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

Peter Murchie, Anika Chowdhury, Sarah Smith, Neil C Campbell, Amanda J Lee, David Linden, Christopher D Burton

Peter Murchie
Senior Lecturer in Primary Care, Division of Applied Health Sciences, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen, AB25 2ZD
Anika Chowdhury
Medical Student, Division of Applied Health Sciences, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen, AB25 2ZD
Sarah Smith
Research Fellow, Division of Applied Health Sciences, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen, AB25 2ZD
Neil C Campbell
Reader in Primary Care, Division of Applied Health Sciences, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen, AB25 2ZD
Amanda J Lee
Professor of Medical Statistics, Division of Applied Health Sciences, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen, AB25 2ZD
David Linden
Director, Detect Cancer Early Programme The Scottish Government, St Andrews House, Regent Rd, Edinburgh EH1 3DG
Christopher D Burton
Senior Lecturer in Primary Care, Division of Applied Health Sciences, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen, AB25 2ZD

Correspondence to Dr Peter Murchie: p.murchie@abdn.ac.uk

Contributors
PM, AC, and NC designed the study of NHS Grampian data; AC and SS conducted the initial descriptive analysis. CB and PM extended the study to NHS England data. CB designed the modelling in collaboration with AL 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.

Transparency Statement
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.

Study Approval

The opinion that formal ethical approval was not required to analyze this completely anonymous dataset 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.

Funding

This study received no direct external funding.

Sponsor

The study sponsor is the University of Aberdeen.

Competing Interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Data Sharing **Word Count:** 3,828

No additional data available.

Short Abstract (219 words)

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

Methods: We analysed individual practice urgent suspected cancer referral (USC) detection (number of practice's cancer detected via USC) and conversion rates (number of practice's USC referrals which are cancer) in routinely collected data on cancer referrals from GP practices in all of England (over four years) and North-east Scotland (over seven 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).

Conclusion: 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.

Abstract (300 words)

Objective:

Publicly available data are reported as showing unacceptable variation in GPs use of urgent suspected cancer referral pathways. We investigated how much of this variation in performance is due to small numbers of cancer cases and random case-mix of cancer presentations.

Design:

Analysis of routinely collected data on cancer referrals from GP practices and modelling of the effects of adding random case-mix to practice variation. Examination of detection rate (proportion of cancers referred by the urgent pathway) and conversion rate (proportion of urgent pathway referrals diagnosed with cancer).

Setting:

955,502 cancer cases from 8,303 practices in NHS England over four years; 10,615 cancer cases from 77 practices in NHS Grampian, Scotland over seven years. Modelling conducted with simulated practices having between 25 and 200 cancer cases.

Results:

Correlations between practice detection rate and conversion rate were weaker when data were aggregated over several years compared to within individual years: NHS England aggregated 0.12 (95% confidence interval 0.09 to 0.14), individual years between 0.24 (0.21 to 0.26) and 0.26 (0.23 to 0.28), NHS Grampian aggregated -0.22 (-0.41 to 0.08), individual years between 0.08 (-0.25 to 0.35) and 0.28 (0.08 to 0.53); Year to year correlation was weaker for detection rate, between 0.20 (0.17 to 0.22) and 0.26 (0.23 to 0.29), than for conversion rate, between 0.53 (0.51 to 0.55) and 0.55 (0.53 to 0.57), implying detection rate is less consistent over time. Adding random case-mix to simulated between-practice variation resulted in the

median proportion of poorly performing practices correctly identified after 25 cancer cases were examined being 20% (IQR 17 to 24) and after 100 cases 44% (40 to 47).

Conclusions: Much apparent variation in general practitioners' use of suspected cancer referral pathways can be attributed to random case-mix. Measures from single years of data are misleading and should not be publically reported.

INTRODUCTION

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 to other developed countries. This effect is widely attributed to longer intervals in the cancer diagnostic and treatment pathway [Coleman et al, 2011; Richards et al, 2009].

The UK NHS has existing fast-track urgent suspected cancer (USC) referral pathways from primary to secondary care [NICE, 2007; Scottish Government, 2009]. In England patients referred from primary care with suspected cancer should be seen in secondary care within two 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 use of this referral route [Baughan et al, 2011]. Similar variation was shown in an analysis of 865,494 referrals in England over one 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). 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 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 [NCIN, 2014]. 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, 2012; 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, so 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) while others typically have non-specific 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 [NCIN, 2014] Even within the same cancer type, some presentations will prompt urgent referral (e.g. lung cancer presenting with haemoptysis) while others may not (e.g. lung cancer presenting with non-specific symptoms) [Birring et al, 2005]. Furthermore pre-symptomatic cancers which 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, so 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.

METHODS

Data sources

The current study used two databases. The first provided data on route of cancer diagnosis, including USC referral from primary care over seven years (2006-12) in the NHS Grampian region of Northeast Scotland. The second contained data on route of cancer diagnosis including two-week referrals over four years (2010-2013) from NHS England. USC in Scotland and two-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 02 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 this 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 [NCIN, 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 datasets, we calculated the detection rate and conversion rate 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' decisions 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 publically available [Meechan et al, 2012; NCIN, 2014].

In both databases, we found that some practices had only a small number of cancer cases. For the NHS Grampian dataset, data from practice-years which 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, so 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 intervals method for both datasets. We calculated the mean (SD) of 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 cases, 76-150, and 151-400).

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 one 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 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 approximately 6000 patients can expect around 25 new cancers in a year.

Random casemix

This represents 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 (3 levels of between-practice variation x 6 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 the 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 which were correctly identified as poorly performing (still in the lowest decile of the distribution). We also recorded the number of practices which were incorrectly identified as poorly performing (ie Practice Detection Rate outside the lowest decile before introducing case-mix, but in the lowest decile afterwards). For both 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 deviations 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 and R version 3.02.

RESULTS

NHS Grampian data included 25,278 USC referrals and 10,615 cancers from 77 practices over a seven year period. When practice-years with no cancers diagnosed via the USC

pathway in the NHS Grampian dataset were excluded there were 24,934 USC referrals and 9,945 cancers. NHS England data included 4,158,358 USC referrals and 955,502 cancers from 8,303 practices over a four year period. When incomplete data (from practice-years with <6 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 6,735 practices. Characteristics of the included practices, including are shown in table 1. The NHS Grampian dataset included a higher proportion of very small practices (list size <3000) than in England (20.8% vs 9.3%). The mean (SD) practice detection and conversion rates for data aggregated over seven years in NHS Grampian were 0.38 (0.10) and 0.18 (0.06) respectively; for NHS England over four years they were 0.48 (0.09) and 0.12 (0.06).

Table 2 shows that overall, 4,003 (37.7%) of cancers in the NHS Grampian dataset were referred by the USC pathway. The detection rate was particularly low in 2006-8, this finding appears to reflect unfamiliarity with implementation of the USC referrals 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, 3,435 out of 6,639 (51.6%) of cancers in NHS Grampian were referred by the USC pathway. In the NHS England dataset 413,718 out of 881,080 (47.0%) of cancers were referred by the USC pathway.

Correlation coefficients between practice detection rates and conversion rates for individual years ranged from 0.08 (95% confidence interval -0.25 to 0.35) and 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 datasets: 2006-12 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 rates and conversion rates is shown graphically in Figure 1 for years common to both datasets and for aggregated data. However, the positive correlations between detection rates and conversion rates in individual years may be spurious. Both detection rate and conversion rate 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 i.e. clinical presentations which clearly meet criteria for USC referral, both detection rate and conversion rate will be relatively high. Conversely, when the proportion of clinically obvious cancers is low, both detection rate and conversion rate will be low. This may lead to a spurious correlation between detection rate and conversion rate which is more likely to occur with small sample sizes and/or substantial 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 rates 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% confidence interval 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 sub-grouped 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 other year, so confidence intervals 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 deviations in the third column are consistently larger than the standard deviations 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 4 relating to "poor performing practices". The first pair show number of practices in the simulations (median with interquartile range) which 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 which 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 (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 datasets.. 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, 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 urgent suspected cancer 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 approximately four years for this number of cancer cases to accrue.

Our choice of two databases allows for both the extra detail and duration of the Northeast Scotland dataset and the breadth and generalisability of the NHS England dataset. The fact that broadly comparable results emerged strengthens our findings and adds credence to the belief that, in both datasets the validity and accuracy of cancer cases and referrals routes are acceptable. While the Scottish and English datasets 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 datasets, and instead, the analysis focuses on within-dataset comparisons showing that similar findings appear in both. Both datasets 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 one 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 inter-practice variation in the use of USC referrals as these measures are currently used in routine reporting and public feedback [Meechan 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 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 rate 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 which 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 eight years to accumulate approximately 100 cases. A medium sized practice of 6000 patients (close to the UK average) will need four 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 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 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 the 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].

Based on 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, mean 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 case review of designated delayed diagnoses as "never events" [de Wet et al, 2014]. While 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 urgent suspected cancer referral pathways based on

annual data may be misleading and should be interpreted with caution until a more robust reporting methodology is in place.

REFERENCES

Baughan P, Keatings J, O'Neill B. Urgent suspected cancer referrals from general practice: audit of compliance with referral guidelines and referral outcomes. *Br J Gen Pract* 2011 **61**:e700–e706.

BBC Wesbite 7th December 2013. Many cancer patients' not referred to specialist by GP. http://www.bbc.co.uk/news/health-25274287 (accessed 23 September 2014)

BBC Wesbite 29th June 2014. GPs who fail to spot cancer should be named. http://www.bbc.co.uk/news/health-28077586 (accessed 23 September 2014)

Birring SS, Peake MD. Symptoms and the early diagnosis of lung cancer. *Thorax* 2005;60:268-269.

Bottle A, Tsang C, Parsons C, Majeed A, Soljak M, Aylin P. Association between patient and general practice characteristics and unplanned first-time admissions for cancer: observational study. *Brit J Cancer* 2012;**107**:1213-1219.

Coleman MP, Forman D, Bryant H, J Butler J, Rachet B, C Maringe C et al. 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. *The Lancet* 2011;**377**:127-138.

de Wet C, O'Donnell C, Bowie P. Developing a preliminary 'never event' list for general practice using consensus-building methods. *Br J Gen Pract* 2014;**64**:e159-167. doi: 10.3399/bjgp14X677536

Department of Health. HSC2000/013 Referral Guidelines for suspected cancer. London: DoH, 2000.

Dua RS, Brown VS, Loukogeorgakis SP, Kallis G, Meleagros L. The two-week rule in colorectal cancer. Can it deliver its promise? *Int J Surg* 2009;7:521-525.

Meechan D, Gildea C Hollingworth L, Richards MA, Riley D, Rubin Variation in use of the 2-week referral pathway for suspected cancer: a cross-sectional analysis. *Brit J Gen Pract* 2012;**62**:e590-e597.

National Cancer Intelligence Network. General Practice Profiles for Cancer. Dec 2012; Available at: http://www.ncin.org.uk/cancer_information_tools/profiles/gp_profiles. (Accessed 23 September 2014).

National Cancer Intelligence Network. Routes to Diagnosis. Available at: http://www.ncin.org.uk/publications/routes_(Accessed 23 September 2014).

NICE. Referral guidelines for suspected cancer. London: National Institute for Health and Clinical Excellence, 2005.

Richards MA. The National Awareness and Early Diagnosis Initiative in England: assembling the evidence. *Brit J Cancer* 2009;**101**:S1-S4.

Scottish Executive Health Department. Scottish referral guidelines for suspected cancer. Edinburgh: Scottish Executive Health Department, 2007.

Scottish Government. Scottish referral guidelines for suspected cancer. Quick reference guide. Edinburgh: Scottish Government, 2009.

Walker K, Neuburger J, Groene O, Cromwell DA, van der Meulen J. Public reporting of surgeon outcomes: low numbers of procedures lead to false complacency. *The Lancet* 2013;**382**:1674-1677.

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

	Northeast Scotland		NHS England	
	Number	**%**	**Number**	**%**
Practice List Size	(N=77)		(N= 6735)	
<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[1]				
less than 0.8	30	39.0	1936	28.7
0.8-1.0	30	39.0	3113	46.2
more than 1.2	17	22.1	1686	25.0
Overall rates[2]	**Mean**	**SD**	**Mean**	**SD**
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

1. The indirectly standardised number of referrals via the Urgent Suspected Cancer pathway relative to a practices size and age and sex composition.
2. Rates for NHS Grampian data include all 7 years (2006-2012).

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

NHS Grampian	Cancers			Referrals	Rates (pooled practices)[1]	
	Total[2]	Urgent (USC)[3]	Urgent (other)[4]	USC	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 06-12	10615	4003	2639	25278	0.377	0.158
Pooled 09-12	6639	3435	1303	20774	0.516	0.165
NHS England	Total[5]	Urgent (USC)		USC	Detection	Conversion
2010	199317	89027	-	772840	0.447	0.115
2011	216957	101260	-	907164	0.467	0.112
2012	229173	109002	-	1007414	0.476	0.108
2013	235634	114429	-	1120988	0.486	0.102
Pooled 10-13	881080	413718	-	3808406	0.470	0.109

[1] These rates are for all patients pooled across practices, they are thus slightly different from the mean practice rates shown in table 1

[2] Includes data from all practice-years whether or not any cancer was referred via the USC (Urgent Suspected Cancer) pathway

[3] Total number of cancers diagnosed after referral through USC pathway

[4] Total number of cancers diagnosed after referral through other urgent pathway. Prior to 2006 this was the only pathway available in NHS Grampian Scotland. This analysis was only available at practice level in Scotland.

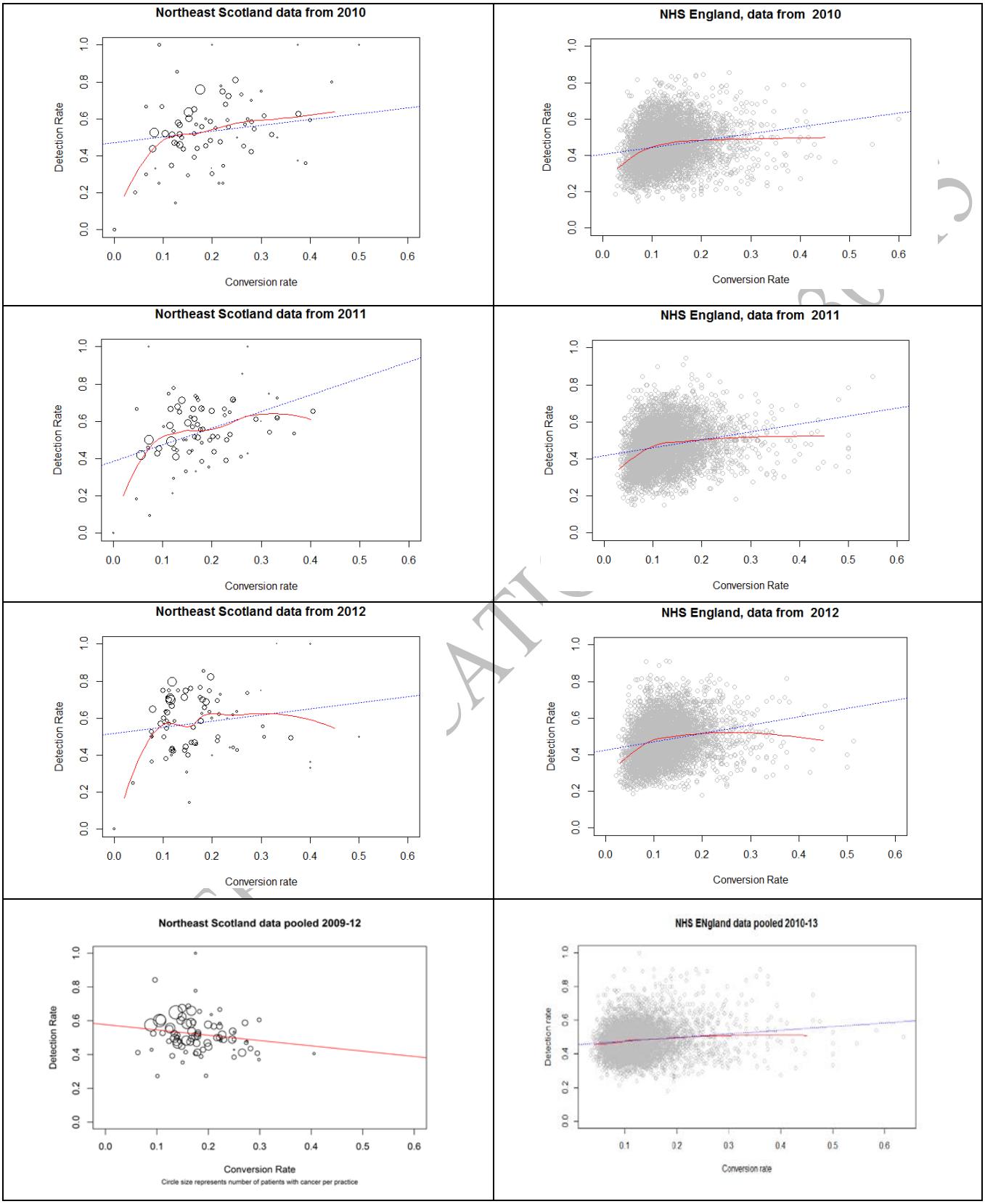
[5] Includes data only from practice years in which full data were available (at least 6 cancer cases referred via the USC pathway)

Table 3. Correlation coefficients (with 95% confidence intervals) between the Detection Rate and Conversion Rate by year and aggregated over all available years for each database

	NHS Grampian			NHS England		
Year	Coefficient[1]	95% CI	P value	Coefficient[1]	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						
2006-12	-0.22	-0.41 to 0.08	0.08			
2009-12	-0.16	-0.34 to 0.01	0.12			
2010-13				0.12	0.09 to 0.14	<0.001

[1] Spearman correlation coefficient with bootstrapped 95% confidence intervals.

Figure 1. Scatter plots of Detection Rate versus Conversion Rate in different years, and aggregated across all available years, for each database



Lines indicate linear regression (dotted) and local polynomial regression (solid)

Northeast Scotland data points (N=77) represent the practice size.

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% CI)[1] 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 (0.27to 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)

[1] Spearman rank correlation coefficient with bootstrapped 95% confidence intervals

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

Model parameters[1]		Observed data (detection rate)[2]			"Poorly performing practices"			
					Correctly identified[3]		Total identified[4]	
	Cases	SD	Min	Max	Median	IQR	Median	IQR
Low variation (SD = 0.025)	25	0.1	0.16	0.8	18	14 to 19	87	80 to 93
	50	0.07	0.24	0.72	19	16 to 21	77	70 to82
	75	0.06	0.27	0.69	24	21 to 27	87	79 to 92
	100	0.06	0.29	0.66	28	24 to 29	87	78 to 94
	150	0.05	0.32	0.64	30	27 to 34	89	84 to 95
	200	0.04	0.34	0.63	35	33 to 38	91	86 to 96
Moderate variation (SD = 0.05)	25	0.11	0.12	0.84	20	17 to 24	58	53 to 64
	50	0.09	0.2	0.76	31	28 to 35	77	69 to 89
	75	0.08	0.23	0.73	39	36 to 43	85	78 to 94
	100	0.07	0.25	0.71	44	40 to 47	89	83 to 95
	150	0.06	0.27	0.69	52	48 to 55	91	88 to 95
	200	0.06	0.28	0.68	57	54 to 59	94	89 to 97
High variation (SD = 0.075)	25	0.12	0.08	0.88	32	29 to 36	75	70 to 83
	50	0.1	0.14	0.82	44	40 to 48	81	75 to 89
	75	0.09	0.16	0.79	52	49 to 55	90	83 to 95
	100	0.09	0.18	0.77	57	54 to 60	91	86 to 97
	150	0.09	0.2	0.76	64	62 to 67	94	90 to 97
	200	0.08	0.21	0.75	68	66 to 71	95	92 to 98

[1] For each set of parameters, 200 sets of 1000 practices were modelled.

[2] Values for SD (standard deviation), minimum and maximum represent median values for all runs of the simulation at each specification of practice variation and number of cancer cases.

[3] Correctly 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.

[4] Total 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.