

Keywords: cancer diagnosis; primary care; referral; healthcare quality assurance

General practice performance in referral for suspected cancer: influence of number of cases and case-mix on publicly reported data

P Murchie^{*1}, A Chowdhury¹, S Smith¹, N C Campbell¹, A J Lee¹, D Linden² and C D Burton¹

¹Division of Applied Health Sciences, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, UK and

²Detect Cancer Early Programme, The Scottish Government, St Andrews House, Regent Road, Edinburgh EH1 3DG, UK

Background: Publicly available data show variation in GPs' use of urgent suspected cancer (USC) referral pathways. We investigated whether this could be due to small numbers of cancer cases and random case-mix, rather than due to true variation in performance.

Methods: We analysed individual GP practice USC referral detection rates (proportion of the practice's cancer cases that are detected via USC) and conversion rates (proportion of the practice's USC referrals that prove to be cancer) in routinely collected data from GP practices in all of England (over 4 years) and northeast Scotland (over 7 years). We explored the effect of pooling data. We then modelled the effects of adding random case-mix to practice variation.

Results: Correlations between practice detection rate and conversion rate became less positive when data were aggregated over several years. Adding random case-mix to between-practice variation indicated that the median proportion of poorly performing practices correctly identified after 25 cancer cases were examined was 20% (IQR 17 to 24) and after 100 cases was 44% (IQR 40 to 47).

Conclusions: Much apparent variation in GPs' use of suspected cancer referral pathways can be attributed to random case-mix. The methods currently used to assess the quality of GP-suspected cancer referral performance, and to compare individual practices, are misleading. These should no longer be used, and more appropriate and robust methods should be developed.

Early detection and treatment of cancer is an important goal for health services. The United Kingdom (UK) and other countries with strong primary care 'gatekeeper' systems persistently display lower cancer survival rates when compared with other developed countries. This effect is widely attributed to longer intervals in the cancer diagnostic and treatment pathway (Richards, 2009; Coleman *et al*, 2011).

The UK NHS has existing fast-track urgent suspected cancer (USC) referral pathways from primary to secondary care (NICE, 2005; Scottish Government, 2009). In England, patients referred from primary care with suspected cancer should be seen in secondary care within 2 weeks (hence the common term '2-week wait referrals'; Meechan *et al*, 2012; Department of Health, 2000) and begin treatment within 62 days (England and Scotland;

Department of Health 2000; Scottish Executive Health Department, 2007).

There is evidence that practices vary in the frequency with which they make USC referrals. In Scotland, an analysis of 18 775 USC referrals in 2008 reported a six-fold variation in the use of this referral route (Baughan *et al*, 2011). Similar variation was shown in an analysis of 865 494 referrals in England over 1 year (Meechan *et al*, 2012). This latter analysis also found a positive correlation between practices' detection rate (the proportion of all cancers referred as USC) and conversion rate (the proportion of all USC referrals resulting in a cancer diagnosis; Supplementary Appendix 1). This correlation has been used as evidence of a quality gradient: with 'high-quality' practices being both more accurate (higher detection rate) and more efficient (higher conversion rate) in their

*Correspondence: Dr P Murchie; E-mail: p.murchie@abdn.ac.uk

Received 24 September 2014; revised 16 February 2015; accepted 25 February 2015; published online 16 April 2015

© 2015 Cancer Research UK. All rights reserved 0007–0920/15

use of USC than others (Meechan *et al*, 2012). In England, the National Cancer Intelligence Network now publishes GP practice profiles for cancer that include annual detection and conversion rates (National Cancer Intelligence Network, 2012). This has led to media reports of unacceptable variations in GP performance culminating in the recent recommendation by the UK Health Secretary that 'poorly performing' practices be publicly named. (BBC Website, 2013; BBC Website, 2014).

Current reporting is based on referral data from a single year and does not distinguish between different cancer types. New cases of cancer are relatively uncommon in primary care, and therefore the number of cases in any year will be small. Furthermore, cancer in primary care is heterogeneous: some cancers typically present with features amenable to prompt recognition and referral (e.g., testicular cancer presenting as a lump), whereas others typically have nonspecific symptoms (e.g., ovarian cancer; Bottle *et al*, 2012). This is reflected in substantial differences in the use of the USC pathway according to cancer type (National Cancer Intelligence Network, 2014). Even within the same cancer type, some presentations will prompt urgent referral (e.g., lung cancer presenting with haemoptysis), whereas others may not (e.g., lung cancer presenting with nonspecific symptoms; Birring and Peake, 2005). Furthermore, presymptomatic cancers that had been detected via national screening programmes are currently counted in the non-USC category, introducing a further source of variation in apparent GP performance.

Current national guidelines dictate which circumstances warrant USC referral; thus, adherence to these will inevitably influence which referral route GPs choose. Thus, depending on case-mix, two practices following guidelines equally well may have different detection and conversion rates with the appearance, based on current metrics, that one is better than the other (Dua *et al*, 2009).

The aim of our study was to investigate the effect of the number of cancer cases and random case-mix on the variation in GP performance in cancer diagnosis and their implication for public reporting. First, we examined the effect of aggregating data for each practice over several years on the proposed 'quality gradient' indicated by the association between detection and conversion rates. Second, we examined year-to-year correlation in detection and conversion rates in order to assess whether practices were consistent over time in their reported rates. Finally, we carried out a simulation modelling study to estimate the variation in USC rates attributable to random case-mix and used this to estimate the likelihood that a poorly performing practice would be correctly identified.

MATERIALS AND METHODS

Data sources. The current study used two databases. The first provided data on the route of cancer diagnosis, including USC referral from primary care over 7 years (2006–2012) in the NHS Grampian region of northeast Scotland. The second contained data on the route of cancer diagnosis including 2-week referrals over 4 years (2010–2013) from NHS England. USC in Scotland and 2-week referrals in England are the broadly equivalent referral routes by which GPs in Scotland and England, respectively, secure an urgent secondary care appointment for patients in whom a strong suspicion of cancer is supported by existing guidelines.

The data from northeast Scotland comprised practice-level data for all GP practices in the NHS Grampian region relating to all cancer diagnoses and all USC referrals made between 2 January 2006 and 30 November 2012. These were obtained from the NHS Grampian Cancer Care Pathway database (CCPd). The CCPd is a detailed clinical database maintained by NHS Grampian, recording information about all cancer referrals made by GPs within the region, as well as information about all cancer diagnoses,

irrespective of route of diagnosis. The start date for these data represents the earliest date of collection of USC referral data. The NHS England data comprised publicly available practice-level data for all GP practices on the NCIN website (National Cancer Intelligence Network, 2014) for the years 2010–2013 inclusive.

Data processing and analysis. With the NHS Grampian data, for each practice and year, we extracted the total number of cancers, the number of cancers detected after USC referral and the total number of USC referrals. We also used demographic data from each practice to calculate age–sex-standardised referral ratios based on the number of USC referrals that would have been expected from the practice relative to other practices in NHS Grampian. With the NHS England data, for each practice and year, we extracted the total number of cancers treated, the number of cancers detected after USC referral and the total number of USC referrals. We also extracted the age–sex-standardised referral rate.

With both extracted data sets, we calculated the detection and conversion rates for each practice for each year and aggregated across years. These are analogous to the sensitivity and positive predictive value, respectively, of a diagnostic test: in this case, the 'diagnostic test' is the practice GPs' decision to refer patients via the USC route. We used these measures because they have been promoted for comparison between practices and are currently being made publicly available (Meechan *et al*, 2012; National Cancer Intelligence Network, 2014).

In both databases, we found that some practices had only a small number of cancer cases. For the NHS Grampian data set, data from practice-years that contained no cancers diagnosed via the USC pathway were included in the analysis of aggregated practice data but excluded from investigation of detection and conversion rates. In the NHS England data, where there were less than six cancers referred or diagnosed in a year for a practice, the exact number was not published; thus, the corresponding record from that year was excluded from all analyses.

For each individual practice-year and for practice data aggregated across all years, we plotted detection and conversion rates as scatter plots, with lines fitted by linear and local polynomial regression, and calculated the correlation coefficient between detection and conversion rates. This followed the method previously used on single-year data (Meechan *et al*, 2012) and used the Spearman rank correlation coefficient with bootstrapped confidence interval (CI) method for both data sets. We calculated the mean and standard deviation of the detection and conversion rates for aggregate data from all practices and by three different case volumes of cancer diagnosis over the study periods (1–75, 76–150 and 151–400 cases).

Year to year correlation. We calculated the year-to-year correlation of practice detection and conversion rates, using the Spearman rank correlation coefficient, for all pairs of adjacent years. To examine the effects of number of cancers on these correlations, we analysed this by subgroups of practices according to the number of new cancer cases in 1 year.

Simulation modelling. In order to examine the effect of random case-mix at practices of different sizes, we generated sets of simulated GP 'practices'. Within each set, we introduced three sources of variation in measured performance: true practice variation, random case-mix and practice case numbers. These were introduced as follows:

True practice variation. This represents the between-practice variation in performance, such as would result from differences in competence, population or organisation. It was introduced by randomly allocating each practice its own practice detection rate, taken from a normal distribution with the mean set at the overall mean detection rate from the NCIN data (0.477). The standard

deviation for this distribution was set to one of three arbitrary values chosen to represent low, moderate and high true practice variation (0.025, 0.05 and 0.075, respectively).

Practice case numbers. This was introduced to examine the effect of different numbers of cancer cases on reported variation. We set practice case numbers at values of 25, 50, 75, 100, 150 and 200 cases. For reference, an average-sized practice with ~6000 patients can expect around 25 new cancers in a year.

Random case-mix. This represents the within-practice variation in performance resulting from the characteristics of individual cancer cases. It assumes that the difficulty in diagnosis lies in each case, such that two equally performing doctors, following guidelines, would vary in observed performance according to the cases they saw. It was introduced by having practices randomly sample their specified number of 'cases' from a larger pool of cases. Cases in this larger pool were all allocated a 'referral route' property in advance (either USC or other), with the proportion of USC cases set at the practice detection rate. As a result, the observed detection rate for each 'practice' represented a single sampling from a binomial distribution whose probability parameter was sampled from a Gaussian distribution.

Modelling procedure. We created 18 model specifications (three levels of between-practice variation \times six practice case numbers). Each specification of the model was constructed for sets of 1000 practices and run 200 times. Within each specification, we recorded the minimum, maximum, mean and standard deviation of the detection rate within each run. We then summarised them by calculating the medians of these measures over all runs.

Within each specification, we designated 'practices' whose practice detection rate was in the lowest decile of the distribution as poorly performing. After adding the effects of random case-mix, we recorded the number of these that were correctly identified as poorly performing (still in the lowest decile of the distribution). We also recorded the number of practices that were incorrectly identified as poorly performing (i.e., practice detection rate outside the lowest decile before introducing the case-mix, but in the lowest decile afterwards). For both of these measures, we reported the median and interquartile range across all of the runs for each model specification.

Comparison with published data. We compared the standard deviation seen in each of the model specifications with the data from both NHS Grampian and NHS England for comparable cancer case numbers (expressed as a range either side of the model specification number). Statistical analyses and modelling were conducted using SPSS for Windows Version 20 (IBM Corp, Armonk, NY, USA) and R version 3.02.

RESULTS

NHS Grampian data included 25 278 USC referrals and 10 615 cancers from 77 practices over a 7-year period. When practice-years with no cancers diagnosed via the USC pathway in the NHS Grampian data set were excluded, there were 24 934 USC referrals and 9945 cancers. NHS England data included 4 158 358 USC referrals and 955 502 cancers from 8303 practices over a 4-year period. When incomplete data (from practice-years with less than six recorded cancer cases referred by USC) were excluded from the NHS England data, there were 3 808 406 USC referrals and 881 078 cancers remaining from 6735 practices. Characteristics of the included practices are shown in Table 1. The NHS Grampian data set included a higher proportion of very small practices (list size <3000) than in England (20.8% vs 9.3%). The mean and standard

deviation practice detection and conversion rates for data aggregated over 7 years in the NHS Grampian were 0.38 (0.10) and 0.18 (0.06), respectively; for NHS England over 4 years, they were 0.48 (0.09) and 0.12 (0.06).

Table 2 shows that, overall, 4003 (37.7%) cancers in the NHS Grampian data set were referred by the USC pathway. The detection rate was particularly low in 2006–2008; this finding appears to reflect unfamiliarity with implementation of the USC referral system as more than half of cancers diagnosed after urgent referral went through generic urgent pathways as opposed to cancer-specific ones (data available on request). From 2009 onwards, 3435 out of 6639 (51.6%) cancers in the NHS Grampian were referred by the USC pathway. In the NHS England data set, 413 718 out of 881 080 (47.0%) cancers were referred by the USC pathway.

Correlation coefficients between practice detection and conversion rates for individual years ranged from 0.08 (95% CI –0.25 to 0.35) to 0.28 (0.08 to 0.53) in NHS Grampian and between 0.24 (0.21 to 0.26) and 0.26 (0.23 to 0.28) in NHS England, as shown in Table 3. When practice data over several years were aggregated, the correlation was weakened in both data sets: 2006–2012 in NHS Grampian –0.22 (–0.41 to 0.08) and in NHS England 0.12 (0.09 to 0.14). Limiting the NHS Grampian data to cancers diagnosed from 2009 onwards had little effect on the pooled correlation: –0.16 (–0.34 to 0.01). The relationship between detection and conversion rates is shown graphically in Figure 1 for years common to both data sets and for aggregated data. However, the positive correlations between detection and conversion rates in individual years may be spurious. Both detection and conversion rates feature the number of cancers diagnosed via the USC pathway as both the numerator and part of the denominator. Consequently, in a practice-year with a high proportion of 'clinically obvious' cancers, that is, clinical presentations that clearly meet criteria for USC referral, both detection and conversion rates will be relatively high. Conversely, when the proportion of clinically obvious cancers is low, both detection and conversion rates will be low. This may lead to a spurious correlation between the detection and conversion rates, which is more likely to occur with small sample sizes and/or substantial

Table 1. Practice, detection and referral characteristics of the two databases

	Northeast Scotland		NHS England	
	Number (N = 77)	%	Number (N = 6735)	%
Practice list size				
<3000	16	20.8	628	9.3
3001–6000	16	20.8	2121	31.5
6001–12000	34	44.2	3079	45.7
>12000	11	14.3	907	13.5
Age-standardised annual referral ratio^a				
<0.8	30	39.0	1936	28.7
0.8–1.0	30	39.0	3113	46.2
>1.2	17	22.1	1686	25.0
Overall rates^b				
	Mean	s.d. ^c	Mean	s.d.
Detection rate	0.382	0.098	0.477	0.085
Conversion rate	0.178	0.058	0.124	0.057
Detection rate by number of cases				
25–75 cases	0.361	0.087	0.485	0.085
76–150 cases	0.353	0.072	0.464	0.073
151–400 cases	0.389	0.057	0.468	0.064

^aThe indirectly standardised number of referrals via the urgent suspected cancer pathway relative to a practice's size and age and sex composition.
^bRates for NHS Grampian data include all 7 years (2006–2012).
^cStandard deviation.

Table 2. Annual data pooled from all practices to show the number of cancers detected by different pathways, number of USC referrals and calculated annual detection and conversion rates

NHS Grampian	Cancers			Referrals USC	Rates (pooled practices) ^a	
	Total ^b	Urgent (USC) ^c	Urgent (other) ^d		Detection	Conversion
2006	1133	144	374	1079	0.127	0.133
2007	1186	167	357	1233	0.141	0.135
2008	1644	257	605	2097	0.156	0.123
2009	1667	643	464	3764	0.386	0.171
2010	1677	919	279	5143	0.548	0.179
2011	1783	975	320	5904	0.547	0.165
2012	1525	898	240	6058	0.589	0.148
Pooled 2006–2012	10615	4003	2639	25278	0.377	0.158
Pooled 2009–2012	6639	3435	1303	20774	0.516	0.165

NHS England	Cancers			Referrals USC	Rates (pooled practices) ^a	
	Total ^e	Urgent (USC)			Detection	Conversion
2010	199 317	89 027	—	772 840	0.447	0.115
2011	216 957	101 260	—	907 164	0.467	0.112
2012	229 173	109 002	—	1 007 414	0.476	0.108
2013	235 634	114 429	—	1 120 988	0.486	0.102
Pooled 2010–2013	881 080	413 718	—	3 808 406	0.470	0.109

^aThese rates are for all patients pooled across practices. They are thus slightly different from the mean practice rates shown in Table 1.
^bIncludes data from all practice-years whether or not any cancer was referred via the USC (urgent suspected cancer) pathway.
^cTotal number of cancers diagnosed after referral through the USC pathway.
^dTotal number of cancers diagnosed after referral through other urgent pathway. Before 2006, this was the only pathway available in NHS Grampian Scotland. This analysis was only available at practice level in Scotland.
^eIncludes data only from practice-years in which full data were available (at least six cancer cases referred via the USC pathway).

case-mix. When numbers of cancers are pooled, this spurious correlation will be diminished.

In the NHS England data, there was a clear difference between detection and conversion rates in their year-to-year correlations. For practice detection rates, practice pooled correlations between pairs of consecutive years were weak, between 0.20 (95% CI 0.17 to 0.22) and 0.26 (0.23 to 0.29), suggesting practice detection rates were not consistent from year to year. In contrast, year-to-year correlations for practice pooled conversion rate were moderately strong, between 0.53 (0.51 to 0.55) and 0.55 (0.53 to 0.57), suggesting greater consistency from year to year. Table 4 shows the results of this analysis with practices subgrouped according to their number of cases. Similar year-to-year analysis on NHS Grampian data was limited by quite small numbers after excluding practices with no cancers in one or the other year, and thus CIs were wide (Table 4).

The simulation modelling of detection rates is reported in Table 5. The first column indicates whether the model specification included low, medium or high between-practice variation. The observed standard deviation values in the second column are consistently larger than the standard deviation values used to simulate true practice variation, indicating that case-mix increases the observed variance. Unsurprisingly, this difference is greater when the number of cases is small. The implications of this increase in variance due to case-mix are shown in the columns of Table 5 relating to 'poor performing practices'. The first pair shows the number of practices in the simulations (median with interquartile range) that were specified as poorly performing before the addition of case-mix and were subsequently detected after the introduction of case-mix variation. The second pair of columns shows the total number of practices that were in the lowest decile for detection rate after introducing random case-mix.

Comparison of the pattern of standard deviations for different levels of between-practice variation (Table 5) with the standard

deviations for practices whose aggregate number of cancers was low (25–75), medium (76–150) or large (151–400) from the empirical data in Table 1 suggests that the model with moderate between-practice variation is most closely matched to the actual data from both empirical data sets. The implication of this is that, with 25 cancer cases per practice, only a median of 20% (IQR 17 to 24) poorly performing practices will be correctly identified and most practices identified as poorly performing will be incorrectly labelled. As the number of cases per practice rises, the accuracy of prediction increases, but only slowly and incompletely: with 100 cases the probability of a poorly performing practice being correctly identified is 44% (95% CI 40 to 47), and with 200 cases it is 57% (54 to 59).

DISCUSSION

This is the first study to examine the effects of number of cases and random case-mix on a publicly reported measure of GP practices' performance in the use of USC pathways. The results indicate that differences between practices, and apparent quality gradients, seen within individual years are weakened when data are aggregated over several years, suggesting that case-mix – different cancers with different referral pathways – rather than actual clinical performance accounts for much of the observed variation. The modelling exercise suggests that at least 100 cancer cases per practice are necessary before the probability is close to 50% that an observed poorly performing practice is actually poorly performing. In terms of an average-sized UK practice, it would take ~4 years for the number of cancer cases to accrue.

Our choice of two databases allows for both the extra detail and duration of the northeast Scotland data set and the breadth and generalisability of the NHS England data set. The fact that broadly comparable results emerged strengthens our findings

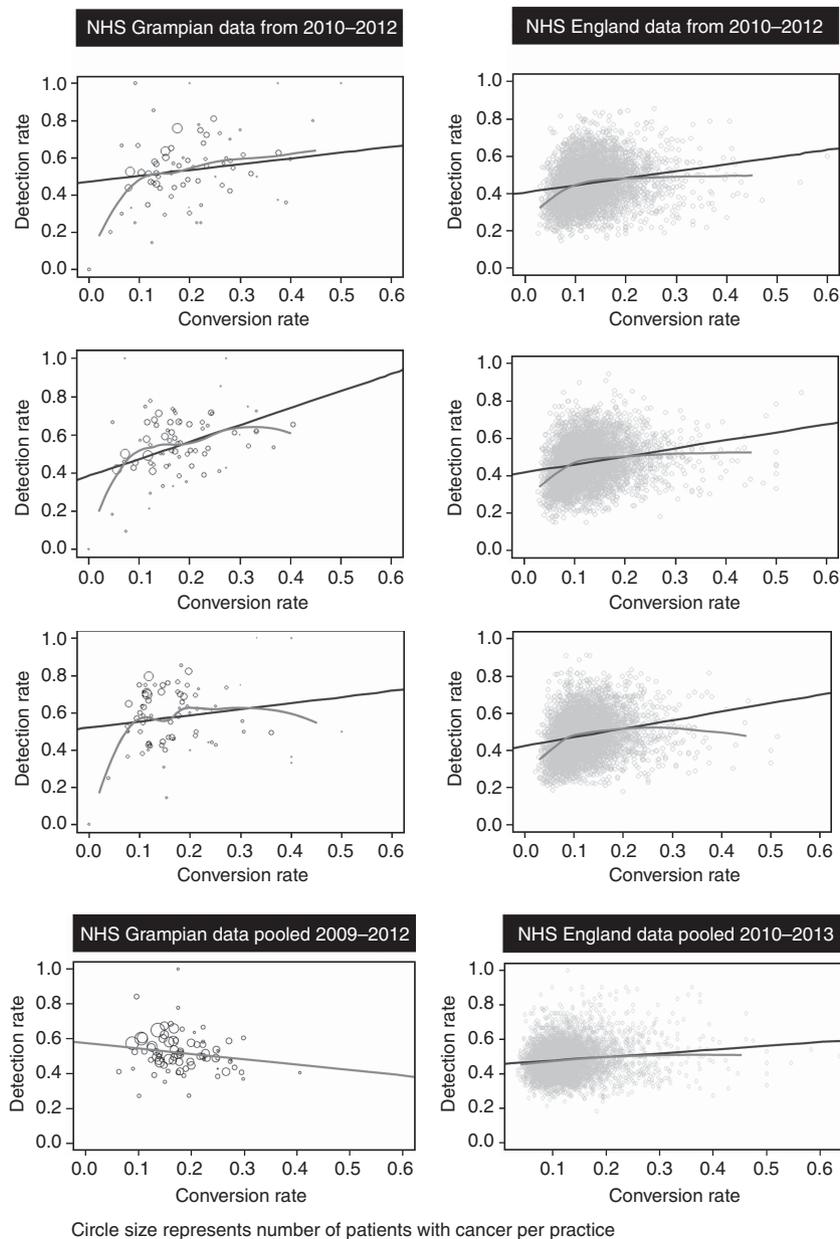


Figure 1. Scatter plots of detection rate vs conversion rate in different years, and aggregated across all available years, for each database. Lines indicate linear regression (blue) and local polynomial regression (red). Northeast Scotland data points ($N = 77$) represent the practice size. A full colour version of this figure is available at the *British Journal of Cancer* journal online.

and adds credence to the belief that, in both data sets, the validity and accuracy of cancer cases and referrals routes are acceptable. Although the Scottish and English data sets were broadly similar, they were not identical in their definitions, nor in the time of the study in relation to introduction of specific cancer referral pathways. Furthermore, the Scottish data were from one region only, whereas the English data were from the whole country. This precludes making comparisons between the data sets, and instead the analysis focuses on within-data set comparisons showing that similar findings appear in both. Both data sets had limitations when it came to small numbers of cancer cases; in the NHS England database, data from practices with less than six USC referrals in 1 year were not available. We did not attempt to impute these data. In addition, we did not attempt to address other possible sources of practice variation, such as differences in practice population rather than GP performance. Within the Scottish data, there were some single-doctor practices with very small list sizes, meaning that it was possible they would have a

year when no cancers were diagnosed. These null data were excluded from the analysis since a meaningful detection rate could not be calculated. However, as our data suggest relatively modest variation in GP performance after adjusting for random case-mix, any adjustment for practice characteristics would be likely to further reduce the variation in intrinsic GP performance.

We deliberately followed previously reported methods for comparing single-year interpractice variation in the use of USC referrals, as these measures are currently used in routine reporting and public feedback (Meehan *et al*, 2012). This is despite the fact that the method has an obvious limitation in that detection and conversion rates are not a naturally complementary pair of measures. One unexpected advantage of this pairing, however, is that, unlike sensitivity and specificity, the relationship between these two measures is not influenced by prevalence. Our finding that the correlation between detection and conversion rates was diminished (or reversed) by aggregating data over several years

Table 3. Correlation coefficients (with 95% CIs) between the detection rate and conversion rate by year and aggregated over all available years for each database

Year	NHS Grampian			NHS England		
	Coefficient ^a	95% CI	P-value	Coefficient ^a	95% CI	P-value
2006	0.18	-0.13 to 0.47	0.21	—	—	—
2007	0.28	-0.05 to 0.48	0.05	—	—	—
2008	0.13	-0.07 to 0.38	0.32	—	—	—
2009	0.28	0.08 to 0.53	0.01	—	—	—
2010	0.27	0.06 to 0.44	0.02	0.24	0.21 to 0.26	<0.001
2011	0.26	0.03 to 0.48	0.03	0.26	0.23 to 0.28	<0.001
2012	0.08	-0.25 to 0.35	0.48	0.25	0.23 to 0.27	<0.001
2013	—	—	—	0.24	0.22 to 0.27	<0.001
Aggregated 2006–2012	-0.22	-0.41 to 0.08	0.08			
2009–2012	-0.16	-0.34 to 0.01	0.12			
2010–2013				0.12	0.09 to 0.14	<0.001

Abbreviation: CI = confidence interval.
^aSpearman correlation coefficient with bootstrapped 95% confidence intervals.

Table 4. Year to year correlation coefficients (95% confidence interval) for detection rate and conversion rate from GP practices in England

Cancer Cases (per year)	Years	Practices	Year to year correlation (95% confidence interval) ^a	
			Detection rate	Conversion rate
N = 6–25	2010 vs 2011	1353	0.21 (0.15 to 0.26)	0.60 (0.56 to 0.63)
	2011 vs 2012	1384	0.16 (0.11 to 0.22)	0.60 (0.56 to 0.64)
	2012 vs 2013	1401	0.16 (0.11 to 0.21)	0.60 (0.56 to 0.64)
N = 26–50	2010 vs 2011	2416	0.28 (0.24 to 0.32)	0.48 (0.45 to 0.51)
	2011 vs 2012	2471	0.20 (0.16 to 0.24)	0.52 (0.49 to 0.55)
	2012 vs 2013	2443	0.22 (0.18 to 0.26)	0.53 (0.50 to 0.56)
N = 51–75	2010 vs 2011	875	0.29 (0.23 to 0.35)	0.55 (0.50 to 0.59)
	2011 vs 2012	1008	0.22 (0.16 to 0.27)	0.50 (0.45 to 0.54)
	2012 vs 2013	1108	0.21 (0.15 to 0.26)	0.52 (0.47 to 0.56)
N > 75	2010 vs 2011	250	0.33 (0.20 to 0.45)	0.58 (0.48 to 0.67)
	2011 vs 2012	316	0.29 (0.19 to 0.39)	0.55 (0.46 to 0.62)
	2012 vs 2013	394	0.35 (0.27 to 0.43)	0.63 (0.56 to 0.69)
All practices	2010 vs 2011	4894	0.26 (0.23 to 0.29)	0.53 (0.51 to 0.55)
	2011 vs 2012	5179	0.20 (0.17 to 0.22)	0.54 (0.52 to 0.57)
	2012 vs 2013	5346	0.21 (0.19 to 0.24)	0.55 (0.53 to 0.57)
All NHS Grampian	2006 vs 2007	38	0.26 (-0.07 to 0.55)	0.3 (-0.01 to 0.58)
	2007 vs 2008	41	0.47 (0.2 to 0.66)	0.48 (0.21 to 0.68)
	2008 vs 2009	62	0.25 (-0.01 to 0.42)	0.47 (0.27 to 0.61)
	2009 vs 2010	74	0.28 (0.09 to 0.44)	0.41 (0.16 to 0.61)
	2010 vs 2011	72	-0.01 (-0.24 to 0.21)	0.05 (-0.18 to 0.25)
	2011 vs 2012	73	0.1 (-0.11 to 0.33)	0.2 (-0.01 to 0.39)

^aSpearman rank correlation coefficient with bootstrapped 95% confidence intervals.

suggests that correlations based on small case numbers may be largely spurious. In our modelling, we focused on the detection rate, as this was the measure highlighted by media and politicians, and also because this showed more year on year variation than the conversion rate. It is not the only measure of diagnostic quality, however, and further work may need to examine conversion rates or other approaches such as imputing specificity.

Our modelling exercise was based on empirical data from NHS England with the detection rate set according to this. It was designed to use sample sizes that are representative of

routine practice. With an annual cancer incidence of around 4.3/1000 per annum, a small practice of 3000 patients can expect only around 13 new cancer cases each year, and will need 8 years to accumulate ~100 cases. A medium-sized practice of 6000 patients (close to the UK average) will need 4 years. Only a very large practice of around 24 000 patients is likely to record 100 cancer cases in a year. Even with 100 cases, our modelling data indicate that the probability of a statistically outlying practice being an intrinsically poor performer is still only around 50%. At present, most practices identified as poorly

Table 5. Results of the modelling of practice detection rates with both specified between-practice variation and random case-mix variation

Model parameters ^a	'Poorly performing practices'						
	Observed data (detection rate) ^b			Correctly identified ^c		Total identified ^d	
Cases	s.d. ^e	Min	Max	Median	IQR	Median	IQR
Low variation (s.d. = 0.025)							
25	0.1	0.16	0.8	18	14 – 19	87	80 – 93
50	0.07	0.24	0.72	19	16 – 21	77	70 – 82
75	0.06	0.27	0.69	24	21 – 27	87	79 – 92
100	0.06	0.29	0.66	28	24 – 29	87	78 – 94
150	0.05	0.32	0.64	30	27 – 34	89	84 – 95
200	0.04	0.34	0.63	35	33 – 38	91	86 – 96
Moderate variation (s.d. = 0.05)							
25	0.11	0.12	0.84	20	17 – 24	58	53 – 64
50	0.09	0.2	0.76	31	28 – 35	77	69 – 89
75	0.08	0.23	0.73	39	36 – 43	85	78 – 94
100	0.07	0.25	0.71	44	40 – 47	89	83 – 95
150	0.06	0.27	0.69	52	48 – 55	91	88 – 95
200	0.06	0.28	0.68	57	54 – 59	94	89 – 97
High variation (s.d. = 0.075)							
25	0.12	0.08	0.88	32	29 – 36	75	70 – 83
50	0.1	0.14	0.82	44	40 – 48	81	75 – 89
75	0.09	0.16	0.79	52	49 – 55	90	83 – 95
100	0.09	0.18	0.77	57	54 – 60	91	86 – 97
150	0.09	0.2	0.76	64	62 – 67	94	90 – 97
200	0.08	0.21	0.75	68	66 – 71	95	92 – 98

Abbreviations: IQR = interquartile range; Max = maximum; Min = minimum.
^aFor each set of parameters, 200 sets of 1000 practices were modelled.
^bValues for s.d., minimum and maximum represent median values for all runs of the simulation at each specification of practice variation and number of cancer cases.
^cCorrectly identified poorly performing practices represents the median number (with interquartile range) of practices, which were in the lowest decile of detection rate before adjusting for case-mix (N = 100) and which were also in the lowest decile of detection rate after introducing random case-mix variation.
^dTotal identified poorly performing practices represents the median number (with interquartile range) of practices, which were in the lowest decile of detection rate after introducing random case-mix variation.
^eStandard deviation.

performing from annual data will be wrongly labelled, and most poorly performing practices will not be detected.

Individual case presentations of cancer vary and this study highlights the need to consider the number of cases and case-mix variation in evaluating the performance of GP practices in their use of USC referral. Simply reporting values for detection and conversion rates on annual data, and considering all cancers together, has a clear potential to mislead both practices and the public. Allied concerns relating to case-mix and small volume caseloads have recently been highlighted in reporting surgeon performance (Walker *et al*, 2013).

On the basis of our findings, we propose that any reporting of practice rates should now be limited to data aggregated over several years and may need to consider each cancer site separately. However, the substantial effect of random case-mix on observed detection rates, even if GPs follow guidelines exactly, means that alternative approaches are needed. We suggest two ways to avoid the influence of case-mix on reported performance. The first is to examine specifically whether cancer cases were referred (or not) in accordance with national guidelines. The second is to adopt a 'confidential enquiry' approach, employing a case review of designated delayed diagnoses as 'never events' (de Wet *et al*, 2014). Although both require more data, and more time to critically reflect on it, they would be more transparent – and more likely to lead to constructive changes in practice – than the current crude approach to identification of supposedly poorly performing practices. In the meantime, the widespread public reporting of GP practice's use of USC referral pathways based on annual data may be misleading and should be interpreted with caution until a more robust reporting methodology is in place.

ACKNOWLEDGEMENTS

The study sponsor is the University of Aberdeen.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DISCLAIMER

The lead author PM affirms that this manuscript is an honest, accurate and transparent account of the study being reported, that no important aspects of the study have been omitted, and that any discrepancies from the study as planned have been explained. The opinion that formal ethical approval was not required to analyse this completely anonymous data set was received from the North of Scotland Research Ethics Committee on 23 May 2012. The study was approved by the Caldicott Guardian of NHS Grampian on 26th February 2013.

AUTHOR CONTRIBUTIONS

PM, AC and NCC designed the study of NHS Grampian data; AC and SS conducted the initial descriptive analysis. CDB and PM extended the study to NHS England data. CDB designed the modelling in collaboration with AJL and PM and conducted the final analysis. SS wrote the initial draft of the manuscript, which was subsequently revised by all contributors. PM is the guarantor of results.

REFERENCES

- Baughan P, Keatings J, O'Neill B (2011) Urgent suspected cancer referrals from general practice: audit of compliance with referral guidelines and referral outcomes. *Br J Gen Pract* **61**: e700–e706.
- BBC (2013) Many cancer patients' not referred to specialist by GP. Available at <http://www.bbc.co.uk/news/health-25274287> (accessed 23 September 2014).
- BBC (2014) GPs who fail to spot cancer should be named. Available at <http://www.bbc.co.uk/news/health-28077586> (accessed 23 September 2014).
- Birring SS, Peake MD (2005) Symptoms and the early diagnosis of lung cancer. *Thorax* **60**: 268–269.
- Bottle A, Tsang C, Parsons C, Majeed A, Soljak M, Aylin P (2012) Association between patient and general practice characteristics and unplanned first-time admissions for cancer: observational study. *Br J Cancer* **107**: 1213–1219.
- Coleman MP, Forman D, Bryant H, Butler J, Rachet B, Maringe C, Nur U, Tracey E, Coory M, Hatcher J, McGahan CE, Turner D, Marrett L, Gjerstorff ML, Johannesen TB, Adolfsson J, Lambe M, Lawrence G, Meechan D, Morris EJ, Middleton R, Steward J, Richards MA (2011) Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *Lancet* **377**: 127–138.
- de Wet C, O'Donnell C, Bowie P (2014) Developing a preliminary 'never event' list for general practice using consensus-building methods. *Br J Gen Pract* **64**: e159–e167.
- Department of Health (2000) HSC2000/013 Referral Guidelines for suspected cancer. London, UK.
- Dua RS, Brown VS, Loukogeorgakis SP, Kallis G, Meleagros L (2009) The two-week rule in colorectal cancer. Can it deliver its promise? *Int J Surg* **7**: 521–525.
- Meechan D, Gildea C, Hollingworth L, Richards MA, Riley D (2012) Rub in Variation in use of the 2-week referral pathway for suspected cancer: a cross-sectional analysis. *Brit J Gen Pract* **62**: e590–e597.
- National Cancer Intelligence Network (2012) General practice profiles for cancer. Available at http://www.ncin.org.uk/cancer_information_tools/profiles/gp_profiles (accessed 23 September 2014).
- National Cancer Intelligence Network (2014) Routes to diagnosis. Available at http://www.ncin.org.uk/publications/routes_ (accessed 23 September 2014).
- NICE (2005) Referral guidelines for suspected cancer. London: National Institute for Health and Clinical Excellence.
- Richards MA (2009) The National Awareness and Early Diagnosis Initiative in England: assembling the evidence. *Br J Cancer* **101**: S1–S4.
- Scottish Executive Health Department (2007) Scottish referral guidelines for suspected cancer. Edinburgh, UK.
- Scottish Government (2009) Scottish referral guidelines for suspected cancer. Quick reference guide. Edinburgh, UK.
- Walker K, Neuburger J, Groene O, Cromwell DA, van der Meulen J (2013) Public reporting of surgeon outcomes: low numbers of procedures lead to false complacency. *Lancet* **382**: 1674–1677.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 4.0 Unported License.

Supplementary Information accompanies this paper on British Journal of Cancer website (<http://www.nature.com/bjc>)