

This is a repository copy of *Should interventions to reduce variation in care quality target doctors or hospitals?*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/129404/>

Version: Accepted Version

Article:

Gutacker, Nils orcid.org/0000-0002-2833-0621, Bloor, Karen Elizabeth orcid.org/0000-0003-4852-9854, Bojke, Christopher orcid.org/0000-0003-2601-0314 et al. (1 more author) (2018) *Should interventions to reduce variation in care quality target doctors or hospitals?* Health Policy. pp. 660-666. ISSN 1872-6054

<https://doi.org/10.1016/j.healthpol.2018.04.004>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Should interventions to reduce variation in care quality target doctors or hospitals?

Nils Gutacker, PhD
Centre for Health Economics
University of York
Heslington
York
YO10 5DD
United Kingdom
Corresponding author. Tel: +44 1904 321443. Fax: +44 1904 1402
Email: nils.gutacker@york.ac.uk

Karen Bloor, PhD
Department of Health Sciences
University of York
United Kingdom
Email: karen.bloor@york.ac.uk

Chris Bojke, PhD¹
Centre for Health Economics
University of York
United Kingdom
Email: c.bojke@leeds.ac.uk

Kieran Walshe, PhD
Alliance Manchester Business School
University of Manchester
United Kingdom
Email: kieran.walshe@manchester.ac.uk

Acknowledgements

This paper is an output from independent research commissioned and funded by the Department of Health Policy Research Programme (PR-R9-0114-11002 Evaluating the development of medical revalidation in England and its impact on organisational performance and medical practice). The views expressed in this publication are those of the author(s) and not necessarily those of the Department of Health.

Conflict of interest

The authors report no conflicts of interest.

¹ Present address: Leeds Institute of Health Sciences, University of Leeds, United Kingdom

Highlights

- Performance management initiatives are increasingly targeting individual doctors as well as hospitals [12 words]
- Less than 25% of variation in clinical outcomes is attributable to providers [12 words]
- More variation in clinical outcomes is associated with doctors than with hospitals [12 words]
- Performance estimates for individual doctors are unreliable due to small samples [11 words]

Should interventions to reduce variation in care quality target doctors or hospitals?

Abstract

Interventions to reduce variation in care quality are increasingly targeted at both individual doctors and the organisations in which they work. Concerns remain about the scope and consequences for such performance management, the relative contribution of individuals and organisations to observed variation, and whether performance can be measured reliably.

This study explores these issues in the context of the English National Health Service by analysing comprehensive administrative data for all patients treated for four clinical conditions (acute myocardial infarction, hip fracture, pneumonia, ischemic stroke) and two surgical procedures (coronary artery bypass, hip replacement) during April 2010 to February 2013. Performance indicators (PIs) are defined as 30-day mortality, 28-day emergency readmission and inpatient length of stay. Three-level hierarchical generalised linear mixed models are estimated to attribute variation in case-mix adjusted indicators to individual doctors and hospital organisations.

Except for length of stay after hip replacement, no more than 11% of variation in case-mix adjusted PIs can be attributed to doctors and organisations with the rest reflecting random chance and unobserved patient factors. Doctor variation exceeds hospital variation by a factor of 1.2 or more. However, identifying poor performance amongst doctors is hampered by insufficient numbers of cases per doctor to reliably estimate their individual performances. Policy makers and regulators should therefore be cautious when targeting individual doctors in performance improvement initiatives.

1 Introduction

Large variations in the quality of health care have been reported over many years, and in many countries [1, 2]. Policymakers and professional bodies have responded to such variations with a variety of mechanisms including measurement ('profiling'), monitoring, public reporting, regulation and incentives (financial and non-financial) [3, 4]. These interventions have mostly been focused on organisational performance, particularly at the level of the hospital or clinical specialty, with the implicit assumption that the variation results from factors that can be influenced or affected by organisations and those who lead them.

Increasingly, interventions to improve care quality and reduce variations operate not just at organisational level but at the level of individual doctors. For example, a number of initiatives have been introduced with the aim of improving hospital specialists' mortality rates through measurement, public reporting and feedback, most notably in cardiac surgery in the UK and US [5, 6]. In the National Health Service (NHS) in England, this has been extended to routine publication of outcome data for consultants (fully-trained hospital specialists) working in 13 specialities [7, 8].

Despite substantial investments in these mechanisms intended to drive improvements in the quality of care, considerable uncertainty exists about whether individual consultants or the organisations in

1 which they work are more important as drivers of variation in the quality of health care. The utility
2 of information derived from administrative data for individual or organisational performance
3 management purposes, and the potential for unintended consequences, remain poorly understood.
4 For example, there is only limited UK evidence on the degree of performance variation among
5 doctors for outcomes other than mortality [9]. In addition, the assessment of performance of
6 individual consultants raises a statistical concern: estimates of their performance are more
7 vulnerable to chance events than those of hospitals because they are based on smaller patient
8 populations [10]. A number of studies have suggested that using indicators at individual level may
9 result in often unreliable estimates of true performance [11-15]. Unreliable estimates may result in
10 incorrect decisions about doctors' performance with potentially adverse consequences for individual
11 careers, the welfare of patients, and the credibility of the measurement process.
12
13
14
15

16 This paper explores these issues in the context of the English NHS, extending a previous analysis of
17 mortality variation in England [14] and also focusing on two performance indicators (PIs) not
18 previously analysed: emergency re-admission within 28 days of discharge and inpatient length of
19 stay. The analysis seeks to answer two questions. First, how much variation in observed PIs can be
20 attributed to individual hospital consultants and how does this compare with that attributable to the
21 organisations in which they work? Second, are performance estimates for individual consultants
22 sufficiently reliable to be useful estimates of their true performance?
23
24
25
26

27 **2 Methods**

28 **2.1 Study population**

29 We used data from Hospital Episode Statistics (HES) on all NHS-funded inpatient care provided in
30 hospitals in England between April 2010 and February 2013. We focused on six
31 conditions/procedures that were selected because they are based on validated indicators used
32 internationally [16, 17], they cover a range of clinical areas and are either part of the consultant-
33 level reporting initiative in England [8] or constitute a substantial proportion of NHS activity:
34 emergency admissions for treatment of acute myocardial infarction (AMI), acute ischemic stroke
35 (AIS), pneumonia and hip fracture; and elective admissions for unilateral primary (i.e. non-revision)
36 hip replacement and isolated coronary artery bypass graft (CABG) surgery. These groups were
37 constructed following US Agency for Healthcare Research and Quality's inpatient quality indicator
38 (IQI) definitions (IQI#12, #14, #15, #17, #19, #20), which were recently amended for use with English
39 NHS data as part of a European study of health care variations [16]. A full list of relevant ICD-10
40 diagnosis codes and OPCS-4 procedure codes are reported in the Appendix. Patients were excluded
41 if they were younger than 18 years at the time of admission (<40 years for CABG surgery; <65 for hip
42 fracture) or were living outside of England.
43
44
45
46
47
48
49
50

51 HES records inpatient activity at the level of Finished Consultant Episodes (FCEs), which we linked to
52 create continuous inpatient spells that cover the entire period from admission to discharge
53 (including transfers between hospitals). Data were extracted on all inpatient activity 365 days before
54 index admission and 28 days after discharge (up to 31st March 2013). Record linkage was based on
55 unique NHS identification numbers. Admission spells were assigned to the first consultant
56 responsible for treatment after the index admission. Consultants who provided care in different
57 hospital organisations were treated as separate units of observation. This issue was most prevalent
58
59
60
61
62
63
64
65

1 in elective hip replacement surgery, where consultants often work both in NHS hospital trusts and
2 privately operated Independent Sector Treatment Centres. Consultants were identified through their
3 unique General Medical Council (GMC) code. These codes were validated against the GMC database
4 of registered specialists and the Electronic Staff Record system and invalid records were excluded
5 from analysis. Consultants (and their patients) were excluded if they treated less than 30 cases over
6 the three-year period. Similarly, hospitals were excluded if they treated less than 90 cases over this
7 period.
8
9

10 **2.2 Performance indicators**

11 We investigated variation in important clinical outcomes and process of care measures that are
12 commonly used as PIs. The clinical outcomes were 28-day all-cause emergency readmission and 30-
13 day all-cause mortality, which was derived from Office for National Statistics date of death data.
14 Length of inpatient stay, measured as the number of overnight stays, was used to approximate the
15 effectiveness of discharge management processes. To reduce the influence of potential miscoding
16 values exceeding the 99th percentile of the distribution of length of stay were replaced with the 99th
17 percentile.
18
19
20
21

22 **2.3 Case-mix adjustment**

23 All PIs were adjusted for age (5-year bands with separate categories for <25 and ≥85; except in the
24 analysis of mortality in which lowest category is <60), sex, age-sex interactions and year of
25 hospitalisation. Severity adjustment was limited to information contained in administrative records
26 and included an indicator for any hospital emergency admission in the previous year, as well as the
27 number of Elixhauser co-morbid conditions (grouped as 0, 1, 2-3, 4+) recorded in secondary
28 diagnosis fields in the index admission or admissions in the previous year. Patients' socio-economic
29 status was approximated by the proportion of residents at small area level (Lower Super Output
30 Area; approximately average population of 1,500 inhabitants) claiming means-tested social security
31 benefits (divided into five quintile groups) [18].
32
33
34
35
36

37 **2.4 Statistical analysis**

38 Three-level hierarchical generalised linear mixed models were fitted to identify variation in PIs due
39 to provider case-mix, additional systematic variation associated with consultants and hospital
40 organisations, and random chance variation [19, 20]. Patient episodes are nested in consultants,
41 which are themselves nested in hospitals. Emergency readmissions and mortality were modelled
42 using logistic regression. Length of stay was modelled as count data using Poisson models with an
43 additional over-dispersion parameter. Separate models were estimated for each patient group and
44 PI. Data were pooled across three years to reflect common practice in performance assessment
45 schemes [14].
46
47
48
49

50 The fixed part of the model captures variation in PIs associated with observable differences in
51 provider case-mix (see section 2.3). The model error term captures the variation in the PI that is not
52 explained by observed patient characteristics and is further partitioned into separate components
53 varying at patient level (i.e. unmeasured patient characteristics or random noise with variance σ^2),
54 consultant level (τ^2) and hospital level (ω^2). From that we calculated variance partition coefficients
55 (VPC) at the response scale by means of simulation [21, 22]. Each VPC measures the proportion of
56 unexplained variation in PIs associated with the respective level of the hierarchy. For example, the
57 VPC at consultant level is defined as
58
59
60
61
62
63
64
65

$$VPC_{consultant} = \frac{\tau^2}{\sigma^2 + \tau^2 + \omega^2}$$

and similarly for other levels. By design, all VPCs must sum to unity. Higher values of VPC therefore indicate a larger influence on PIs relative to other levels.

In most performance assessment schemes, case-mix adjusted performance estimates are obtained by means of indirect standardisation. There is a risk that any performance estimates for individual consultants conflate true variation across consultants with random noise. The reliability (R) of performance estimates for individual consultants is a function of their case-load N and the $VPC_{consultant}$. It is calculated as

$$R_{consultant} = \frac{N \times VPC}{1 + [N - 1] \times VPC}$$

with $0 < R \leq 1$ [10]. Higher values of R indicate that estimates of individual consultants' performance are less subject to unrelated variation and are thus more suitable for performance assessment purposes. Values of ≥ 0.7 are often required for low-stakes applications such as confidential reports to clinicians with limited risk of punitive actions [15]. Conversely, values of ≥ 0.9 have been suggested for high-stakes applications such as public reporting of performance or pay-for-performance schemes. The minimum level of activity required for a given level of reliability R can be obtained by solving the above equation for N . We calculated minimum activity thresholds and the proportion of consultants fulfilling these thresholds to achieve reliability of 0.7 and 0.9, respectively.

All statistical analyses were performed in Stata 13 (StataCorp LP, College Station, TX, USA) and MLwiN 2.36 (Centre for Multilevel Modelling, University of Bristol, UK).

3 Results

A total of 1,211,983 patients were included in the initial sample. Of these, 172,826 (14.3%) patients did not fulfil the inclusion criteria, leaving 1,039,157 patients for further analysis (Table 1). These patients received care from 7,197 consultants (6,731 unique GMC codes) in 240 hospitals. The number of patients per consultant varied substantially within and across conditions. The lowest case-load was observed for consultants treating AIS patients (median=55; IQR=38-150) and the highest consultant case-load was for CABG surgery (median=104; IQR=72-158).

[Table 1 here]

Patients in our sample were on average 73 years old and approximately half were male. The overall 28-day emergency readmission and 30-day mortality rates were 12.0% and 11.0% respectively, and patients stayed in hospital for 12.5 nights on average. There was marked variation in patient characteristics and PIs across conditions. Patients admitted for planned care were on average younger (68 vs. 75 years), stayed shorter in hospital (5.8 vs. 14.1 nights) and were at lower risk of readmission (6.2 % vs. 13.4%) and mortality (0.2% vs 13.5%).

3.1 Variation across hospitals and consultants

All coefficients on case-mix variables show the expected sign and internally consistent ranking of magnitudes. The McKelvey-Zavoina Pseudo R^2 statistics [10, 23, 24] measure the proportion of

1 variance in PIs explained by observed patient characteristics and range from 16.7% to 26.9% for
2 mortality, 2.7% to 4.4% for emergency readmission, and 5.7% to 22.8% for the number of inpatient
3 days. More detail on regression coefficients and explained variance are provided in the Appendix.

4
5 Our primary interest is in the proportion of variation not explained by case-mix and how this relates
6 to consultants and hospital organisations. Figure 1 shows the estimated VPCs at consultant and
7 hospital level (stacked) for each of the three PIs and by condition. Approximately 0.6% to 4.1% of
8 unexplained variation in the case-mix adjusted probability of readmission can be attributed to
9 hospitals and consultants. The remainder reflects random variation at patient level that is not
10 associated with observed patient characteristics. VPCs for mortality are of similar magnitude and
11 range from 0.3% to 2.0%. Conversely, hospitals and consultants have a relatively larger influence
12 over patients' length of stay. Between 1.9% and 22.6% of unexplained variation in length of stay is
13 associated with either consultant or hospital. Note that the noticeably larger variation in length of
14 stay after planned hip replacement surgery may reflect differences in the performance of public and
15 private hospitals [25]; with hip replacement being the only condition studied for which this
16 distinction is relevant.

17
18 The proportion of unexplained variation at consultant level exceeds that at hospital level by a factor
19 of 1.2 or more, except for emergency readmission after AMI. It was not possible to differentiate
20 consultant and hospital variation for mortality after planned hip replacement surgery as part of the
21 estimation procedure, and the presented number should therefore be interpreted as a composite.

22
23 [Figure 1 about here]

30 31 **3.2 Reliability of consultant and hospital performance estimates**

32 Table 2 shows the reliability of consultant performance estimates for the three PIs, the level of
33 activity required to achieve reliability of at least 0.7 and 0.9, and the proportion of consultants that
34 fulfil this requirement. The reliability of consultants' emergency readmission rates as indicators of
35 their performance ranges from 0.19 to 0.71. The required 3-year activity to achieve a reliability of
36 ≥ 0.7 lies between 92 to 563 admissions for the six conditions studied. Very few consultants achieve
37 such case-loads. By extension, even fewer consultants reach case-loads required for a reliability of
38 ≥ 0.9 . A noteworthy exception is hip replacement surgery, where more than half of consultants treat
39 a sufficient number of patients to obtain reliable estimates at $r \geq 0.7$.

40
41 Estimates of reliability and required case-load for 30-day mortality follow the same pattern.

42
43 [Table 2 here]

44
45 The reliability of consultant performance estimates for length of stay is significantly higher. At
46 median activity level, the reliability is estimated to range from 0.46 to 0.93. For each of the six
47 conditions studied, at least 25% of consultants treat enough patients to achieve a reliability of at
48 least 0.7. In some cases, such as cardiac surgeons performing CABG surgery, this is true for more
49 than 90% of consultants. Between 0.4% and 70% of consultants achieve a reliability of 0.9 or more.

50
51 Table 3 reports the same information for hospital performance estimates. As hospital organisations
52 are usually held accountable for all variation that is not attributable to case-mix and random noise,
53 including variation that derives from consultants working for them, the reported estimates are
54
55
56
57
58
59
60
61
62
63
64
65

1 based on the pooled VPC, calculated as $VPC_{Consultant} + VPC_{Hospital}$. Unsurprisingly, performance
2 estimates at hospital level are significantly more reliable than those calculated for consultants due to
3 the substantially larger case-loads and the increased VPC. The reliability of performance estimates
4 for a hospital of median activity levels exceeded 0.85 for all indicators and conditions. A large share
5 of hospitals fulfils volume requirements to achieve reliability of 0.9, ranging from 29% of hospitals
6 for emergency readmission after AMI to 100% for length of stay after bypass surgery.
7
8
9

10 4 Discussion

11 This study demonstrated that for the performance indicators and conditions chosen, the amount of
12 case-mix adjusted variation that is attributable to consultants generally exceeds that which is
13 attributable to organisations, although both are substantially outweighed by random variation at
14 patient level that is not explained by the observed patient characteristics. In addition, we found that
15 a large proportion of consultants do not treat a sufficient volume of patients for performance
16 estimates based on these measures to represent reliably their underlying performance.
17
18
19
20

21 Commentators have considered the estimated proportion of variability in performance indicators at
22 levels higher than patients (including physicians, groups and organisations) as low or even trivial and
23 have raised concerns about the purpose of performance management [26]. However, we wish to
24 stress that such judgements must consider not only the amount of unwarranted variation but also
25 the value of the performance indicators and the direct and indirect costs of initiatives aimed at
26 eradicating it [27]. For example, assuming an average cost of an emergency readmission in the
27 English NHS of £2,100 [28], we estimate the overall value of improving consultant performance to
28 match that of the current average for our sample alone to be approximately £8.4 million. This
29 ignores any patient health benefits associated with a reduced risk of readmissions. The organisations
30 in which consultants work also play a role in determining outcomes, albeit less than consultants.
31 Hence, the possible benefit of reducing unwarranted variation between consultants and/or
32 organisations is unlikely to be negligible, although this does not necessarily imply that any such
33 effort is a cost-effective use of resources.
34
35
36
37
38
39

40 As the amount of case-mix adjusted variation between consultants generally exceeds that which
41 occurs between organisations, a focus on individual doctors' performance may be thought justified.
42 In practice, however, there are obstacles to realising the potential benefit of consultant-level
43 performance information. In particular, efforts to identify poorly performing consultants using
44 outcome measures such as readmission and mortality derived from routine data are likely to
45 encounter measurement problems: a large proportion of consultants do not treat a sufficient
46 number of patients over a three-year period for these performance estimates to be reliable
47 representations of their individual underlying performances. There are several ways in which the
48 reliability of individual performance estimates can be improved, although each comes with their
49 own problems. Firstly, most consultants provide a variety of treatments for different patient groups
50 and this can be exploited to generate more comprehensive performance profiles on larger, and thus
51 more reliable, patient samples [29]. This, however, requires a more complex case-mix adjustment
52 strategy and may hide differential performances among the components of the composite for
53 individual consultants [30]. An alternative approach is to employ *shrinkage estimators*, which take
54 into account reliability to generate estimates that are less subject to random variation and
55 regression-to-the-mean [31, 32]. This means, however, that resulting estimates of consultant
56
57
58
59
60
61
62
63
64
65

1 performance are overly conservative and biased towards the average [33]. The implication is that
2 poorly performing consultants with smaller caseloads would be less likely to be identified correctly
3 as negative outliers.

4
5 These results suggest that policymakers seeking to manage performance and reduce unwarranted
6 variation pursue the right target but do so by the wrong means. While the variation across
7 consultants *overall* is larger than between hospitals, the performance of an *individual* consultant is
8 difficult to establish reliably. This suggests that performance management approaches seeking to
9 leverage routinely collected data on individual consultants' performances risk generating a non-
10 trivial number of false positive warnings, which may undermine trust in the validity and fairness of
11 the assessment. Until methods to increase the reliability of individual consultants' performance
12 estimates have been agreed and implemented, approaches to performance management may be
13 best aimed at the entire population of consultants (e.g. through enhanced professional regulation)
14 rather than a subset identified by unreliable means.
15
16
17
18

19 There are a number of limitations to our study. First, in line with current UK health policy we have
20 chosen consultants (fully trained hospital specialists) as the unit of analysis. However, consultants
21 generally lead teams of healthcare professionals and we cannot observe the actions taken by each
22 individual. It may therefore not be the consultant that had a measurable effect on outcomes;
23 although some may argue that, as leaders of these teams, they remain ultimately responsible.
24
25 Second, as in all observational studies, our results may be subject to unobserved confounding. Most
26 importantly, length of stay and emergency readmission may be determined by local supply factors,
27 such as the availability of primary care services or care home places. This may explain some of the
28 variation observed across hospitals but is unlikely to explain variation between consultants within
29 the same hospital. Thus, consultant-level variance partition coefficients and the reliability of
30 individual performance estimates may be underestimated. Similarly, performance estimates may be
31 biased by unobserved differences in case-mix. If, for example, more severely ill patients are more
32 likely to seek treatment from providers offering reportedly better services then the estimated
33 variation in performance would be biased downwards. This is clearly of less concern for emergency
34 care where patients have limited ability to choose and so may affect estimates differently across
35 conditions. Third, variation among healthcare providers in dichotomous outcomes (mortality,
36 readmission) may be more difficult to estimate than in continuous outcomes (length of stay) for a
37 given sample size. Since the probability of mortality and readmission, rather than the actual event,
38 can never be observed, this constitutes an inherent limitation of these metrics. Fourth, we have
39 focussed on a number of high-volume procedures and conditions that form part of performance
40 assessment initiatives in England or elsewhere and for which validated performance indicators exist.
41 But these conditions necessarily capture only a subset of all inpatient activity in English hospitals and
42 it is, therefore, unclear in how far our results can be generalised to other patient populations.
43
44 Finally, while our analysis provides estimates of the degree of variation in patient outcomes and
45 length of inpatient stay that is associated with consultants and hospitals it was not designed to
46 identify the influences and decisions that result in this variation. For example, some of the observed
47 variation at hospital level may be due to differences in infrastructure, which may be difficult to
48 resolve in the short run or is outside the control of the organisation entirely.
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

5 Conclusions

1 Policy makers, healthcare regulators and professional bodies in the UK and elsewhere are
2 increasingly targeting both organisations and individual hospital consultants through a variety of
3 performance management schemes and mechanisms. Our study shows that consultants vary in
4 terms of their clinical outcomes and resource utilisation, and that in general the proportion of
5 unexplained variation at consultant level exceeds that at hospital level. However, both consultant
6 and hospital factors explain only a small fraction of the variation in risk-adjusted patient outcomes
7 and process measures (length of stay, mortality and readmissions) compared with unmeasured
8 patient characteristics and random noise, which seems to suggest that the potential impact of these
9 performance management schemes aimed at organisations, individual consultants or both is likely to
10 be relatively limited. In addition, relatively small patient samples per consultant make it difficult to
11 form reliable judgements about consultants' individual performance, and suggest that producing
12 and publishing such comparisons may be at best uninformative and at worst misleading.
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Appendix 1: Sample definition – diagnosis and procedure codes

Sample	ICD-10 diagnosis codes	OPCS4 procedure codes
AMI	I21 - I22	-
		K40 - K46
		<i>Exclusion: K35 - K38, K49 -</i>
CABG	-	K50
Hip fracture	S72.0 - S72.2, S72.9	-
		W37 - W39, W46 - W48, W52
		- W54, W58.1, W93 - W95
		<i>Exclusion: Z94.1 + Z94.2,</i>
		<i>Z94.3 or any code indicating</i>
Hip replacement	-	<i>revision surgery</i>
Stroke	H34.1, I63 - I64	-
Pneumonia	J12 - J18	-

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Appendix 5: Proportion of variance explained by observed patient characteristics (Pseudo-R²)

Condition	Pseudo-R ²		
	28-day emergency readmission	30-day mortality	Length of stay
AMI	21.0%	4.4%	11.7%
CABG	26.9%	3.3%	13.7%
Hip fracture	18.5%	3.5%	12.7%
Hip replacement	26.3%	4.3%	22.8%
Pneumonia	16.7%	4.4%	8.0%
Stroke	17.1%	2.7%	5.7%

Notes: McKelvey-Zavoina Pseudo-R² statistics is defined at the latent scale and is calculated as the ratio of the variance of the linear predictor to the sum of all variance components and the variance of the linear predictor.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

6 References

1. Wennberg, J. and A. Gittelsohn, *Small Area Variations in Health Care Delivery: A population-based health information system can guide planning and regulatory decision-making*. Science, 1973. **182**(4117): p. 1102-1108.
2. Corallo, A., et al., *A systematic review of medical practice variation in OECD countries*. Health Policy, 2014. **114**(1): p. 5-14.
3. Smith, P.C., *Performance Measurement in Health Care: History, Challenges and Prospects*. Public Money & Management, 2005. **25**(4): p. 213-220.
4. Oliver, A., *Incentivising improvements in health care delivery*. Health Economics, Policy and Law, 2015. **2015**(10): p. 327-343.
5. Bridgewater, B., et al., *Has the publication of cardiac surgery outcome data been associated with changes in practice in northwest England: an analysis of 25,730 patients undergoing CABG surgery under 30 surgeons over eight years*. Heart, 2007. **93**(6): p. 744-8.
6. Hannan, E.L., et al., *The New York State Cardiac Registries: History, Contributions, Limitations, and Lessons for Future Efforts to Assess and Publicly Report Healthcare Outcomes*. Journal of the American College of Cardiology, 2012. **59**(25): p. 2309-2316.
7. NHS England, *Major breakthrough in NHS Transparency as consultant mortality data goes online for first time*. 2013.
8. HQIP. *Clinical Outcomes Publication*. 02/10/2017]; Available from: <http://www.hqip.org.uk/national-programmes/clinical-outcomes-publication/>.
9. Varagunam, M., A. Hutchings, and N. Black, *Relationship Between Patient-reported Outcomes of Elective Surgery and Hospital and Consultant Volume*. Medical Care, 2015. **53**(4): p. 310-316.
10. Snijders, T.A.B. and R.J. Bosker, *Multilevel Analysis - An Introduction to Basic and Advanced Multilevel Modeling*. 2 ed. 2012, Los Angeles: Sage.
11. Hofer, T.P., et al., *The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease*. JAMA, 1999. **281**(22): p. 2098-2105.
12. Dimick, J.B., D.O. Staiger, and J.D. Birkmeyer, *Ranking Hospitals on Surgical Mortality: The Importance of Reliability Adjustment*. Health Services Research, 2010. **45**(6p1): p. 1614-1629.
13. Adams, J.L., et al., *Physician Cost Profiling — Reliability and Risk of Misclassification*. New England Journal of Medicine, 2010. **362**(11): p. 1014-1021.
14. Walker, K., et al., *Public reporting of surgeon outcomes: low numbers of procedures lead to false complacency*. The Lancet, 2013. **382**(9905): p. 1674-1677.
15. Eijkenaar, F. and R.C.J.A. van Vliet, *Profiling Individual Physicians Using Administrative Data From a Single Insurer: Variance Components, Reliability, and Implications for Performance Improvement Efforts*. Medical Care, 2013. **51**(8): p. 731-739.
16. Bernal-Delgado, E., et al., *ECHO: health-care performance assessment in several European health systems*. European Journal of Public Health, 2015. **25**(Supplement 1): p. 3-7.
17. Agency for Healthcare Research and Quality. *Inpatient Quality Indicators Technical Specifications Updates - Version v7.0 (ICD 10), September 2017*. 2017 13/10/2017]; Available from: http://www.qualityindicators.ahrq.gov/Modules/IQI_TechSpec_ICD10_v70.aspx.
18. McLennan, D., et al., *The English Indices of Deprivation 2010*, Department for Communities and Local Government, Editor. 2011: London.
19. McCulloch, C.E., *Generalized linear mixed models*. 2003, Beachwood, Ohio: Institute of Mathematical Statistics.
20. Rodriguez, G., *Multilevel Generalized Linear Models*, in *Handbook of Multilevel Analysis*, J. de Leeuw and E. Meijer, Editors. 2008, Springer: New York.
21. Goldstein, H., W. Browne, and J. Rasbash, *Partitioning variation in multilevel models*. Understanding Statistics, 2002. **1**(4): p. 223-231.

22. Browne, W.J., et al., *Variance partitioning in multilevel logistic models that exhibit overdispersion*. Journal of the Royal Statistical Society: Series A (Statistics in Society), 2005. **168**(3): p. 599-613.
23. McKelvey, R.D. and W. Zavoina, *A statistical model for the analysis of ordinal level dependent variables*. The Journal of Mathematical Sociology, 1975. **4**(1): p. 103-120.
24. Nakagawa, S. and H. Schielzeth, *A general and simple method for obtaining R2 from generalized linear mixed-effects models*. Methods in Ecology and Evolution, 2013. **4**(2): p. 133-142.
25. Siciliani, L., P. Sivey, and A. Street, *Differences in length of stay for hip replacement between public hospitals, specialised treatment centres and private providers: selection or efficiency?* Health Economics, 2013. **22**(2): p. 234-242.
26. Fung, V., et al., *Meaningful Variation in Performance: A Systematic Literature Review*. Medical Care, 2010. **48**(2): p. 140-148.
27. Meacock, R., S.R. Kristensen, and M. Sutton, *The cost-effectiveness of using financial incentives to improve provider quality: a framework and application*. Health Economics, 2014. **23**(1): p. 1-13.
28. Billings, J., et al., *Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30)*. BMJ Open, 2012. **2**(4).
29. Smith, K.A., et al., *Improving the Reliability of Physician "Report Cards"*. Medical Care, 2013. **51**(3): p. 266-274.
30. Gutacker, N., et al., *Hospital Variation in Patient-Reported Outcomes at the Level of EQ-5D Dimensions: Evidence from England*. Medical Decision Making, 2013. **33**(6): p. 804-818.
31. Efron, B. and C. Morris, *Stein's Estimation Rule and Its Competitors--An Empirical Bayes Approach*. Journal of the American Statistical Association, 1973. **68**(341): p. 117-130.
32. Skrondal, A. and S. Rabe-Hesketh, *Prediction in multilevel generalized linear models*. Journal of the Royal Statistical Society. Series A (Statistics in Society), 2009. **172**(3): p. 659-687.
33. Austin, P.C., D.A. Alter, and J.V. Tu, *The Use of Fixed-and Random-Effects Models for Classifying Hospitals as Mortality Outliers: A Monte Carlo Assessment*. Medical Decision Making, 2003. **23**(6): p. 526-539.

Figures and tables

Figure 1: Proportion of variation attributable to consultants and hospitals; case-mix adjusted

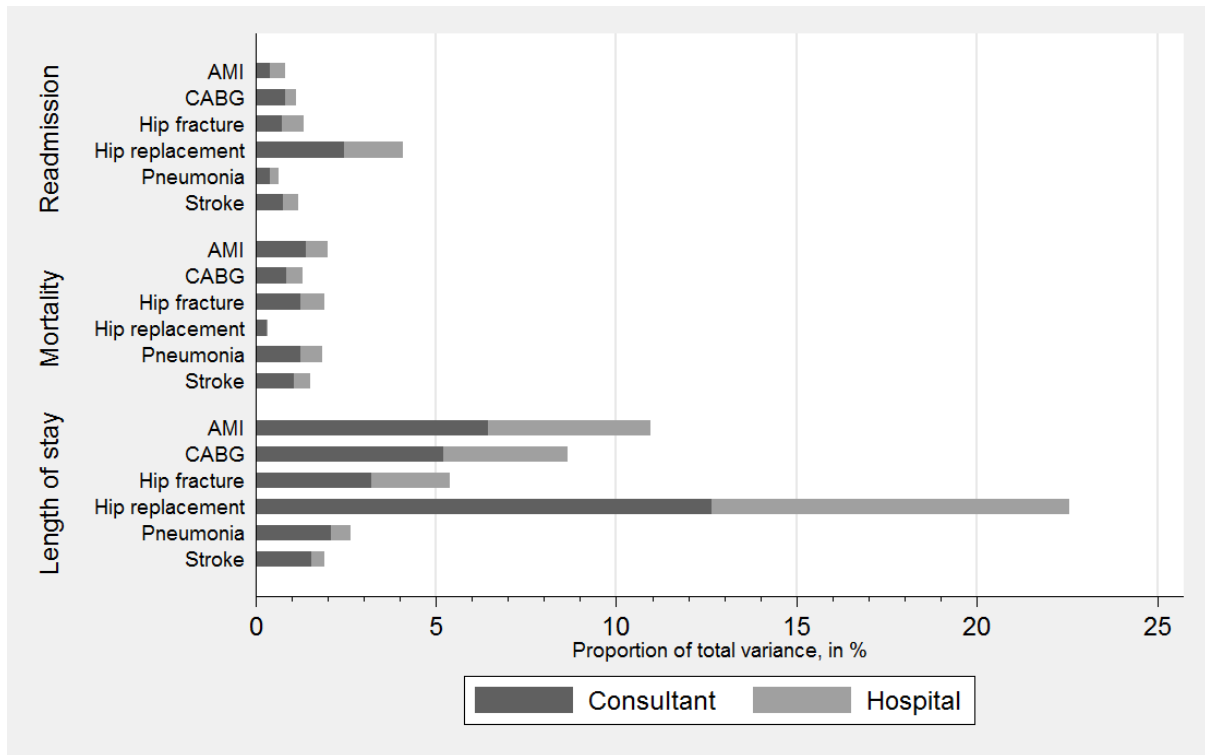


Table 1: Descriptive statistics of patient sample (April 2010 to February 2013)

	AMI		CABG		Hip fracture		Hip replacement		Pneumonia		Stroke		Total	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
<i>Patient level</i>														
28-day emergency readmission (yes/no)	0.15	0.36	0.12	0.32	0.12	0.32	0.05	0.22	0.15	0.35	0.10	0.30	0.12	0.33
30-day mortality (yes/no)	0.07	0.26	0.01	0.09	0.06	0.23	0.00	0.03	0.19	0.40	0.11	0.32	0.11	0.31
Length of stay (in days)	7.74	9.09	9.08	6.67	23.51	21.34	5.32	3.99	10.62	13.07	19.65	26.17	12.52	16.83
Patient age (in years)	69.75	14.14	66.07	9.35	81.00	11.42	67.96	11.51	73.80	16.63	75.37	13.20	73.42	14.81
Male (yes/no)	0.65	0.48	0.83	0.37	0.27	0.44	0.40	0.49	0.51	0.50	0.49	0.50	0.48	0.50
Elixhauser: 0 comorbidities	0.15	0.36	0.06	0.23	0.15	0.36	0.32	0.47	0.10	0.30	0.11	0.31	0.15	0.36
Elixhauser: 1 comorbidity	0.21	0.41	0.11	0.32	0.22	0.41	0.28	0.45	0.14	0.34	0.20	0.40	0.19	0.39
Elixhauser: 2-3 comorbidities	0.33	0.47	0.35	0.48	0.35	0.48	0.29	0.45	0.29	0.45	0.37	0.48	0.32	0.46
Elixhauser: 4+ comorbidities	0.31	0.46	0.48	0.50	0.28	0.45	0.11	0.31	0.48	0.50	0.32	0.47	0.34	0.47
Emergency admission in last year (yes/no)	0.26	0.44	0.38	0.49	0.33	0.47	0.09	0.29	0.51	0.50	0.31	0.46	0.35	0.48
Number of patients	138,044		24,505		156,145		170,678		405,671		144,114		1,039,157	
<i>Consultant level</i>														
Number of consultants	1,746		212		1,735		1,325		3,760		1,214		9,992	
Case-load: Median	56		104		86		95		83		55		78	
Case-load: 25th percentile	39		72		60		56		52		38		47	
Case-load: 75th percentile	94		158		112		167		131		149		125	
<i>Hospital level</i>														
Number of hospitals	148		30		148		229		152		144		851	
Case-load: Median	787		734		1000		649		2471		946		946	
Case-load: 25th percentile	505		616		705.5		224		1794		632.5		570	
Case-load: 75th percentile	1214.5		953		1337.5		985		3350		1348.5		1571	

Table 2: Reliability of consultant performance estimates

Condition	Estimated variance components				Case-load (median)	R	Case-load required		% Consultants with sufficient case-load over 35 months	
	σ^2	τ^2	ω^2	VPC			R=0.7	R=0.9	R=0.7	R=0.9
28-day emergency readmission										
AMI	0.23955	0.00102	0.00095	0.4%	56	0.19	552	2131	0.0%	0.0%
CABG	0.21058	0.00178	0.00065	0.8%	104	0.47	277	1068	0.5%	0.0%
Hip fracture	0.23855	0.00180	0.00144	0.7%	86	0.39	312	1203	0.3%	0.1%
Hip replacement	0.20456	0.00527	0.00348	2.5%	95	0.71	92	355	51.0%	3.8%
Pneumonia	0.24823	0.00103	0.00060	0.4%	83	0.26	563	2171	0.7%	0.0%
Stroke	0.24581	0.00188	0.00107	0.8%	55	0.29	307	1183	8.9%	0.0%
30-day mortality										
AMI	0.143812	0.002077	0.000857	1.4%	56	0.45	163	627	10.0%	0.0%
CABG	0.065208	0.000578	0.000289	0.9%	104	0.48	264	1020	0.5%	0.0%
Hip fracture	0.186773	0.002375	0.001295	1.2%	86	0.52	185	713	2.1%	0.1%
Pneumonia	0.204589	0.002588	0.001311	1.2%	83	0.51	186	716	10.5%	0.4%
Stroke	0.174487	0.001901	0.000806	1.1%	55	0.37	215	830	16.6%	0.2%
Length of stay										
AMI	118.748	8.623	5.987	6.5%	56	0.79	34	130	89.6%	14.5%
CABG	64.090	3.648	2.417	5.2%	104	0.85	43	164	93.4%	19.3%
Hip fracture	2270.457	77.193	51.934	3.2%	86	0.74	70	271	65.6%	0.4%
Hip replacement	30.123	4.921	3.855	12.7%	95	0.93	16	62	100.0%	70.0%
Pneumonia	2354.658	50.766	13.211	2.1%	83	0.64	109	420	35.1%	1.5%
Stroke	28799.816	453.625	105.254	1.5%	55	0.46	149	573	25.0%	1.6%

Notes: R = Reliability; VPC = Variance partition coefficient at consultant level. Median case-load is measured over the period April 2010 to February 2013. Variation in mortality after hip replacement at consultant level could not be differentiated from that at hospital level and the corresponding statistics are therefore not recorded.

Table 3: Reliability of hospital performance estimates

Condition	Estimated variance components				Case-load (median)	R	Case-load required		% Consultants with sufficient case-load over 35 months	
	σ^2	τ^2	ω^2	VPC*			R=0.7	R=0.9	R=0.7	R=0.9
28-day emergency readmission										
AMI	0.23955	0.00102	0.00095	0.8%	787	0.87	284	1095	89.9%	29.1%
CABG	0.21058	0.00178	0.00065	1.1%	734	0.89	202	779	100.0%	46.7%
Hip fracture	0.23855	0.00180	0.00144	1.3%	1000	0.93	172	664	96.6%	79.1%
Hip replacement	0.20456	0.00527	0.00348	4.1%	649	0.97	55	210	97.8%	77.3%
Pneumonia	0.24823	0.00103	0.00060	0.7%	2471	0.94	356	1371	98.7%	86.2%
Stroke	0.24581	0.00188	0.00107	1.2%	946	0.92	194	750	91.7%	63.9%
30-day mortality										
AMI	0.143812	0.002077	0.000857	2.0%	787	0.94	114	441	94.6%	80.4%
CABG	0.065208	0.000578	0.000289	1.3%	734	0.91	176	677	100.0%	56.7%
Hip fracture	0.186773	0.002375	0.001295	1.9%	1000	0.95	119	458	98.6%	87.8%
Pneumonia	0.204589	0.002588	0.001311	1.9%	2471	0.98	122	472	99.3%	97.4%
Stroke	0.174487	0.001901	0.000806	1.5%	946	0.94	150	580	92.4%	77.8%
Length of stay										
AMI	118.748	8.623	5.987	11.0%	787	0.99	19	73	100.0%	95.3%
CABG	64.090	3.648	2.417	8.6%	734	0.99	25	95	100.0%	100.0%
Hip fracture	2270.457	77.193	51.934	5.4%	1000	0.98	41	158	99.3%	96.6%
Hip replacement	30.123	4.921	3.855	22.6%	649	0.99	8	31	100.0%	99.6%
Pneumonia	2354.658	50.766	13.211	2.6%	2471	0.99	86	331	100.0%	98.7%
Stroke	28799.816	453.625	105.254	1.9%	946	0.95	120	464	93.1%	80.6%

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Notes: R = Reliability; VPC = Sum of variance partition coefficients at consultant and hospital levels. Median case-load is measured over the period April 2010 to February 2013. Variation in mortality after hip replacement at consultant level could not be differentiated from that at hospital level and the corresponding statistics are therefore not recorded.*