of the relevant unit or hospital. Far more studies have looked at hospital volume than have looked at surgeon volume, but there is no consensus on which has the stronger association with outcome. A related point is whether the surgeon/hospital needs to be 'high volume' for the

procedure in question, or is high volume in general associated with good outcomes for all procedures? Urbach and Baxter (2004), for example, argue that volume in general is more important than volume for the specific procedure.

A final important issue is the fact that the volume/outcome literature looks at average effects. Although high volume is associated with good outcome in general, there are many low volume hospitals whose outcomes are superior to typical high volume hospitals and there are high volume surgeons with poor results who work within high volume hospitals.

6. Conclusions

Returning to the three conclusions extracted from the York Review in Section 2, I would say that the concerns over case-mix adjustment no longer hold. There is now a core of studies of adequate methodological quality to establish striking volume/outcome associations in certain complex high risk surgical procedures and more modest but clinically relevant effects in a wide range of common procedures.

There is only limited evidence to suggest that the observed associations are causal, and that interventions to manipulate volume can lead to better outcomes. It is important to note that the issue here is that evidence is sparse, rather than there being strong evidence of a lack of a causal association. The relevance of the observed volume/outcome relationships to health service planning depends crucially on how one interprets the underlying mechanisms which generate the associations.

My interpretation of the recent literature strengthens the final conclusion of the York Review, namely it does indeed appear that any benefits which would arise through manipulation of volume would be exhausted at a relatively low threshold level, which may be consistent with increased sub-specialisation within disciplines.

References

Begg CB, Riedel ER, Bach PB, et al. Variations in morbidity after radical prostatectomy. *N Eng. J Med* 2002; **346**: 1138-1144.

Birkmeyer JD and Dimick JB. Potential benefits of the new Leapfrog standars: effect of process and outcome measures. *Surgery* 2004; **135**: 569-575.

Birkmeyer JD, Stukel TA, Siewers AE, et al. Surgeon volume and operative mortality in the United States. *N Eng J Med* 2003; **349**: 2117-2127.

Cook JA, Ramsay CR and Fayers P. Statistical evaluation of learning curve effects in surgical trials. *Clinical Trials* 2004; **1**: 421-427.

Gandjour A, Bannenberg A and Lauterbach KW. Threshold volumes associated with higher survival in health care. A systematic review. *Med Care* 2003; **41**: 1129-1141.

Halm EA, Lee C and Chassin MR. Is volume related to outcome in health care? A systematic review and methodological critique of the literature. *Ann Intern Med* 2002; **137**: 511-520.

Jarman B, Gault S, Alves B, et al. Explaining differences in English hospital death rates using routinely collected data. *BMJ* 1999; **318**: 1515-1520.

Nathens AB, Brunet FP and Maier RV. Development of trauma systems and effect on outcomes after injury. *Lancet* 2004; **363**: 1794-1801.

Nobilia L, Fortuna D, Vizioli M, et al. Impact of regionalisation of cardiac surgery in Emilia-Romagna, Italy. *J Epidemiol Community Health* 2004; **58**: 97-102.

Panageas KS, Schrag D, Riedel E et al. The effect of clustering of outcomes on the association of procedure volume and surgical outcomes. *Ann Intern Med* 2003; **139**: 658-665.

Ramsay CR, Grant AM, Wallace SA, et al. Statistical assessment of the learning curves of health technologies. *Health Technology Assessment* 2001; **5** (12): 1-98.

Shahian DM and ST Normand. The volume-outcome relationship: from Luft to Leapfrog. *Ann Thorac Surg* 2003; **75**: 1048-1058.

Sowden AJ, Grilli R and Rice N. The relationship between hospital volume and quality of health outcomes. CRD report 8, part 1. York: Centre for Reviews and Dissemination, 1997.

The UK Neonatal Staffing Study Group. Patient volume, staffing, and workload in relation to risk-adjusted outcomes in a random stratified sample of UK neonatal intensive care units: a prospective evaluation. *Lancet* 2002; **359**: 99-107.

Urbach DR and Baxter NN. Does it matter what a hospital is "high volume" for? Specificity of hospital volume-outcome associations for surgical procedures: analysis of administrative data. *BMJ* 2004; **328**: 737-740.