Comparison of key practice characteristics between general practices in England and Wales and general practices in the QRESEARCH database

Report Title	**Comparison of key practice characteristics between general practices in England and Wales and general practices in the QRESEARCH database**
Document Name	**Characteristics of QRESEARCH practices (database version 8) v1.0.doc**
Document Version	**1.0**
Authors	**Julia Hippisley-Cox, Yana Vingradova, Carol Coupland, Mike Pringle**
Status	**Final Draft**
Submitted to	**Dave Roberts, NHS Information Centre**
Date Submitted	**20 Nov 2005**

1 Executive summary

This report compares data concerning patients registered with general practices in England using three data sources:

- The extract from the Attribution Data Set (ADS) supplied by the Department of Health (31st March 2004)
- The QMAS practice prevalence data (31 March 2005)
- Version 8 of the QRESEARCH database (31 March 2004 and 31 March 2005)

The comparison was commissioned in order to understand the extent to which QRESEARCH practices represent the overall national picture as found in the extract from the ADS and QMAS data sets. This information is needed to assess whether the results of analyses conducted on QRESEARCH are likely to be generalisable.

The key findings are:

- QRESEARCH is broadly comparable with the findings in the national databases and hence the results of analyses on QRESEARCH are likely to be generalisable.

- QRESEARCH has a good geographical spread through the English regions, with two exceptions: there has been an excess QRESEARCH recruitment in the East Midlands (reflecting the history of the QRESEARCH team in using East Midlands practices for research) and under recruitment in London.

- These differences in practice recruitment are also, as expected, found in the numbers of patients in each region available for analysis.

- QRESEARCH practices are slightly larger – have higher mean numbers of registered patients – than the full national picture would predict. This is likely to reflect the preparedness of larger practices to share their clinical databases.

- The age-sex distribution of people in QRESEARCH practices closely matches that in the national sample (extract from the ADS).

- Although there are some differences in the prevalence of chronic diseases in a comparison between QRESEARCH and QMAS, the differences are relatively small, with QRESEARCH showing slightly higher prevalence for all disease areas except cancer.

This work has confirmed earlier analyses which suggest that the characteristics of QRESEARCH's practices are biased by a differential recruitment of slightly larger practices, but the characteristics of the patients registered with QRESEARCH's practices closely reflects those of the overall national picture.

2 Table of contents

3 Tables

4 Figures

5 Purpose of document

This report compares data concerning patients registered with general practices in England using three data sources:

- The extract from the Attribution Data Set (ADS) supplied by the Department of Health (31st March 2004)
- The QMAS practice prevalence data (31 March 2005)
- Version 8 of the QRESEARCH database (31 March 2004 and 31 March 2005)

6 Aim

The aim of this report is to compare characteristics of practices in England and QRESEARCH to determine how representative QRESEARCH practices are compared with practices nationally. This information is needed to inform and interpret analyses undertaken on QRESEARCH for the NHS Information Centre and other organisations.

7 Methodology

7.1 Attribution Data Set

An extract of the Attribution Data Set was supplied by the Delivery Analysis Team, Department of Health. This is a practice level dataset which contains the following variable for each practice in England and Wales.

- Age sex bands (0-4 years, 5-14 years, 15-44 years, 45-64 years, 65-74 years, 75-84 years and 85 plus years)
- Total registered patients on 1st April 2004
- SMR under 65 - Standardised mortality ratio for those aged under 65.
- Local age structure (assumed to be from the 2001 Census) is applied to national age specific death rates, then standardised by national population.
- New registrations - to GP's list in Apr 2003 - Mar 2004 on the extract from the ADS
- Nursing and residential home patients - estimated by matching patients in the extract from the extract of the ADS with the same postcode as nursing and residential homes in the A-Z of Care Homes database (http://www.carehome.co.uk/disk.cfm) and scaling these figures to the number of beds in each nursing and residential home

7.2 QMAS data for practice prevalence rates

Practice level prevalence data for practices in England on 1st April 2005 was downloaded from the Information Centre website.

http://www.ic.nhs.uk/services/qof/data/

The source of QOF tables published by the Health and Social Care Information Centre is the Quality Management and Analysis System (QMAS), a national IT system that supports the QOF payment process. This new single national system ensures consistency in the calculation of quality achievement and prevalence. QMAS also gives GP practices, Primary Care Trusts and Strategic Health Authorities objective evidence and feedback on the quality of care delivered to patients.

7.3 QRESEARCH data

Comparative data for this preliminary report have been taken from the QRESEARCH database version 8 which contains data up to date until 1st Oct 2005.

The population data are from 31st March 2004 (to get the closest comparison with the extract from the extract of the ADS) and only includes practices which met the practice inclusion criteria for the GMS formula review. These criteria were a list size of at least one thousand registered patients and a consultation rate of at least one consultation per person per year.

The prevalence data are from 31st March 2005 (to get the closest comparison with QMAS data) and included all practices with complete data on that date in QRESEARCH.

8 Results

8.1 Study practices

Overall, there were 454 practices in England and Wales which met our inclusion criteria i.e. had complete data for April 2003 to March 2004 and had a list size of at least one thousand patients and a minimum consultation rate of one consultation per person per year.

8.2 Study population

There were 3,824,522 patients registered at any point during 1st April 2003 to 31st March 2004. Of these, 3,351,050 patients were registered on 31st March 2004.

There was a substantial turnover of patients registered during the 12 month period as shown in the table below. During the 12 month period, there were 33,727 deaths, 97,239 temporary residents and 319,435 new patient registrations.

Table 1 Number of patients by registration length and type between 1st April 2003 and 31st March 2004 in 454 QRESEARCH practices in England and Wales (database version 8)

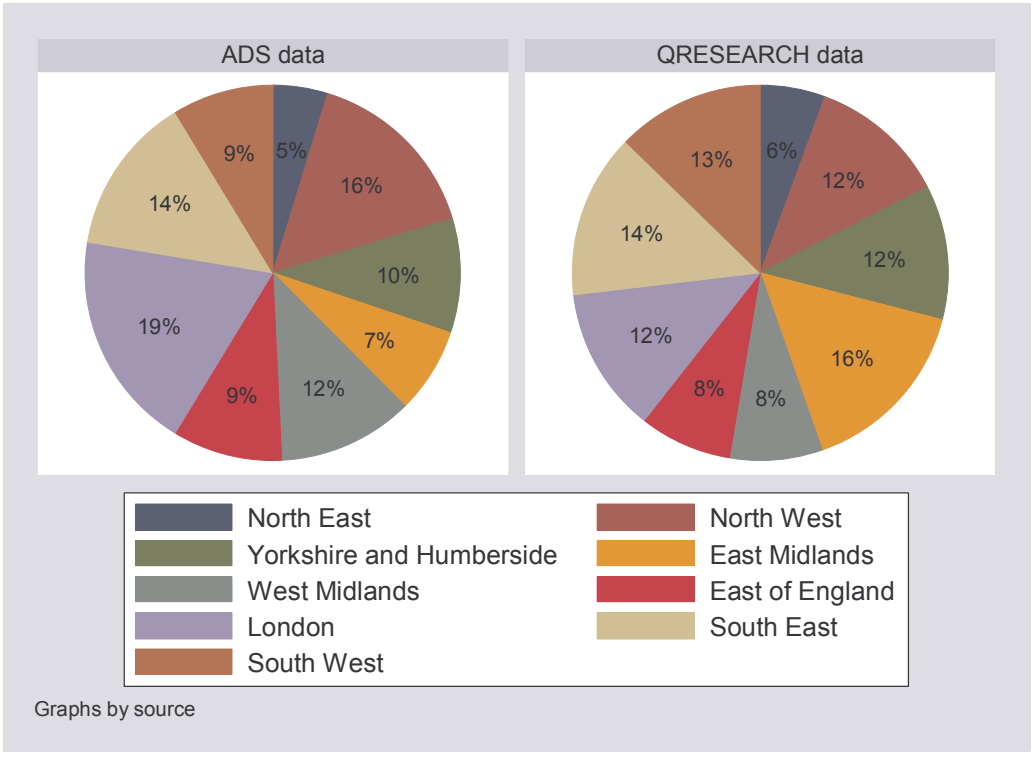
	Number of patients	Column % of patients registered at any time
Number of patients registered for 24 plus months on 1st April 2003	2,857,545	74.7
Number of patients registered for 12-23 months on 1st April 2003	247,169	6.5
Number of patients registered for less than 12 months on 1st April 2003	303,134	7.9
Number of new registrations during 1st April 2003 to 31st March 2004	319,435	8.4
Number of temporary residents during 1st April 2003 to 31st March 2004	97,239	2.5
Total	*3,824,522*	100

8.3 Geographical location

Table 1 (and the pie chart) shows the distribution of the 444 practices in England by government office region. There were 10 practices from Wales.

Some regions, North West, West Midlands, East of England and London are represented by larger proportions of practices in the extract from the Attribution Data Set (ADS), particularly in London, 18.8% vs. 11.9% in QRESEARCH.

Figure 1 Proportion of practices in England by geographical region in the ADS and in QRESEARCH on 31 March 2004



The other regions have a larger proportion of data in QRESEARCH: the most noticeable difference is in the East Midlands region, 7.4% in the extract of the Attribution Data Set vs. 15.1% in QRESEARCH. The over-representation in the East Midlands reflects where the original pilot work for QRESEARCH was done.

Table 2. Distribution of practices in England from QMAS and QRESEARCH (version 8) by Government Office Region on 31st March 2004

Name of region	Number of practices in QMAS	Col% for QMAS	Col % for practices in the ADS extract	Number of practices in QRESEARCH	Col % for QRESEARCH
North East	405	4.8	4.7	25	5.6
North West	1,321	15.6	15.6	49	11.0
Yorkshire and Humberside	822	9.7	9.9	52	11.7
East Midlands	639	7.5	7.4	67	15.1
West Midlands	991	11.7	11.7	35	7.9
East of England	807	9.5	9.5	38	8.6
London	1,592	18.8	18.9	53	11.9
South East	1,167	13.8	13.7	66	14.9
South West	742	8.7	8.7	59	13.3
Total	**8,486**	**100**	**100**	25	5.6

Table 3 shows the distribution of patients within each government office region. It is similar to the distribution of practices except for South East region: the proportion of patients is lower in QRESEARCH.

Table 3. Distribution of patients in England and QRESEARCH by Government Office Region on 31st March 2004

Name of region	Col % for patients in the ADS extract	Number of patients in QMAS	Col % for QMAS	Number of patients in QR	Col % for QR
North East	5.0	2,632,634	5.0	203,007	6.1
North West	13.6	7,176,170	13.6	324,354	9.7
Yorkshire and Humberside	10.1	5,261,058	10	396,419	11.8
East Midlands	8.2	4,357,892	8.2	483,766	14.4
West Midlands	10.6	5,591,065	10.6	263,429	7.9
East of England	10.8	5,728,270	10.8	307,218	9.2
London	15.8	8,322,132	15.8	397,213	11.9
South East	16.1	8,535,874	16.2	542,137	16.2
South West	9.8	5,228,489	9.9	433,507	12.9
Total	**100**	**52,833,584**	**100**	3,351,050	100.00

8.4 List size by region

The next table shows the median list size (with inter-quartile range) for practices in both data sets by government office region and overall as of 31 March 2004. The median list size for practices in QRESEARCH is larger than that for all practices over all and in every region. The largest difference in list size is found in London and the West Midlands.

Table 4 Median list size (IQR) for practices in England and QRESEARCH by region (version 8) for 31st March 2004.

Name of region	Median list size (IQR) in the Attribution Data Set	Median list size (IQR) in QMAS Data Set	Median list size (IQR) in QRESEARCH Data Set
North East	5,824 (3,234 – 8,825)	5,879 (3,401 – 9,130)	7,030 (4,662-9,439)
North West	4,533 (2,725 – 7,341)	4,493 (2,746 – 7,330)	5,771 (4,283-8,093)
Yorkshire and Humberside	5,697 (3,163 – 8,743)	5,582 (3,063 – 8,595)	7,005 (4,845-8,996)
East Midlands	5,945 (3,408 – 9,052)	5,993 (3,423 – 9,053)	5,709 (3,844-9,760)
West Midlands	4,559 (2,806 – 7,832)	4,558 (2,818 – 7,632)	6,844 (4,138-9,638)
East of England	6,545 (3,484 – 9,976)	6,387 (3,492 – 9,867)	8,117 (5,696-10,228)
London	4,384 (2,831 – 6,905)	4,356 (2,806 – 6,773)	6,717 (5,217-9,223)
South East	6,861 (3,804 – 10,028)	6,747 (3,690 – 9,911)	7,965 (5,314-9,735)
South West	6,543 (4,066 – 9,372)	6,472 (4,056 – 9,249)	6,615 (4,577-10,475)
Total*	5,394 (3,118 – 8,529)	5,335 (3,092 – 8,453)	6,816 (4725-9691)

8.5 Age-sex distribution

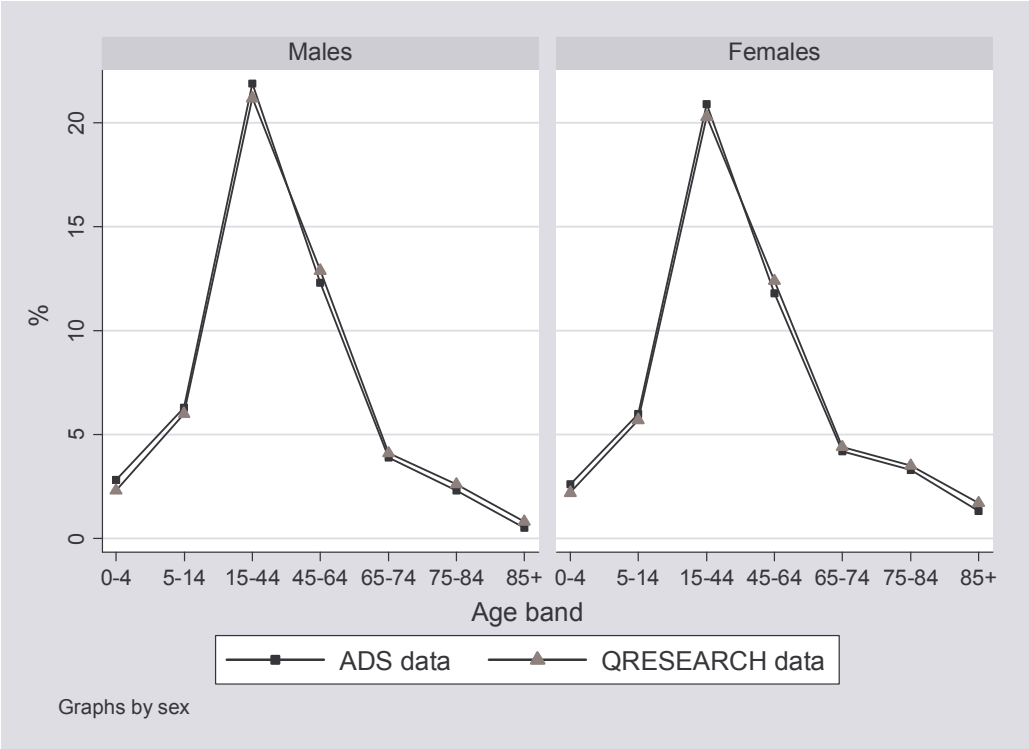
The next table shows the age-sex distribution of patients in England and in QRESEARCH for patients registered on 31st March 2004. The population in the extract from the Attribution Data Set is slightly younger than the population in QRESEARCH: 60.5% of people younger than 45 years in the extract from the Attribution Data Set data vs. 57.7% of the same age in QRESEARCH and consequently 15.5% of older than 65 years in the extract from the Attribution Data Set vs. 17.1% in QRESEARCH.

Table 5 Age-sex distribution of registered patients in England and QRESEARCH (version 8) on 31st March 2004

sex	ageband	ADS Col %	QRESEARCH (1st April 2004) Col %
M	0-4	2.8	2.4
M	5-15	6.3	6.1
M	15-44	21.9	21.3
M	45-64	12.3	12.7
M	65-74	3.9	4.1
M	75-84	2.3	2.5
M	85 plus	0.5	0.7
M	**total**	**50**	**50**
F	0-4	2.6	2.3
F	5-15	6.0	5.8
F	15-44	20.9	20.6
F	45-64	11.8	12.2
F	65-74	4.2	4.4
F	75-84	3.3	3.5
F	85 plus	1.3	1.6
F	**total**	**50**	**50**
	overall total	**100**	**100.0**

The following figure shows the age distribution in the extract from the ADS and QRESEARCH constructed for males and females separately. The lines are very close, however, a very slightly bigger proportion of older people can be seen in QRESEARCH data, which is compensated by a very smaller proportion of younger people.

Figure 2. Proportions for different age bands in the extract from the ADS and QRESEARCH



The next table shows the number of patients in each age-sex band of eligible practices in England and Wales by region in the QRESEARCH database on 31st March 2004.

Table 6. Age-sex distribution of registered patients in Wales and in England by Government Office Region QRESEARCH on 31st March 2004.

AGE_BAND	NE	NW	YH	EM	WM	EE	L	SE	SW
total	**203,007**	**324,354**	**396,419**	**483,766**	**263,429**	**307,218**	**397,213**	**542,137**	**433,507**
total females	**102,880**	**162,396**	**200,193**	**243,746**	**133,618**	**153,588**	**199,288**	**271,040**	**218,283**
0-4	4,032	7,204	8,478	10,269	5,743	6,993	12,237	12,172	8,514
5-14	10,767	19,389	23,346	28,585	15,803	17,829	21,867	32,282	24,304
15-44	45,258	67,103	76,233	96,113	49,348	61,818	103,818	108,636	80,718
45-65	23,878	38,919	51,172	61,681	34,342	37,615	36,423	67,670	57,538
65-74	9,349	14,086	19,216	21,829	13,378	13,749	11,471	22,497	20,820
74-85	6,872	11,023	15,378	17,342	10,565	10,646	9,178	18,637	17,697
85 plus	2,724	4,672	6,370	7,927	4,439	4,938	4,294	9,146	8,692
total males	**100,127**	**161,958**	**196,226**	**240,020**	**129,811**	**153,630**	**197,925**	**271,097**	**215,224**
0-4	4,311	7,497	8,878	10,619	6,112	7,378	12,859	12,486	8,925
5-14	11,038	20,757	24,056	29,869	16,314	18,527	22,762	34,298	25,705
15-44	45,746	70,790	78,880	99,687	51,230	64,728	103,771	114,433	85,230
45-65	24,765	40,258	52,855	62,867	34,564	39,377	39,886	71,102	58,875
65-74	8,259	13,048	17,824	20,626	12,188	13,219	10,592	21,412	19,703
74-85	4,885	7,762	11,031	12,797	7,534	8,119	6,212	13,201	12,878
85 plus	1,123	1,846	2,702	3,555	1,869	2,282	1,843	4,165	3,908

8.6 Distribution of patients by deprivation score

The following table shows the number and percentage of patients in each quintile of Townsend score. The cut offs for the Townsend quintiles were based on the national distribution of Townsend scores for the output area from the 2001 census. Townsend scores were observed for 96% of patients. There was a slight over representation of patients from the most affluent quintile in QRESEARCH – 23% of patients registered on 31st March 2004 were allocated to the most affluent quintile and 19% were allocated the most deprived quintile (compared with 20% in each group nationally).

Table 7 Distribution of patients by quintile of Townsend score (quintiles defined using national data for output area from the 2001 census)

	Number of patients registered on 31st March 2004 in each Townsend quintile	Column % of total
Quintile 1 (Most affluent)	762,379	23.2
Quintile 2	678,783	20.6
Quintile 3	631,650	19.2
Quintile 4	584,281	17.8
Quintile 5 (most deprived)	633,226	19.2

8.7 Prevalence of chronic diseases

The next table shows the median practice prevalence (and inter-quartile range) for each of the diseases in the new GP contract comparing data from QMAS for 1st April 2005 with that from QRESEARCH for the same date. The prevalence definitions have been determined using the nationally agreed set of computer recorded Read Codes for the new GMS contract.

Table 8 Median practice prevalence of diseases in QMAS & QRESEARCH (v8 1st April 2005)

	QMAS data			QRESEARCH data		
	No practices	Median (IQR) prevalence per 1000	Range (min-max)	No practices	Median (IQR) prevalence per 1000	Range (min-max)
coronary heart disease	8486	35 (26 – 44)	(0-346)	488	39 (29 – 47)	(0.2-97)
left ventricular dysfunction	8486	4 (2 – 6)	(0-44)	488	5 (3 - 7)	(0-16)
stroke	8486	14 (9 – 18)	(0-250)	488	17 (12- 22)	(0.2-83)
hypertension	8486	112 (90 – 133)	(0-513)	488	124 (102 – 147)	(0-278)
diabetes	8486	33 (28 – 39)	(0-165)	488	35 (29 – 40)	(0-75)
COPD	8486	12 (8 – 18)	(0-135)	488	16 (11 - 22)	(0-83)
epilepsy	8486	6 (4 – 7)	(0-55)	488	6 (5 – 7)	(0-25)
mental health problems	8486	5 (3 – 7)	(0-156)	488	5 (3 – 7)	(0-53)
thyroid	8486	21 (15 – 27)	(0-173)	488	21 (14 – 26)	(0-53)
asthma	8486	57 (47 – 67)	(0-224)	488	67 (57 – 77)	(0-163)
Cancer	8486	5 (3 – 6)	(0-26)	488	2.2 (1 - 4)	(0-83)

The prevalence of seven diseases is slightly higher in QRESEARCH (coronary heart disease, left ventricular dysfunction, stroke, hypertension, diabetes, COPD and asthma). This may be because these diseases are associated with old age and QRESEARCH has a slightly older population. The prevalence for three diseases (epilepsy, mental health problems and hypothyrodism) were identical between QRESEARCH and QMAS. QRESEARCH had a slightly lower prevalence of cancer.

The next charts are histograms of the prevalence of the diseases for practices in England. For better presentation, the highest 8 values of prevalence were removed when drawing each histogram. The number of such observations did not exceed 0.1% of the sample and therefore did not affect the shape of the histogram.

Figure 3. Prevalence of coronary heart disease

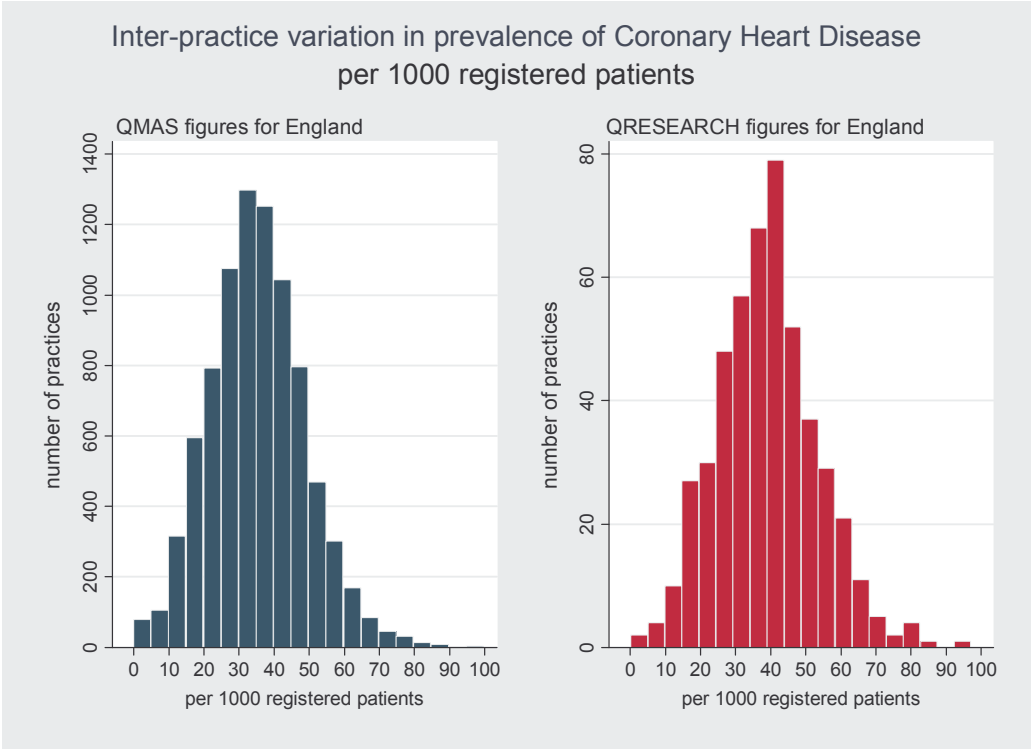


Figure 4. Prevalence of left ventricular dysfunction

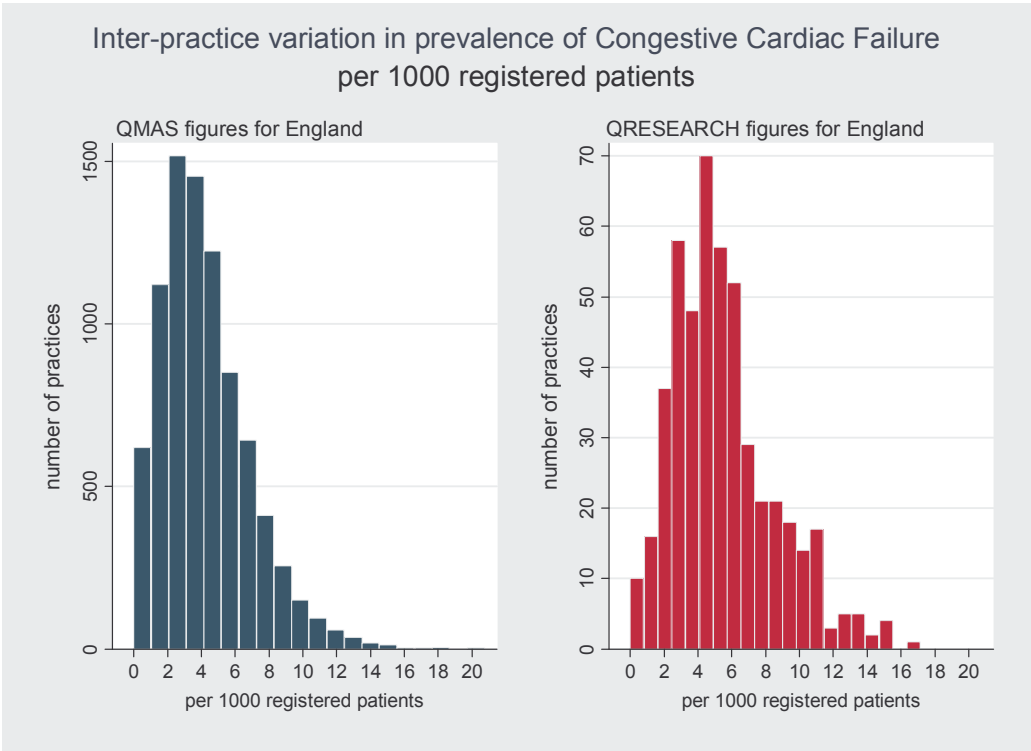


Figure 5. Prevalence of stroke or TIA

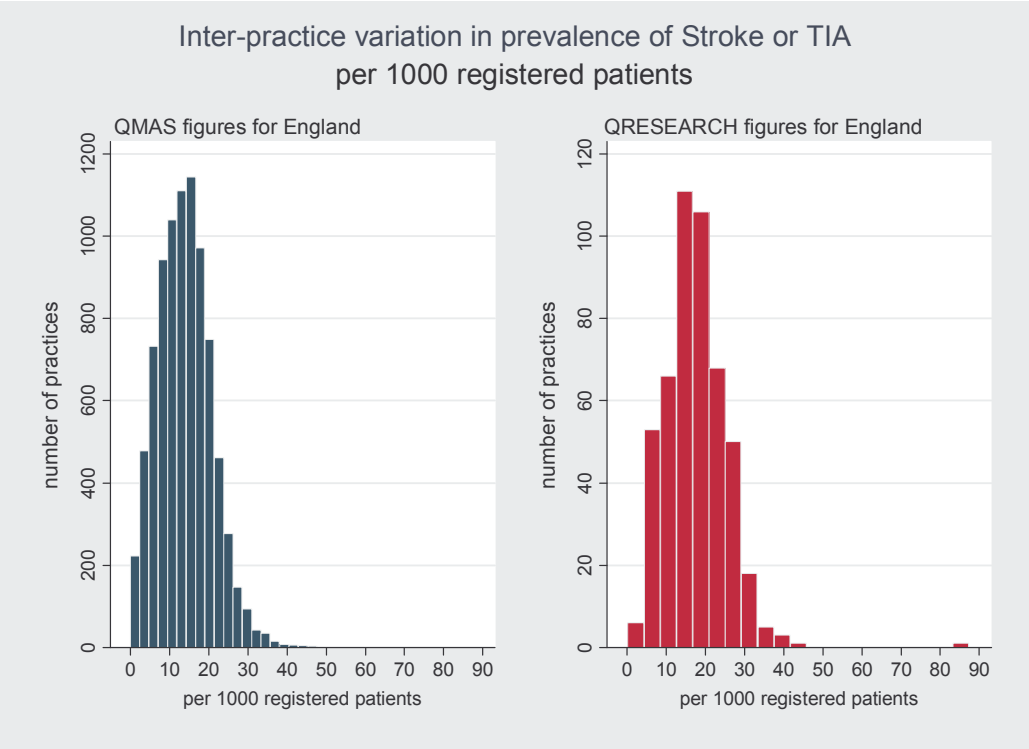


Figure 6. Prevalence of hypertension

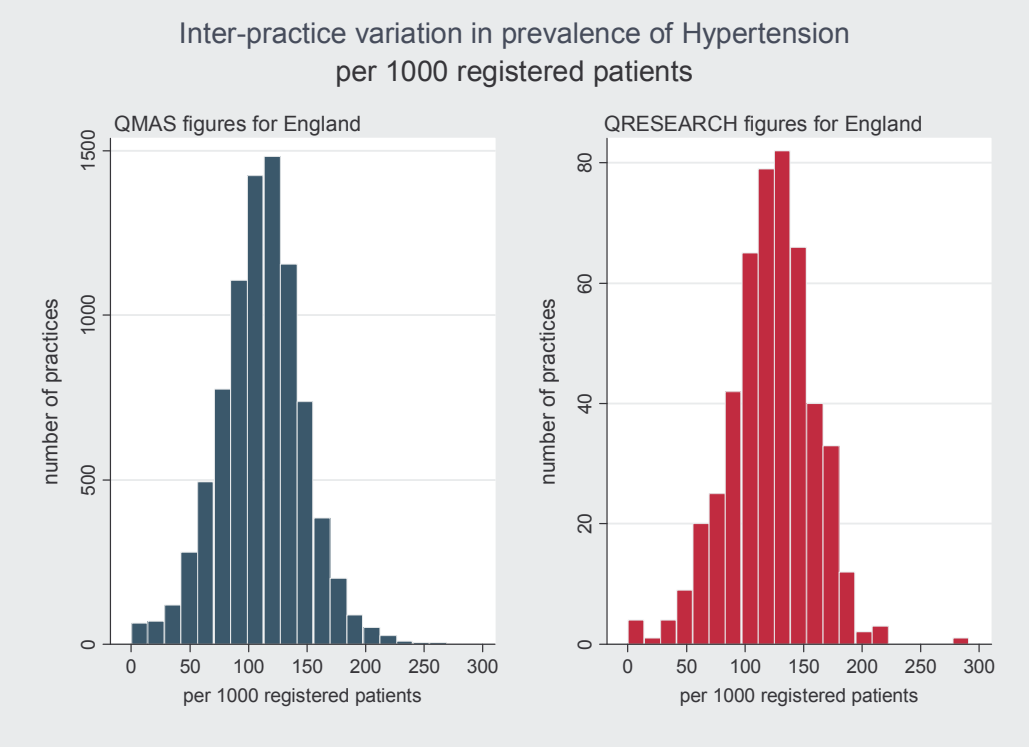


Figure 7. Prevalence of diabetes in patients over 15

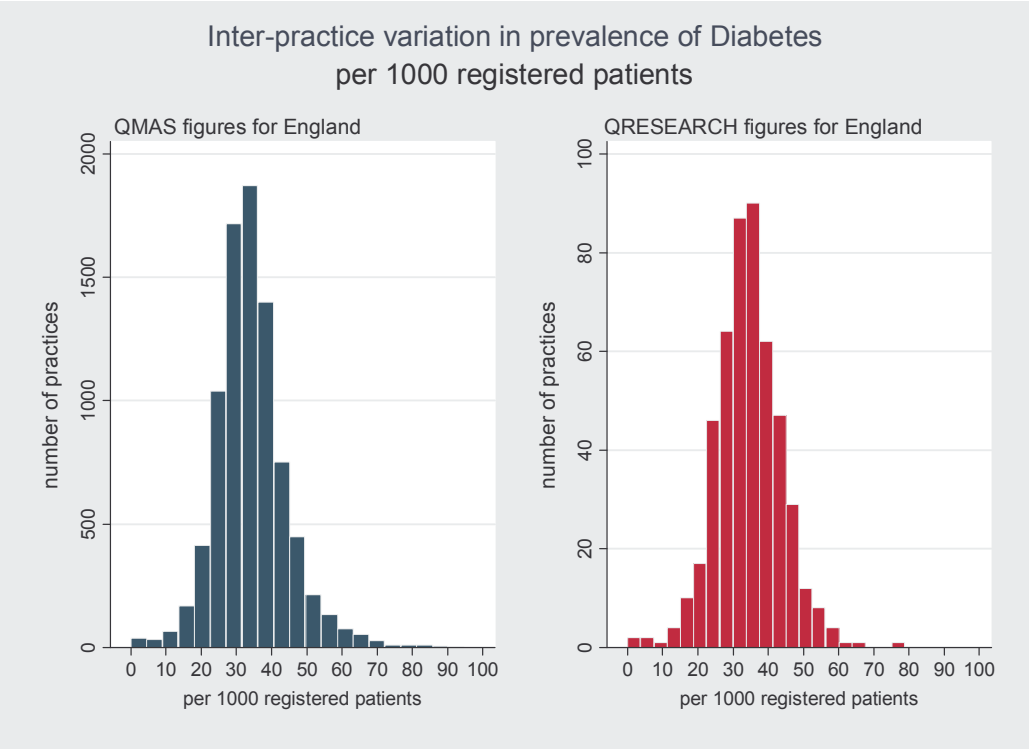
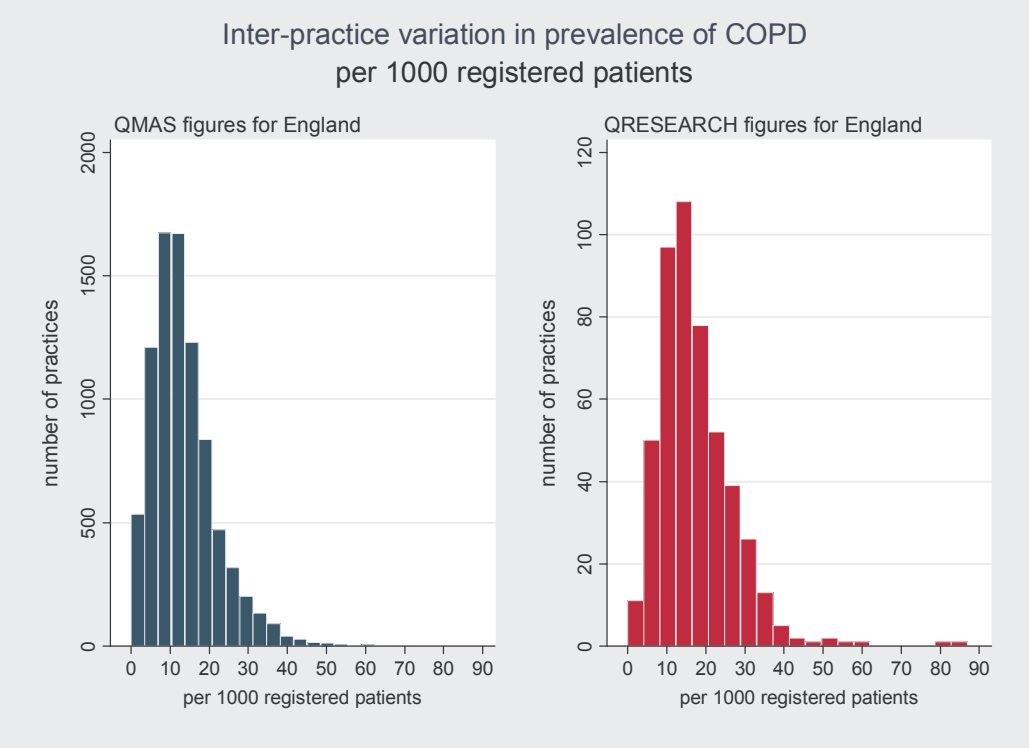


Figure 8. Prevalence of Chronic Obstructive Pulmonary Disease



Authors: Julia Hippisley-Cox, Yana Vinogradova, Carol Coupland, Mike Pringle

Figure 9. Prevalence of treated epilepsy in patients over 15

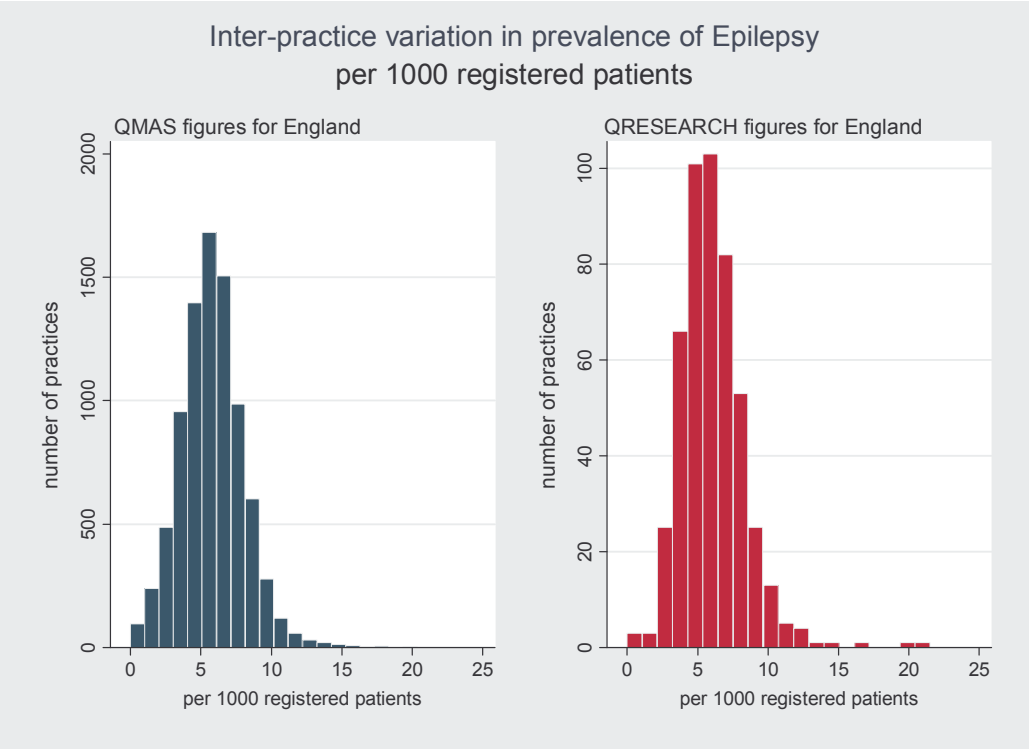


Figure 10. Prevalence of mental health problems

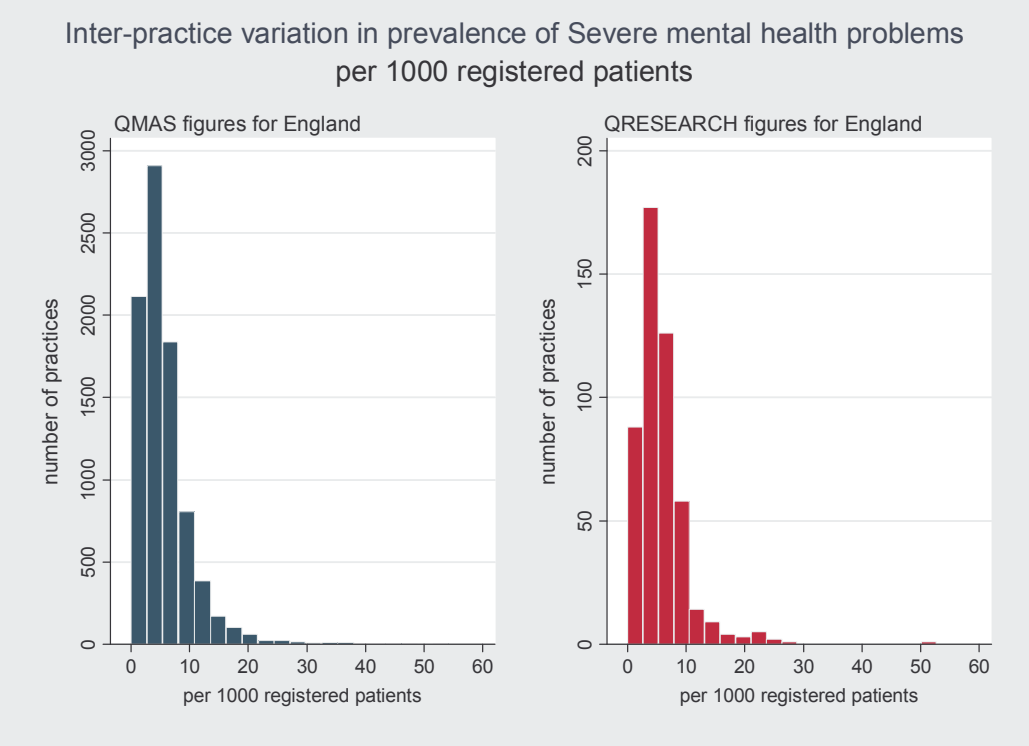


Figure 11. Prevalence of treated hypothyroidism

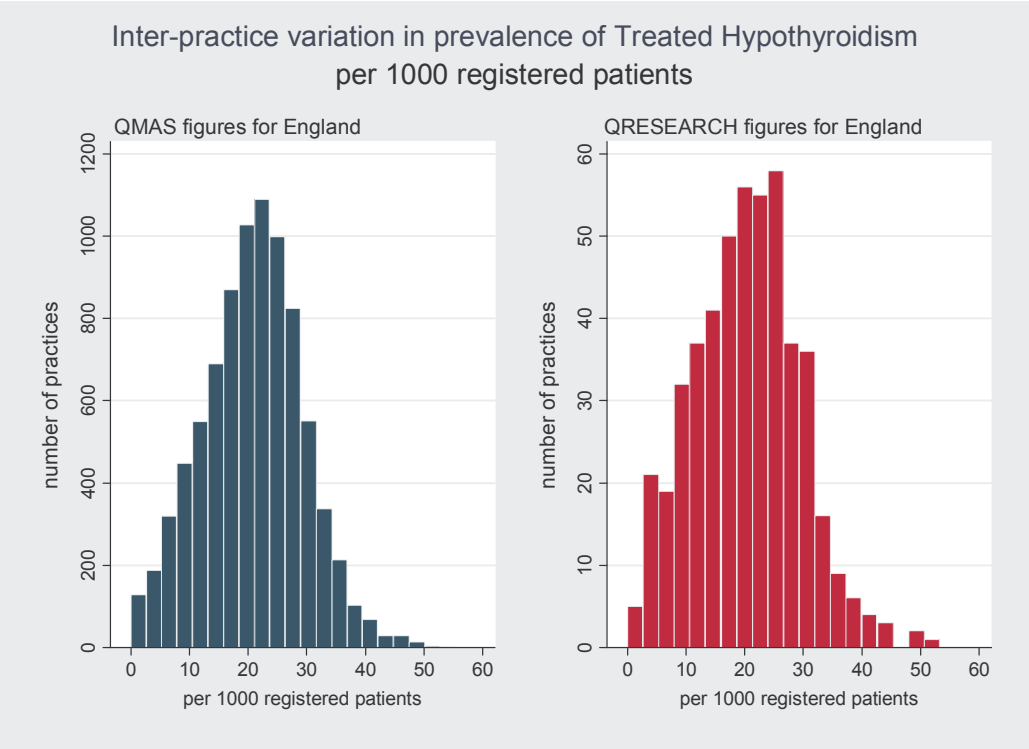


Figure 12. Prevalence of treated asthma

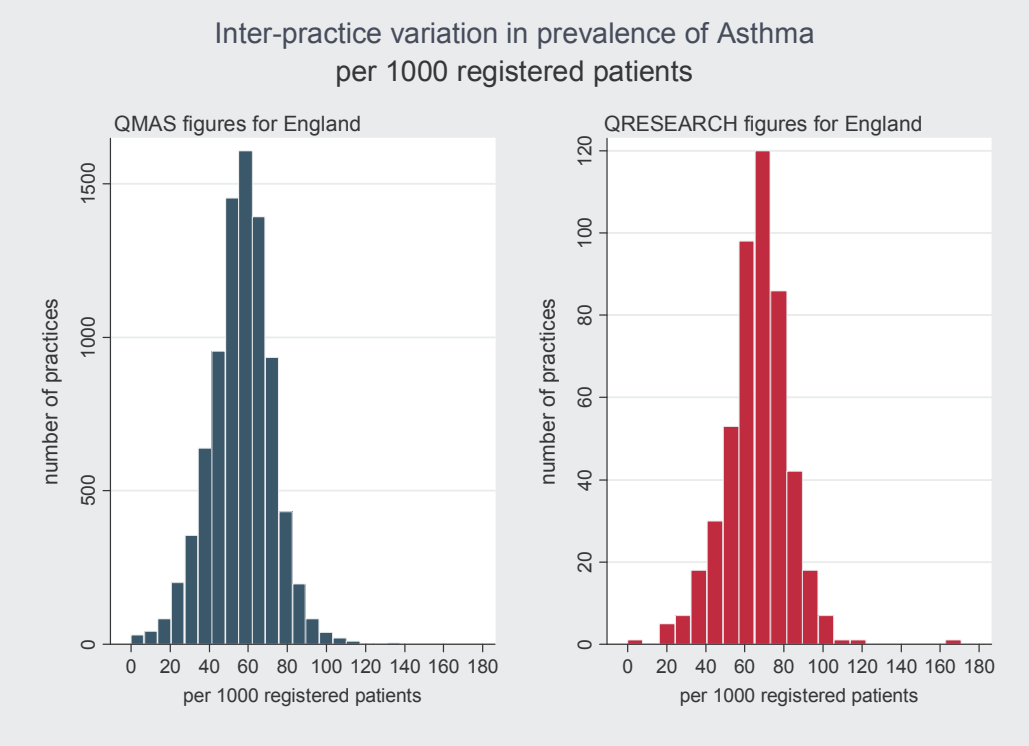


Figure 13. Prevalence of newly diagnosed cancer (after April 2003)

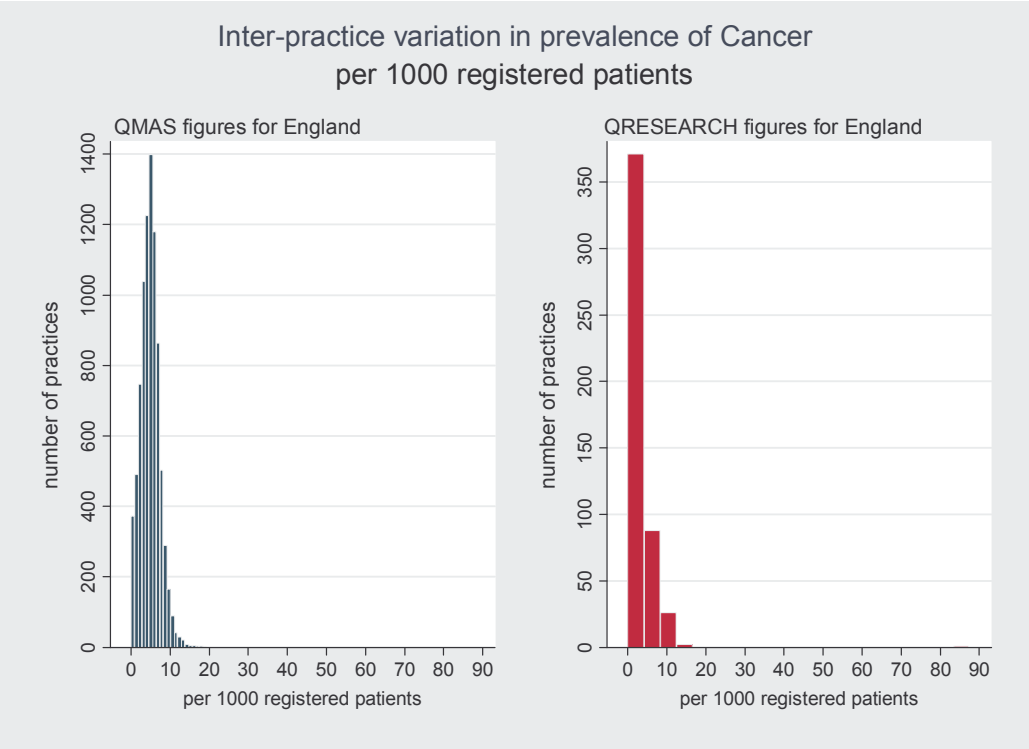
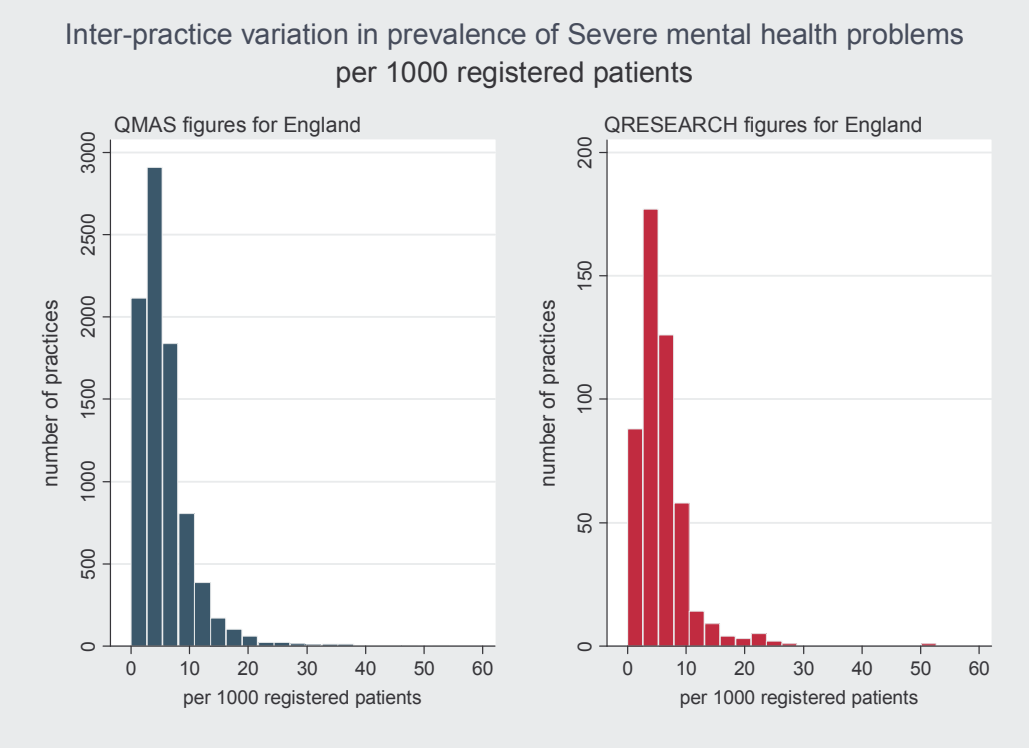


Figure 14. Prevalence of severe mental health problems



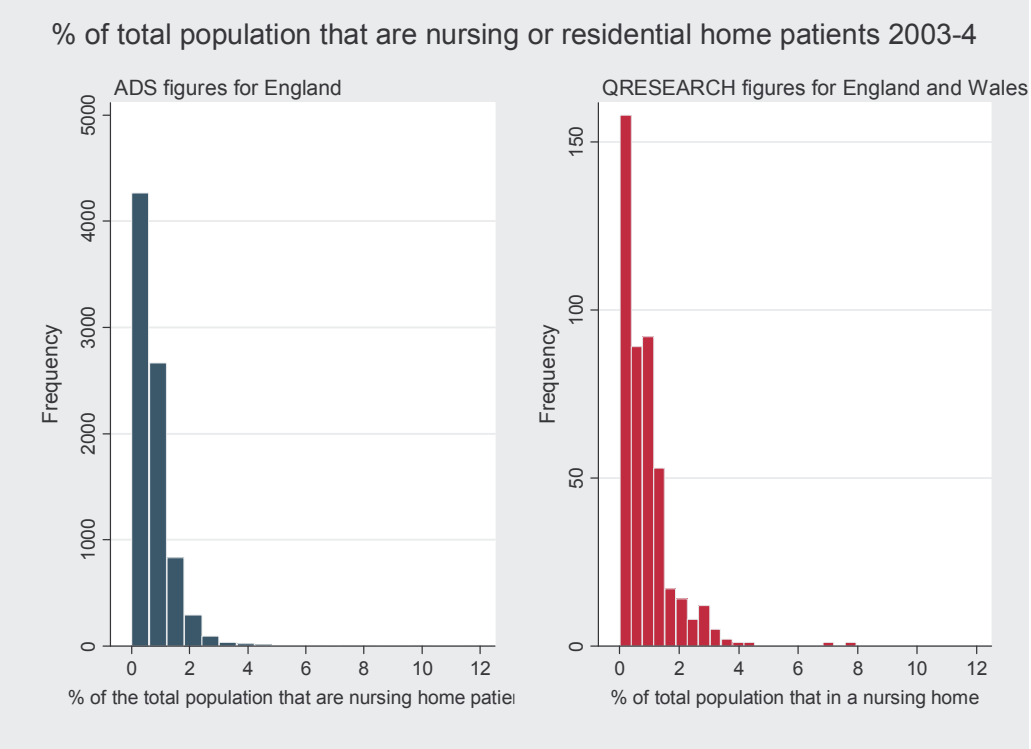
Authors: Julia Hippisley-Cox, Yana Vinogradova, Carol Coupland, Mike Pringle

8.8 Nursing/residential homes

The median practice percentage of patients estimated to be in a nursing home or residential home in the extract from the ADS was 0.6% (IQR 0.3% to 1%).

In QRESEARCH, there is not an exact equivalent to the definition for nursing/residential home used in the extract from the ADS. Therefore we explored developing a proxy measure for nursing homes based on the unique anonymised household ID. Our proxy measure was based on a multiple occupancy household with four or more patients aged 65 years or over. Whilst there was a good correspondence in the median percentage of the population in a nursing/residential home using this proxy (0.7% (IQR 0.2% to 1.2%) and the distribution looks similar (as shown in the next graph), the methodology was not sufficiently specific. The validation study showed that this methodology over identified nursing/residential homes and hence its use is not recommended.

Figure 15. Percentage of the total population that are nursing or residential home patients



9 Summary of findings

This report compares data concerning patients registered with general practices in England using three data sources:

- The extract from the Attribution Data Set (ADS) supplied by the Department of Health (31st March 2004)
- The QMAS practice prevalence data (31st March 2005)
- Version 8 of the QRESEARCH database (31st March 2004 and 31st March 2005)

The comparison was commissioned in order to understand the extent to which QRESEARCH practices represent the overall national picture as found in the extract from the ADS and QMAS data sets.

The key findings are:

- QRESEARCH is broadly comparable with the findings in the national databases

- QRESEARCH has a good geographical spread through the English regions, with two exceptions: there has been an excess QRESEARCH recruitment in the East Midlands (reflecting the history of the QRESEARCH team in using East Midlands practices for research) and under recruitment in London

- These differences in practice recruitment are also, as expected, found in the numbers of patients in each region available for analysis

- QRESEARCH practices are slightly larger – have higher mean numbers of registered patients – than the full national picture would predict. This is likely to reflect the preparedness of larger practices to share their clinical databases

- The age-sex distribution of people in QRESEARCH practices closely matches that in the national sample (extract from the ADS)

- Although there are some differences in the prevalence of chronic diseases in a comparison between QRESEARCH and QMAS, the differences are relatively small, with QRESEARCH showing slightly higher prevalence in all disease areas except cancer

This work has confirmed earlier analyses which suggest that the characteristics of QRESEARCH's practices are biased by a differential recruitment of slightly larger practices, but the characteristics of the patients registered with QRESEARCH's practices closely reflects those of the overall national picture.