GIG CYMRU NHS WALES
Arsyllfa Iechyd Cyhoeddus Cymru
Public Health Wales Observatory

CHD prevalence modelling by GP cluster

Date: 4th June 2015	**Version:** 2a

Publication/ Distribution:

- Public Health Wales Observatory Website (Intranet)
- Public (Internet)

Purpose and Summary of Document:

This document describes the modelling of coronary heart disease (CHD) prevalence by GP cluster, originally produced for the Inverse Care Law Project areas but expanded here for the whole of Wales.

The Welsh Health Survey was used to model the prevalence of self reported treated heart disease in GP cluster populations. This modelled prevalence was compared to the diagnosed CHD prevalence in GP cluster populations based on the QOF CHD register.

The modelled self reported treated heart disease prevalence was higher than the QOF CHD register in every GP cluster. An additional project would be required to investigate this in more detail.

However, these results may be of use to GP Clusters as a starting point for any work to investigate CHD case ascertainment.

Contents

Key Messages

- This analysis compared self reported treated CHD prevalence with recorded QOF CHD register prevalence.
- The self reported prevalence, modelled from Welsh Health Survey data, was higher in every GP cluster than the QOF register prevalence.
- Although the model over-estimated total self reported prevalence, there was substantial variation in the gap between the two prevalence measures.
- Whilst this approach does not identify undiagnosed cases, it may prove a useful starting point for case ascertainment exercises.

1 Background

1.1 Inverse Care Law Project

The Public Health Wales Observatory (PHWO) was asked to produce modelled coronary heart disease (CHD) prevalence estimates as part of the Inverse Care Law Project in Cwm Taf and Aneurin Bevan University Health Boards. Similar estimates had previously been published for England by the Association of Public Health Observatories (APHO), who commissioned the work from Imperial College London (Walford *et al*, 2011). Those APHO estimates were used in the work presented by Professor Chris Bentley at the 1000lives+ workshop in March 2013.

This PHWO report is an expanded version, to include all GP clusters in Wales, of the report delivered to the Inverse Care Law Project Board.

1.2 Modelled prevalence

Whilst the modelled prevalence estimates presented in this PHWO report were produced using the method described in the APHO report (Walford *et al*, 2011) the results must be viewed with caution. The fit of the model has been tested but its results have not been validated against an independent source of self reported treated heart disease prevalence.

When the national estimate from the model was compared against the national estimate of the source upon which it was based, namely the Welsh Health Survey (WHS), it was found that the model over estimated the prevalence by around eight per cent. Further investigation revealed the same phenomenon in the work by Walford *et al* (2011), though no mention of it was made in their APHO report, nor was any attempt to account for it.

This method for producing small area prevalence estimates works best when detailed population characteristics for the small areas are known and available. As most individual characteristics of registered populations are unknown, many potentially informative factors had to be excluded from both this and the APHO model.

In summary, the modelled prevalence estimates presented in this report should not be assumed to be a more accurate representation of true population prevalence than the QOF register. However, the difference between the two may help inform local case ascertainment investigations.

2 Project definition

2.1 Project aims

The aim of this project was to produce CHD prevalence estimates, by General Practice (GP) cluster, using a method based on self reported health survey data, and to compare those estimates with the QOF CHD register. In doing so, the results of this project may be a useful starting point for local investigations of case ascertainment.

2.2 Project objectives

- To identify and use a method for modelling CHD prevalence.
- To compare the modelled prevalence with the QOF CHD register.
- To conduct the necessary analytical work and ensure its quality.
- To produce a report detailing the method and results.

2.3 Project scope

The scope of this project was limited to the production of modelled prevalence estimates for CHD only and by GP cluster only, not for other conditions or by any other administrative or organisational geography.

3 Measuring CHD prevalence

Within this project the prevalence of CHD has been considered in two ways, both of which have potential limitations.

The basis of this model is self reported prevalence of treatment for heart attack and/or angina from the WHS, the original APHO model having used the Health Survey for England. The exact questions from the WHS are detailed in appendix A and are almost identical to those used in the APHO model. It is important to note that this is not a measure of undiagnosed CHD, in fact quite the opposite in that, by definition, it has, or is, being treated and hence it is a measure of diagnosed CHD. It is also important to remember that the WHS does not include any clinical measurement or confirmation, and that it is liable to various biases including:

- Response rate (79% in the period used here)
- Exclusion of care home residents
- Problems with self reporting health status

The basis for the comparison of the modelled prevalence is the QOF CHD register. For a patient to be on the QOF CHD register there must be at

least one of the specified codes in their electronic health record. The full set of codes can be seen in Appendix B. It is important to remember that, in general, prevalence based on clinical diagnosis depends upon people presenting with such symptoms and hence may not capture some cases.

Other methods for modelling prevalence, not used here, range from crude applications of rates in higher level geographies to lower level population to complex Bayesian regression models, though caution is always required, summed up by the eminent statistician George Box who famously proposed *"all models are wrong, but some are useful"* (Box & Draper, p424).

4 Method

4.1 Data sources

The WHS was used as the data source for the CHD prevalence model. The outcome of interest was defined as "ever been treated for heart attack" and/or "currently being treated for angina". A variable to represent the outcome of interest was created using the respective variables (hrtatbi and angbi) in the WHS dataset.

The cluster populations were extracted from the Welsh Demographic Service database of General Practice registrations. Smoking prevalence in each combination of age group, sex and deprivation fifth was attributed from the equivalent all Wales fractions derived from the WHS.

4.2 Model construction

The variables for consideration in the CHD prevalence model were:

- age group (10 year age groups from 25 to 75+),
- sex,
- Welsh Index of Multiple Deprivation (WIMD) fifth,
- smoking status and
- ethnic group

since these are the known risk factors for CHD for which population estimates at a local level could be obtained. Data for 16 to 24 year olds were excluded from the analysis since it was found that less than 0.2% of the self reported treated heart disease cases were in this age range.

Stata12 software was used to assess the effect of each risk factor on the outcome of interest and was also used to construct the CHD prevalence model using logistic regression, taking into account the survey design of the WHS dataset.

Effect modification or interaction was assessed for each combination of the risk factors using the Wald test. The goodness of fit test for design-based

logistic regression models was used to assess the fit of the model when considering the inclusion of borderline interaction variables.

4.3 Internal validation

The performance of the model was assessed by validating its ability to predict the response for each subject. Due to the nature of the WHS dataset neither a classification table nor Receiver Operating Characteristic (ROC) curve could be produced directly from the validation sample. An alternative approach was taken to use forced ordinary logistic regression methods as recommended by Hosmer *et al*, 2013. A ROC curve was therefore produced using these methods and then examined to determine the models discrimination in the validation sample.

4.4 External validation

The model has not been subject to any formal external validation and hence its results must be viewed with caution. External validation is beyond the scope of this project. However, the results were compared to those published by APHO and are entirely consistent with them.

It was observed in both these and the APHO results that the model over estimated prevalence in terms of a comparison between the modelled total and the total obtained from the respective health surveys. Further investigation indicated that the degree of over or under modelling varied geographically (Appendix C). To illustrate the effect that this may have had on the cluster prevalences, the modelled results were adjusted by a factor representing the respective local authority difference (Appendices D & E). However, these adjusted results should not be presumed to be more accurate than the original modelled estimates.

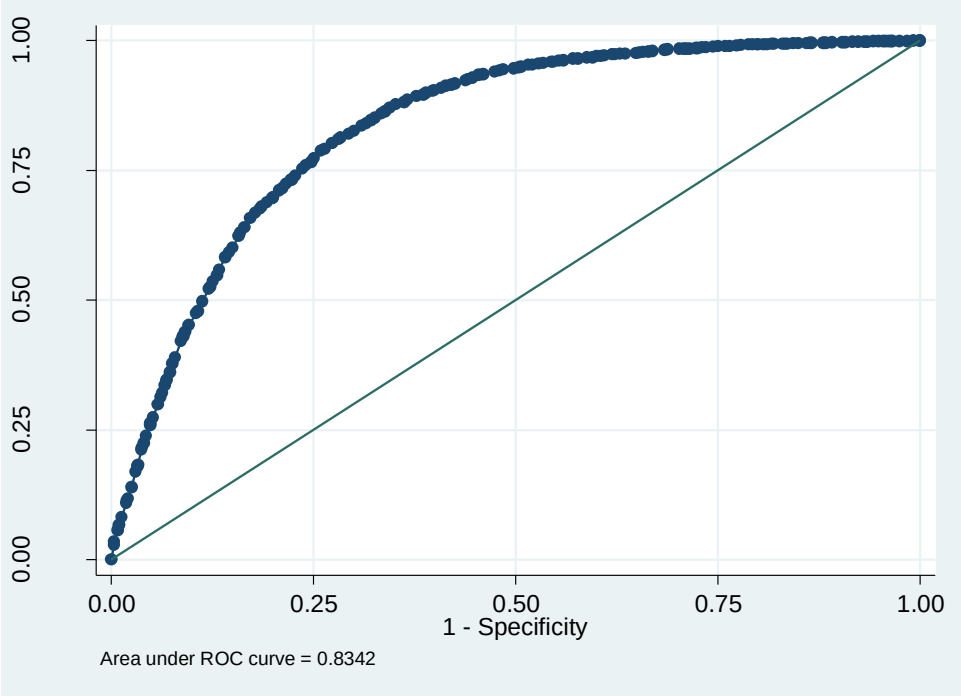
5 Results

All of the risk factors except ethnic group ($p>0.05$) were associated with CHD in our dataset. As such ethnic group was excluded from any further analysis and further examination suggested that the lack of association could be due to over 95% of respondents being in the white ethnic group.

The assessment of interactions showed a possible interaction between smoking and sex. However it was decided to exclude this from the final model since it was only significant for one level of the interaction and its inclusion reduced the goodness of fit of the model.

The results of the Pearson goodness-of-fit test showed evidence of a good fit of the final model to the data ($p=0.33$) and the ROC curve showed that the model discriminated well in the validation sample (area under the ROC curve = 0.83, figure 1).

Figure 1. ROC curve showing the final models discrimination in the validation sample



The results of the final model can be found in Table 1.

The odds ratios were applied to the model for each combination of age group, sex, smoking status and level of deprivation to determine the odds in each strata. The prevalence in each strata was then derived from the odds using the formula: prevalence =odds/1+odds.

The estimated prevalence can then be applied to the corresponding strata in populations and summed to the appropriate level to determine the number of subjects expected to have CHD (see section 6).

Table 1. Odds ratios for the final model

Risk Factor	Odds Ratio	Std. Error	t	P>t	95% Confidence Interval	
Age 25-34	1.000					
Age 35-44	2.511	0.81	2.86	0.00	1.34	4.72
Age 45-54	7.669	2.28	6.85	0.00	4.28	13.74
Age 55-64	22.489	6.51	10.75	0.00	12.75	39.68
Age 65-74	53.360	15.41	13.77	0.00	30.29	93.99
Age 75+	107.857	31.17	16.20	0.00	61.21	190.05
Never smoked	1.000					
Ex-smoker	1.570	0.08	9.05	0.00	1.42	1.73
Current smoker	1.544	0.10	6.55	0.00	1.36	1.76
Female sex	1.000					
Male sex	1.763	0.08	12.80	0.00	1.62	1.92
WIMD quintile 1	1.000					
WIMD quintile 2	1.117	0.08	1.53	0.13	0.97	1.29
WIMD quintile 3	1.266	0.09	3.34	0.00	1.10	1.45
WIMD quintile 4	1.462	0.11	5.21	0.00	1.27	1.69
WIMD quintile 5	2.266	0.16	11.38	0.00	1.97	2.61
Constant	0.001	0.00	-22.51	0.00	0.00	0.00

6 Application of results to GP clusters

The results obtained from the model were applied to populations representing each combination of age, sex, deprivation fifth and smoking status within each cluster to give strata specific prevalence estimates. Summing those strata specific prevalence estimates gave cluster level prevalence estimates (Table 2).

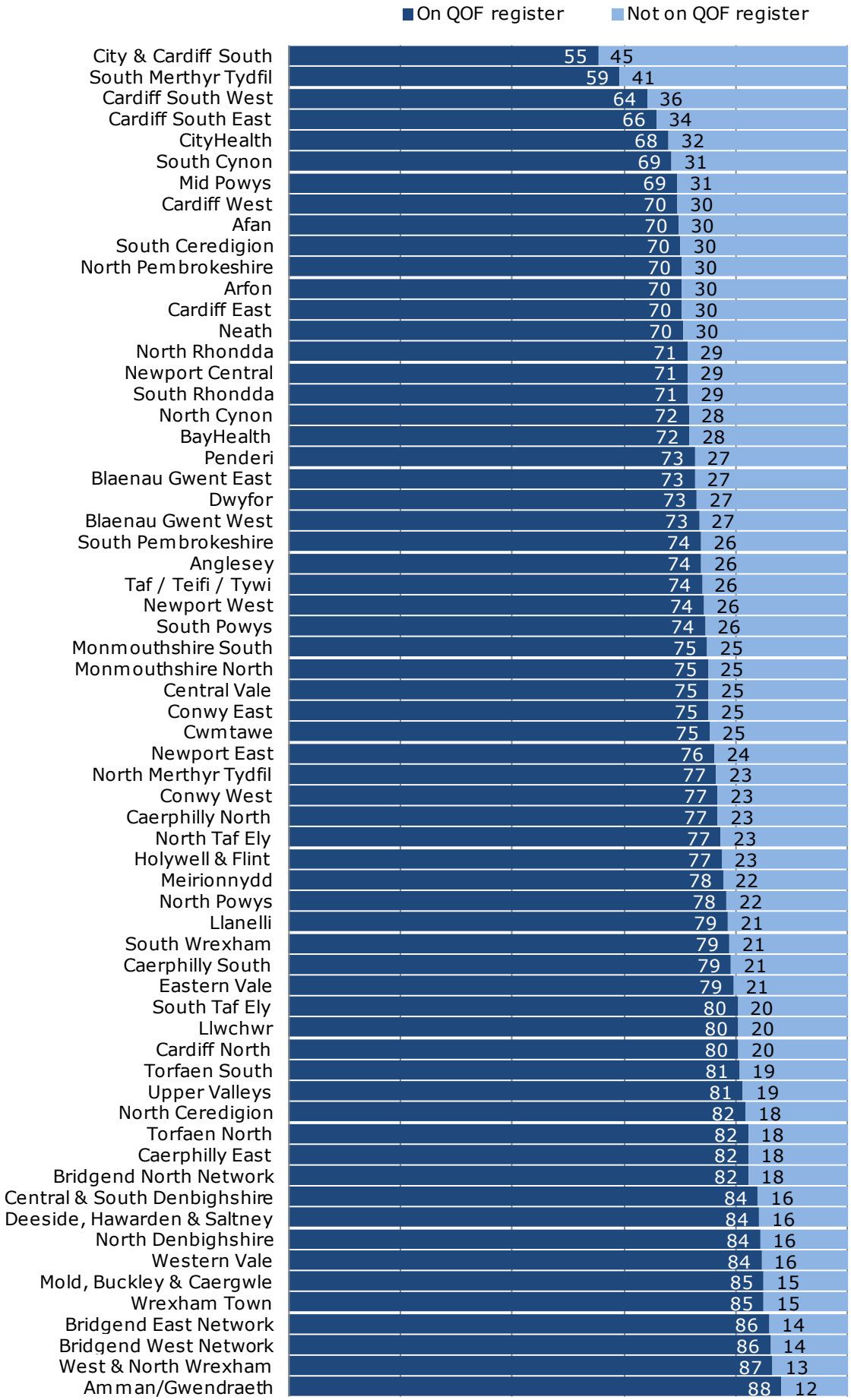
It can be seen that, for every cluster, the model estimated higher prevalence than the QOF register reported. The QOF register prevalence as a per cent of the modelled prevalence for each practice is shown in figure 2.

For example, in City and Cardiff South, the modelled prevalence estimate was 1353 and the QOF register prevalence was 748. Therefore the QOF register prevalence as a per cent of the modelled prevalence was 748/1353x100=55. It follows then that 100-55=45 per cent of the modelled prevalence could not be on the QOF register.

Table 2. All Wales modelled prevalence and QOF CHD register counts

	Prevalence			Prevalence	
cluster	**modelled**	**QOF**	**cluster**	**modelled**	**QOF**
Afan	3269	2278	Mid Powys	1775	1233
Amman/Gwendraeth	3342	2946	Mold, Buckley & Caergwle	2268	1927
Anglesey	4026	2972	Monmouthshire North	2812	2107
Arfon	2907	2044	Monmouthshire South	2237	1674
BayHealth	3553	2547	Neath	3275	2309
Blaenau Gwent East	2370	1720	Newport Central	2280	1627
Blaenau Gwent West	2122	1556	Newport East	2471	1880
Bridgend East Network	3151	2706	Newport West	2655	1971
Bridgend North Network	3022	2484	North Ceredigion	2049	1674
Bridgend West Network	2041	1757	North Cynon	2376	1699
Caerphilly East	2912	2393	North Denbighshire	3655	3082
Caerphilly North	4003	3072	North Merthyr Tydfil	1861	1424
Caerphilly South	2763	2186	North Pembrokeshire	3561	2502
Cardiff East	2510	1766	North Powys	3414	2673
Cardiff North	3701	2979	North Rhondda	2414	1721
Cardiff South East	2041	1342	North Taf Ely	2275	1758
Cardiff South West	2903	1860	Penderi	2017	1464
Cardiff West	2220	1543	South Ceredigion	2968	2081
Central & South Denbighshire	2141	1797	South Cynon	1270	870
Central Vale	2985	2240	South Merthyr Tydfil	1568	928
City & Cardiff South	1353	748	South Pembrokeshire	3229	2381
CityHealth	2758	1874	South Powys	2558	1904
Conwy East	3581	2688	South Rhondda	2702	1930
Conwy West	3651	2802	South Taf Ely	2450	1966
Cwmtawe	2162	1627	South Wrexham	2665	2101
Deeside, Hawarden & Saltney	2761	2323	Taf / Teifi / Tywi	3155	2331
Dwyfor	1516	1107	Torfaen North	2596	2133
Eastern Vale	1695	1347	Torfaen South	2216	1786
Holywell & Flint	2001	1550	Upper Valleys	1735	1410
Llanelli	3593	2821	West & North Wrexham	1833	1586
Llwchwr	2199	1769	Western Vale	1212	1024
Meirionnydd	1978	1535	Wrexham Town	2392	2032

Figure 2. All Wales CHD QOF register as a per cent of the modelled prevalence



7 Discussion

It proved to be possible to replicate the work of Walford *et al* (2011) and in doing so produce modelled CHD prevalence estimates for Wales. An issue was identified with both this work and the work of Walford *et al* (2011) in that both models appear to over-estimate the prevalence when compared to the respective health survey estimates from which they were derived.

The exact causes of the model overestimating CHD at the national and local level would require detailed examination beyond the scope of this project. However, as an initial assessment of the possible effect, the cluster level model results were rescaled according to the difference between the model and WHS results at the local authority level (Appendix C). This crude adjustment reduced the gap between modelled and QOF register prevalence (Appendix D). In three clusters this adjustment made the QOF prevalence greater than the modelled prevalence (Appendix E). However, these should not be regarded as official results of this exercise and are shown for illustrative purposes only.

Additional investigations, comparing the WHS actual and modelled results at the strata level, suggested that the model tended to most overestimate CHD prevalence in older smokers of both sexes and in the oldest male ex-smokers and never smokers in the more deprived fifths. The model tended to most underestimate CHD prevalence in the oldest female ex and never smokers and in the oldest male ex and never smokers in the least deprived fifths.

7.1 Assumptions

A number of assumptions have been made in deriving these prevalence estimates:

- The proportion of smokers in each age, sex and deprivation strata locally is the same as the national equivalent. There is no actual smoking prevalence estimate available for registered populations and the WHS cannot provide local strata specific estimates. Therefore the smoking prevalence used in this model is an estimate.
- The QOF CHD prevalence in the under 25's is very low. Therefore their exclusion from the modelled prevalence estimates and inclusion in the QOF data would have little effect on the comparison of the two.
- The exclusion of ethnic origin does not have an effect on most clusters. However, the modelled estimates for clusters with significant ethnic origin populations, which are known to have higher prevalence of CHD, may be underestimates.
- The WHS responses are an accurate representation of treated heart disease in the wider Welsh population.

7.2 Using Cluster level modelled prevalence

Extreme caution is urged when comparing the modelled prevalence estimates with the QOF register.

Any differences between the modelled prevalence estimates and the QOF register cannot be attributed to any cause without further investigation. Possible causes include, but are not limited to:

- Deficiencies in the model, or the data on which the model is built, including the reliability of existing data and the unavailability of some potentially important data, such as true local smoking prevalence.

- Untypical populations invalidating at least one of the assumptions upon which the model is based.

This method in effect is a comparison of (self reported) treated prevalence and general practice register diagnosed disease prevalence. It does not in any way identify undiagnosed cases, though may identify gaps between conditions being managed in different care settings.

The fact that the model does not return the same results as the survey upon which it is based, and the somewhat systematic strata level discrepancies, suggests that further work would be required to refine and improve a CHD prevalence model.

The true test of any such model would be to conduct a programme of active case finding within one or more clusters. Whilst this model may provide an indication of where this may be more likely to yield a greater number of cases, it cannot be assumed that the numbers modelled definitely exist within the community.

If, bearing in mind the issues mentioned previously including the reliability of the self reporting of disease in the WHS and the overall over estimation of the model, the modelled prevalence is closer to the true prevalence than the QOF register is, then the gap between the two measures may be somewhat indicative of a shortfall in case ascertainment.

8 References

Walford H, Ramsay L, Soljak M, Gordon F & Birger R (2011) *CHD prevalence modelling briefing document*, APHO. Available at http://www.apho.org.uk/resource/item.aspx?RID=111141

Box G & Draper N (1987), *Empirical Model Building and Response Surfaces*, John Wiley & Sons, New York.

Hosmer DW, Lemeshow S & Sturdivant RX (2013) Applied logistic regression (3rd Ed.). Hoboken, N.J.: Wiley.

9 Appendices

9.1 Appendix A: Derived definition of CHD in WHS

For the WHS data, a composite variable was created which indicated CHD if either (or both) of the following WHS questions were answered "Yes":

Have you **ever** been treated for heart attack?

Are you **currently** being treated for angina?

9.2 Appendix B: Definition of CHD in QOF

The Read codes v2 QOF definition of CHD is G3... – G309., G30B. - G330z (excluding G310.), G33z. - G3401, G342. – G35X., G38.. – G3z.. and Gyu3.% (excluding Gyu31).

G3...	Ischaemic heart disease
G30..	Acute myocardial infarction
G300.	Acute anterolateral infarction
G301.	Other specified anterior myocardial infarction
G3010	Acute anteroapical infarction
G3011	Acute anteroseptal infarction
G301z	Anterior myocardial infarction NOS
G302.	Acute inferolateral infarction
G303.	Acute inferoposterior infarction
G304.	Posterior myocardial infarction NOS
G305.	Lateral myocardial infarction NOS
G306.	True posterior myocardial infarction
G307.	Acute subendocardial infarction
G3070	Acute non-Q wave infarction
G3071	Acute non-ST segment elevation myocardial infarction
G308.	Inferior myocardial infarction NOS
G309.	Acute Q-wave infarct
G30B.	Acute posterolateral myocardial infarction
G30X.	Acute transmural myocardial infarction of unspecified site
G30X0	Acute ST segment elevation myocardial infarction
G30y.	Other acute myocardial infarction
G30y0	Acute atrial infarction
G30y1	Acute papillary muscle infarction
G30y2	Acute septal infarction
G30yz	Other acute myocardial infarction NOS
G30z.	Acute myocardial infarction NOS

G31..	Other acute and subacute ischaemic heart disease
G311.	Preinfarction syndrome
G3110	Myocardial infarction aborted
G3111	Unstable angina
G3112	Angina at rest
G3113	Refractory angina
G3114	Worsening angina
G3115	Acute coronary syndrome
G311z	Preinfarction syndrome NOS
G312.	Coronary thrombosis not resulting in myocardial infarction
G31y.	Other acute and subacute ischaemic heart disease
G31y0	Acute coronary insufficiency
G31y1	Microinfarction of heart
G31y2	Subendocardial ischaemia
G31y3	Transient myocardial ischaemia
G31yz	Other acute and subacute ischaemic heart disease NOS
G32..	Old myocardial infarction
G33..	Angina pectoris
G330.	Angina decubitus
G3300	Nocturnal angina
G330z	Angina decubitus NOS
G33z.	Angina pectoris NOS
G33z0	Status anginosus
G33z1	Stenocardia
G33z2	Syncope anginosa
G33z3	Angina on effort
G33z4	Ischaemic chest pain
G33z5	Post infarct angina
G33z6	New onset angina
G33z7	Stable angina
G33zz	Angina pectoris NOS
G34..	Other chronic ischaemic heart disease
G340.	Coronary atherosclerosis
G3400	Single coronary vessel disease
G3401	Double coronary vessel disease
G342.	Atherosclerotic cardiovascular disease
G343.	Ischaemic cardiomyopathy
G344.	Silent myocardial ischaemia
G34y.	Other specified chronic ischaemic heart disease
G34y0	Chronic coronary insufficiency
G34y1	Chronic myocardial ischaemia
G34yz	Other specified chronic ischaemic heart disease NOS
G34z.	Other chronic ischaemic heart disease NOS
G34z0	Asymptomatic coronary heart disease
G35..	Subsequent myocardial infarction
G350.	Subsequent myocardial infarction of anterior wall
G351.	Subsequent myocardial infarction of inferior wall

G353.	Subsequent myocardial infarction of other sites
G35X.	Subsequent myocardial infarction of unspecified site
G38..	Postoperative myocardial infarction
G380.	Postoperative transmural myocardial infarction of anterior wall
G381.	Postoperative transmural myocardial infarction of inferior wall
G382.	Postoperative transmural myocardial infarction of other sites
G383.	Postoperative transmural myocardial infarction of unspecified site
G384.	Postoperative subendocardial myocardial infarction
G38z.	Postoperative myocardial infarction, unspecified
G39..	Coronary microvascular disease
G3y..	Other specified ischaemic heart disease
G3z..	Ischaemic heart disease NOS
Gyu30	[X]Other forms of angina pectoris
Gyu32	[X]Other forms of acute ischaemic heart disease
Gyu33	[X]Other forms of chronic ischaemic heart disease
Gyu34	[X]Acute transmural myocardial infarction of unspecified site
Gyu35	[X]Subsequent myocardial infarction of other sites
Gyu36	[X]Subsequent myocardial infarction of unspecified site
Gyu3.	[X]Ischaemic heart diseases

9.3 Appendix C: Local Authority model and WHS reported prevalence reconciliation

Calculated as 1 – [(ModPrevRate-RepPrevRate)/ModPrevRate]

Where ModPrevRate = modelled prevalence / registered pop
RepPrevRate = WHS reported prevalence / resident pop

Local Authority	Adjustment factor
Torfaen	0.7220
Pembrokeshire	0.7795
Ceredigion	0.8021
Anglesey	0.8267
Powys	0.8467
Cardiff	0.8470
Vale of Glamorgan	0.8669
Wrexham	0.8710
Denbighshire	0.8780
Monmouthshire	0.8911
Gwynedd	0.8949
Merthyr Tydfil	0.9194
Flintshire	0.9286
Carmarthenshire	0.9393
Newport	0.9468
Conwy	0.9475
Caerphilly	0.9631
Bridgend	0.9775
Blaenau Gwent	0.9806
Neath Port Talbot	0.9975
Swansea	1.0017
Rhondda Cynon Taf	1.0963
Wales	0.9237

9.4 Appendix D: Adjusted modelled prevalence and QOF CHD register counts

cluster	Prevalence modelled	QOF	cluster	Prevalence modelled	QOF
Afan	3261	2278	Mid Powys	1503	1233
Amman/Gwendraeth	3139	2946	Mold, Buckley & Caergwle	2106	1927
Anglesey	3328	2972	Monmouthshire North	2506	2107
Arfon	2602	2044	Monmouthshire South	1993	1674
BayHealth	3559	2547	Neath	3267	2309
Blaenau Gwent East	2324	1720	Newport Central	2159	1627
Blaenau Gwent West	2080	1556	Newport East	2339	1880
Bridgend East Network	3080	2706	Newport West	2514	1971
Bridgend North Network	2954	2484	North Ceredigion	1644	1674
Bridgend West Network	1995	1757	North Cynon	2605	1699
Caerphilly East	2804	2393	North Denbighshire	3209	3082
Caerphilly North	3855	3072	North Merthyr Tydfil	1711	1424
Caerphilly South	2661	2186	North Pembrokeshire	2776	2502
Cardiff East	2126	1766	North Powys	2890	2673
Cardiff North	3135	2979	North Rhondda	2646	1721
Cardiff South East	1729	1342	North Taf Ely	2494	1758
Cardiff South West	2459	1860	Penderi	2021	1464
Cardiff West	1880	1543	South Ceredigion	2381	2081
Central & South Denbighshire	1880	1797	South Cynon	1392	870
Central Vale	2588	2240	South Merthyr Tydfil	1442	928
City & Cardiff South	1146	748	South Pembrokeshire	2517	2381
CityHealth	2763	1874	South Powys	2166	1904
Conwy East	3393	2688	South Rhondda	2962	1930
Conwy West	3460	2802	South Taf Ely	2687	1966
Cwmtawe	2165	1627	South Wrexham	2321	2101
Deeside, Hawarden & Saltney	2564	2323	Taf / Teifi / Tywi	2963	2331
Dwyfor	1357	1107	Torfaen North	1875	2133
Eastern Vale	1470	1347	Torfaen South	1600	1786
Holywell & Flint	1858	1550	Upper Valleys	1730	1410
Llanelli	3375	2821	West & North Wrexham	1596	1586
Llwchwr	2203	1769	Western Vale	1051	1024
Meirionnydd	1770	1535	Wrexham Town	2083	2032

9.5 Appendix E: QOF register as a per cent of the adjusted modelled prevalence