Technical Briefing 8

January 2011

ASSOCIATION OF PUBLIC HEALTH OBSERVATORIES
APHO

Prevalence Modelling

Purpose

This is the eighth in a series of technical briefings produced by the Association of Public Health Observatories (APHO), designed to support public health practitioners and analysts and to promote the use of public health intelligence in decision making.

In this briefing we discuss the need for prevalence modelling and provide an overview of various methods of generating estimates of the prevalence of diseases or risk factors in local populations. Some of the most common methods for assessing the accuracy and robustness of models are summarised and some of the challenges in developing and using prevalence estimates are considered.

Further material to support the technical briefing series (and disease prevalence estimates developed by APHO) are available from www.apho.org.uk

Contents

Authors

Hannah Walford
Ben Kearns
Steve Barron

Contributors

Helen Bolton
Anne Cunningham (NHS Blackburn with Darwen)
Mark Dancox
Julian Flowers
Paul Fryers
Gyles Glover
Naomi Holman
James Hollinshead
John Kemm
Michael Soljak (Imperial College London)

What is prevalence modelling?

Prevalence modelling is a technique used to estimate the number of people with a particular condition or risk factor in a population when direct evidence is not available. Direct evidence may be lacking because surveys or data collection have not been undertaken, are technically impractical, or are unreliable.

Methods for generating synthetic or modelled estimates range in complexity from simple to highly sophisticated. Crude estimates of the number of cases can be generated by applying known prevalence rates to a different population, for example applying national rates measured in a large survey to a local population; or applying local rates for a recent year to a projected future population. However, many factors such as age, gender, deprivation and ethnicity can influence the prevalence of a behaviour, risk factor or disease and more complex epidemiological modelling techniques are required in order to take such factors into account.

The need for prevalence modelling

In many cases, routine data are not available to measure directly the frequency and distribution of diseases or behaviours in local populations. Modelling is often the best alternative for quantifying prevalence in the absence of reliable direct measures. Typically, direct measures are not available at local level for lifestyle behaviours such as smoking or alcohol consumption, or for diseases that are generally managed in primary care, for example diabetes or hypertension.

Understanding the distribution of behaviours that affect health (either positively or negatively) is increasingly important in the allocation of public health resources and the delivery of interventions. Prevalence modelling can be used to assess need and help identify those communities that will most benefit from public health initiatives. Modelled estimates of prevalence can also be helpful in explaining variations in care utilisation and outcomes.

The quality and completeness of routine datasets, such as the Quality and Outcomes Framework (QOF) for primary care, are improving, and QOF is now a reasonable basis for prevalence estimates of many diseases. However, the measured prevalence is limited to diagnosed disease. Modelled estimates that include undiagnosed disease in the population can offer additional information that can inform case-finding initiatives and highlight areas where under-diagnosis could be an issue.

There is considerable interest in obtaining estimates of expected prevalence at various geographies and for different subgroups of the population, for example ethnic groups or age cohorts, to assist in understanding and tackling health inequalities.

Methods

Many methods exist for creating synthetic estimates of prevalence, and in many cases methodologies are combined and adapted to make best use of the information and data available. There is often a balance to be struck between increasing the complexity of the model by incorporating more contributory factors and the availability of good quality data at local level to populate the model. These input requirements of a model are often restricted by what information is available. Complex models can also suffer from difficulty of interpretation, which negates the benefit of increased accuracy.

All models are based on assumptions. Good models clearly state the assumptions that have been made and good interpretation of modelled estimates takes into account the limitations of the assumptions.

1) Prevalence estimates from studies and trials applied to local populations

Although this can be a quick and simple method, its usefulness depends on the size of the studies and hence the precision of the prevalence estimates, and whether or not they include prevalence estimates for population subgroups. Incorporating different prevalence rates for subgroups by deprivation, gender and/or age is usually useful and technically straightforward. However, increasing the specificity of the model by using prevalence estimates in many subgroups (for example gender, age, ethnicity, smoking status and employment status) can limit the range of local levels to which it can be applied, because there is seldom sufficient alignment between data routinely available at local level and all the variables used in the source study.

2) Regression models using demographic characteristics from large surveys

Multiple variables from large surveys can be used to model the risk factors for a behaviour or disease, using techniques such as multinomial logistic regression. However, it is important to limit the factors considered to those for which data are available in the population of interest. For example, cholesterol level or family history of disease may be important risk factors which were recorded in the source survey, but such information is not usually available at population level and therefore these are not appropriate variables to be included in a disease prevalence model.

National surveys are usually limited to people living in private households and omit populations such as the homeless, those living in institutional care, 'special populations' (armed forces and prisoners) and people particularly likely to decline to participate. For some disease areas, notably some types of mental illness, these omitted populations can be particularly important. Despite this limitation, national surveys are often the best source of prevalence information available, but where possible should be used in conjunction with other evidence about the likely extent to which they miss cases. Models can then be adjusted to take account of the resultant under-estimation of prevalence.

Although regression models most commonly use survey data, other data sources, for example information recorded in general practices, can also be used to create this type of prevalence model.

3) Capture-recapture methods

Capture-recapture methods are used to estimate the number of people with a disease or behaviour, for example the total number of injecting drug users, including those unknown to any services. A random sample of people is taken ('captured') from the whole population, and examined for the characteristic of interest. 'Sample 1' is the number of individuals found to have the characteristic. A second random sample of the whole population is then taken and 'Sample 2' is defined similarly as those found to have the characteristic. Some people will appear in both Sample 1 and Sample 2 and the proportion of Sample 2 that were also in Sample 1 is calculated. This proportion is assumed to be equivalent to the proportion of all the people with the characteristic in the whole population that were captured in Sample 1. Hence, by dividing Sample 1 by this proportion an estimate of the total number with the characteristic is obtained. (Table 1)

Table 1: Example of capture-recapture sampling

Number in Sample 1	100
Number in Sample 2	105
Number of individuals in both Sample 1 and Sample 2	21
Proportion of individuals in Sample 2 that were also in Sample 1	21/105 = 0.2
Size of population with characteristic	100/0.2 = 500

The methodology is more complex when multiple data sources are used to capture samples of the population, and weightings have to be introduced to take into account interdependence of datasets. Extensive details of the use of capture-recapture methods in estimating the prevalence of problem drug use are given by Hay et al.[1]

4) Combining multiple sources

Often there are several estimates of prevalence rates available from larger and smaller scale epidemiological studies, which need to be integrated. For example, regional prevalence rates from large national surveys can provide control totals for smaller geographies for which synthetic estimates are generated. Each source will have strengths and weaknesses: national surveys may have robust sampling and include a wide range of risk factors, but can lack local detail, whereas local studies may use more elaborate methods, for example capture-recapture techniques, but may focus on unrepresentative areas. Combining prevalence estimates requires critical appraisal of the appropriateness of each source and development of mathematical methodology to integrate the variance estimates from unrelated sources to produce an overall confidence interval for the synthetic estimate.

Meta-analysis techniques have been developed to combine multiple estimates of prevalence, each of which may have data quality issues, to produce one triangulated estimate with improved quality at small area level.[2,3] Estimates from a wide range of sources can be combined, including prevalence estimates from surveys, data from primary care and modelled synthetic estimates. Further details can be found in *APHO Technical Briefing 7: Measuring smoking prevalence in local populations.*[4]

Bayesian statistical methods can be employed to synthesise a diverse set of available data into a prevalence estimate. For example, Goubar et al[5] combined an array of information, including routine surveillance data and anonymous surveys, to estimate HIV prevalence in various risk groups using Markov chain Monte Carlo simulation.

Validation, confidence intervals and robustness

The accuracy of model outputs depends on the predictive power of the model and on the accuracy of the input data. Models should be subjected to validation checks to ascertain their robustness and general applicability.

Estimates of the accuracy of prevalence estimates based on simple models can be generated by combining the uncertainty in prevalence rates from the source study or trial with the stochastic variation expected given the size of the local population. This approach results in a range estimate for the prevalence, rather than confidence intervals. The range estimates are calculated using the same methods as those used to derive the control limits for funnel plots. These are described in *Technical Briefing 2: Statistical process control methods in public health intelligence.*[6] However, if there is uncertainty around both the population data and the input data, the calculation of confidence intervals can be complex. Bootstrapping methods are commonly used in such situations.[7]

The APHO preferred method for calculating confidence intervals when using the capture-recapture method is that described by Cormack.[1,8,9]

Four ways of validating models are described here.

1) Sensitivity testing

Sensitivity testing can be useful in assessing how the uncertainty in input data affects prevalence estimates. For some models, very small variations in the input data will have a large effect on the results. Other models may be relatively insensitive to variability in input data. For example,[10] different sources of practice level smoking prevalence data were input into the APHO chronic obstructive pulmonary disease (COPD) models. Estimated COPD prevalence in general practices ranges from 1% to 7%. In 92% of cases, changing the source of smoking prevalence data made an absolute difference of less than 0.3% in the COPD prevalence estimate. Estimates generated using different smoking prevalence source data were strongly correlated with each other ($r^2 > 0.95$).

One-way sensitivity analysis such as this evaluates the impact of a change in one variable on the model results. Multi-way sensitivity analysis is more powerful and can be used to assess the impact of changing two or more variables simultaneously.[11]

2) Internal validation

One method of assessing the performance of a model is to use it to predict the response for each subject in the source data (e.g. a large survey). These predictions are called fitted values. The differences between the fitted and the observed values are called residuals. Residual analysis can be used to check the adequacy of any assumptions used when creating the model. It can also be used to identify whether any additional factors should be included.

To check the accuracy of the model, the predicted ‘classification’ of each individual (i.e. whether or not they have the disease or behaviour that is being modelled) can be compared with their actual classification. This will result in a ‘misclassification’ (also known as a ‘contingency’ or ‘confusion’) table (Table 2).

Table 2: Misclassification table of modelled results

		Actual	
		Negative	Positive
Modelled	Negative	A True negative	B False negative
	Positive	C False positive	D True positive

Specificity = true negative rate = A/(A + C)
= 1 – false positive rate= 1 – C/(A + C)
Sensitivity = true positive rate= D/(B + D)

3) Receiver-Operating Characteristic (ROC) Analysis

ROC analysis is a useful way of assessing the accuracy of a model by understanding the trade-off between the sensitivity (in this sense referring to the true positive rate; Table 2) and the specificity (the true negative rate).[12] The method was developed to assess the accuracy of distinguishing signal from noise in radar systems and has since been applied in many other settings, including clinical diagnostic testing and the evaluation of regression models that classify cases into two categories, for example diseased and non-diseased. Sensitivity is plotted against 1-specificity (specificity subtracted from one) over a range of values and the area under the curve (AUC or AUROC) is used as a summary of the predictive or diagnostic accuracy. A ‘perfect’ model that accurately predicts every case has AUROC = 1. Typically, models have a convex ROC curve and an AUROC between 0.5 (equivalent to random chance) and 1. A model with AUROC < 0.5 is less accurate than random chance.

4) External validation

Modelled estimates can be compared with observed prevalence where such measures exist. For example, modelled estimates for small areas can be aggregated to regional or national level and compared with survey measures of prevalence. In some models, local values are adjusted so that regional or national aggregates are consistent with observed prevalence.

Projections and forecasting using models

Prevalence models can often be adapted to predict future prevalence. The sophistication of projected prevalence estimates depends on the modelling methodology adopted, and falls into three broad categories:

- Same risk, changing (e.g. increasing and/or ageing) population. Use the same model coefficients or risks of disease but incorporate population projections. For example, what will be the prevalence of coronary heart disease (CHD) in 2020 if we assume that the age-specific risks do not change but we take into account the aging population? This is sometimes called the ‘prevalence ratio method’.

- Same population, changing risk. Use the same demographic information but change the risk profile. For example, what will be the prevalence of CHD if smoking prevalence reduces?

- Modify the population and the risks to produce ‘scenario models’ e.g. what will be the CHD prevalence in 2025 if the population ages and the smoking prevalence reduces?

One of the characteristics of complex systems such as health is that no matter how tightly the present state of the system is specified the future state cannot be confidently predicted. Extra care should be taken in interpreting modelled estimates of projected prevalence as the assumptions inherent to the model may not hold in the future.

Using prevalence models

A collection of case studies to illustrate the use of APHO prevalence models (see Box 1) by Primary Care Trusts (PCTs) has been put together by the Department of Health Informing Healthier Choices programme and is available at www.apho.org.uk/resource/item.aspx?RID=86900

It is important to remember that prevalence figures generated by models are synthetic estimates of the expected prevalence of disease. They are not ‘real’ measures of prevalence. Remember that ‘all models are wrong but some are useful.’[13] Discrepancies between modelled estimates and other sources of data (such as primary care disease registers) may be due to local variations not captured by the model and cannot be solely attributed to weaknesses in directly measured prevalence data. For local populations that differ significantly from a 'typical' population (e.g. a large black and minority ethnic (BME) population that has a very different smoking pattern to the national average) the assumptions of a model may not apply and discrepancies may occur. Local expert opinion (e.g. local GPs' knowledge of the pattern of disease) can be invaluable in interpreting and applying synthetic estimates of prevalence correctly and usefully. A typical use of prevalence estimates is to compare expected prevalence with recorded prevalence, for example from the QOF in England.[14,15] Such an approach needs to be taken with care. Are the two populations comparable, or are you trying to compare adult prevalence with all-age prevalence? Is the definition of disease used in the modelled estimates the same as the clinical definition used for diagnosis in primary care? Does the model include an estimate of undiagnosed disease or not? An understanding of these issues and differences is vital in interpreting any comparisons made between synthetic estimates and measured prevalence.

Because modelled prevalence is an estimate of expected prevalence, generally with the assumption that the local area behaves in the same way as the population from which the source data were derived, it is not straight forward to use synthetic estimates to evaluate the impact of a local intervention. For example, low modelled prevalence of binge drinking in a local area that has invested heavily in action to decrease alcohol misuse is not proof that the investment has reduced binge drinking. It is only an indication that the area can expect a low prevalence, given its demographic characteristics. Local interventions or prioritisation of an issue may explain discrepancies between modelled and directly measured prevalence, but the discrepancy does not prove that an intervention or policy is effective. It is not advisable to use prevalence models for performance management or to evaluate the impact of a local programme.

It is also inappropriate to use modelled estimates to monitor changes over time. Changes in estimated prevalence could be due to updated local input data (e.g. demographics) or changes in the source data used to generate a new version of the model. There may also have been adjustments in the modelling methodology used if source data have been re-modelled.

Box 1: The APHO prevalence modelling project

APHO has worked with partners and stakeholders including the Department of Health and NHS Diabetes to develop a suite of disease prevalence models to support health improvement and commissioning. To date, APHO has produced the models listed below. The models are intended to be used for:

- comparing service provision (including preventive services) with population need, and commissioning health and social care services
- planning future health and social care service requirements
- undertaking health equity audits
- clinical governance, e.g. comparing complication rates or admission rates after adjustment for variation in expected prevalence
- assessing the completeness of disease registers in primary care and the completeness of case finding

	Geographical coverage			**Populations**		
Disease area	**PCT**	**LA**	**GP**	**Age bands**	**Sex**	**Ethnicity**
Coronary heart disease	✓	✓	✓	✓	✓	✓
Stroke	✓	✓	✓	✓	✓	
Hypertension	✓	✓	✓	✓	✓	✓
Cardiovascular disease	✓	✓		✓	✓	✓
Chronic obstructive pulmonary disease	✓	✓	✓	✓	✓	✓
Diabetes	✓	✓	†	✓	✓	✓
Chronic kidney disease	✓	✓	†	✓	✓	
Dementia	✓	✓		✓	✓	
Common mental health problems	✓	✓		✓	✓	

† Users can enter practice data into an interactive tool to generate modelled estimates.
All of the APHO prevalence models are available at www.apho.org.uk/resource/view.aspx?RID=48308

Strengths and limitations

Understanding the strengths and limitations of any model is crucial if the results are to be appropriately used. However, this understanding is often hampered by the limited amount of information published about the methodology, development and testing of models. Unal et al[16] give the following checklist of ten areas that should be reported and discussed in a modelling paper. Although the subject of their paper is intervention models, the list applies equally well to prevalence modelling.

1. Aims of the project
2. Structure and methods of the model
3. Data quality (data availability, how up to date, comprehensive, any gaps in certain population groups or interventions, reasons for selecting or excluding specific data sources)
4. Methodological limitations
5. The assumptions used to address these deficiencies
6. Sensitivity analyses (one-way or preferably multi-way)
7. Whether the validity of the model was checked (with real observational data or with other models)
8. Replication of the model in different populations
9. Model results and comparisons with other studies
10. Social and economic policy implications of model outcomes

Box 2: Overview of prevalence models available at local level

Note that this is not a comprehensive list. Results of many of the models are also available through resources such as NHS Comparators (http://www.ic.nhs.uk/nhscomparators).

Source	Diseases/behaviours	Link	Comments
APHO	See Box 1	http://www.apho.org.uk/resource/view.aspx?RID=48308 http://www.inispho.org/publications/makingchronicconditionscount	Models were developed initially to cover England only. INIsPHO has adapted the stroke, CHD, hypertension, COPD and diabetes models for Ireland and Northern Ireland.
APHO	Smoking, binge drinking, fruit and vegetable consumption, obesity	http://www.apho.org.uk/resource/view.aspx?RID=91736	APHO commissioned an update to these models using 2006-2008 HSE data. Similar models based on older HSE data are available from the NHS Information Centre.
Doncaster PCT QOF benchmarking tool	All diseases (except depression) measured in QOF	http://www.doncaster.nhs.uk/about-us/our-roles-directories/public-health/public-health-intelligence-evaluation-team/tools-resources/qof-benchmarking-tool/	Produced as a rapid response to the QOF requirement for prevalence lists in general practices, in the absence of national models. Several of the models have been superseded by the more recent and robust APHO models. The ability to generate small area and practice-level prevalence estimates has led to popularity and widespread use.
POPPI (Projecting Older People Population Information)	Older peoples' health status, social care need and determinants of health	http://www.poppi.org.uk	Provides data at English LA level. Uses prevalence rates to estimate the impact of a wide range of diseases and conditions for populations aged 65+ by age, sex, ethnic group and a range of other socio-economic indicators.
PANSI (Projecting Adult Need and Service Information)	Physical and learning disabilities and mental health conditions	http://www.pansi.org.uk	Provides data at English LA level. Uses prevalence rates to estimate the impact of a wide range of diseases and conditions for populations aged 18-64 by age, sex, ethnic group and disability living allowance status.
NEHEM (National Eye Health Epidemiological Model)	Age-related macular degeneration, glaucoma, cataract, low vision	http://www.eyehealthmodel.org.uk	Estimates of the number of cases and prevalence by LA and PCO/local health board for the whole of the UK.
National Treatment Agency	Opiate and/or crack cocaine use	http://www.nta.nhs.uk/facts-prevalence.aspx	Estimates of the number of problem drug users in each Drug Action team (DAT) area and Government Office Region in England. Data from drug treatment services and the criminal justice system are combined to create the prevalence estimates.

Table 3: Summary of strengths and limitations of prevalence models

Strengths	Limitations
Prevalence estimates can be generated at any geographical level or for any population for which input data are available.	Input data are often difficult to obtain or unreliable for small areas. Assumptions made in the input data may be invalid, for example it may be necessary to assume that a risk factor is equally common across all age bands.
Can take into account known risk factors for disease.	Risk factors can themselves be hard to quantify. Risk factor data may not be available and may also require modelling. For example modelled estimates of smoking prevalence are used as an input to the APHO CHD prevalence model.
Models can incorporate estimates of undiagnosed disease prevalence	Definition of disease is dependent on the definition used in the source data (e.g. survey). This can be a particular issue with diseases where the clinical diagnosis differs from the measurement available in the survey. For example, hypertension is clinically diagnosed after raised blood pressure has been measured over a period of time, whereas the Health Survey for England (HSE), used as the basis for the APHO hypertension model identifies hypertension based on blood pressure measured on just one day.
Model results can (and should) include measures of uncertainty around the estimates and give range estimates rather than just point estimates.	Accurately measuring all of the sources of uncertainty that affect modelled estimates and combining them into a confidence interval can be methodologically difficult.
Prevalence models can be re-run to produce updated prevalence estimates (e.g. using updated population data), provided the original assumptions remain valid.	Modelled estimates may be based on out-of-date source data. This is a more important issue for risk factors or diseases whose prevalence changes rapidly (e.g. within 5–10 years). If models are re-derived using updated source data care must be taken in interpreting changes over time. For example, it is not valid to compare modelled estimates of smoking prevalence based on HSE 2000-2002 with those based on HSE 2003-2005 due to differences in modelling methodology.
Modelled prevalence estimates are straightforward to interpret and apply (as long as assumptions and methodology are transparent and clearly stated).	Some models are 'black boxes'. If assumptions, methodology and input data are not made clear, modelled estimates should be treated with increased caution.

Glossary

Bayesian inference: Statistical methods in which evidence from observations is used in conjunction with prior knowledge to infer and update the probability that a hypothesis is true. In traditional statistical inference, evidence is used simply to accept or reject a hypothesis depending on whether the probability that it is true is above or below an arbitrary value. A simple introduction is available at http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/What_is_Bay_stats.pdf

Bootstrapping: Statistical method for testing a model for stability, and sensitivity to statistical variability in the input data. The original sample is randomly re-sampled a large number of times, based on the statistical properties of the input variables, to give many possible alternative results, from which the distribution of possible model outcomes is derived. This method can be used to provide confidence intervals when other methods are not applicable.

Input data: In this document, the term input data denotes the explanatory factors which the model takes into account in estimating prevalence.

Markov chain Monte Carlo simulation: Algorithms used in Bayesian models to generate simulated samples from a statistical distribution.

Meta-analysis: Statistical techniques used to combine the results of several studies addressing the same research question.

Multinomial logistic regression: Multinomial logistic regression is used when the variable being modelled is categorical and has more than two categories. Standard logistic regression is used when the variable being modelled is binary – i.e. it has only two outcomes (e.g. presence or absence of a characteristic).

Prevalence: The number of cases of a disease or characteristic that are present in a particular population at a given time.

QOF: Quality and Outcomes Framework, the mechanism for rewarding general practitioners in England for meeting a defined set of quality criteria (http://www.ic.nhs.uk/qof).

Regression: A statistical method for determining the statistical relationship between a dependent variable (the outcome being modelled) and one or more independent variables (the input data on factors thought to affect the dependent variable).

Stochastic variation: Random statistical variation.

References

1. Hay G, Gannon M, MacDougall J, Millar T, Eastwood C, McKeganey N. Local and national estimates of the prevalence of opiate use and/or crack cocaine use (2004/05). In Singleton N, Murray R, Tinsley L (eds). Measuring different aspects of problem drug use: methodological developments (2nd ed). London: Home Office; 2006. Available at http://www.homeoffice.gov.uk/rds/pdfs06/rdsolr1606.pdf
2. Manzi G, Spiegelhalter DJ, Flowers J, Turner RM, Thomson SG. Combining small-area smoking prevalence estimates from multiple surveys. In RSS 2008 Conference Abstracts Booklet. London: RSS; 2008. Available at http://www.rss.org.uk
3. Eastern Region Public Health Observatory. Experimental statistics on adult smoking prevalence for East of England local authorities 2005. ERPHO; 2008. Available at http://www.erpho.org.uk/viewResource.aspx?id=17893
4. Davies C, Jenner D. Technical Briefing 7: Measuring smoking prevalence in local populations. York: APHO; 2010. Available at http://www.apho.org.uk/resource/item.aspx?RID=87192
5. Goubar A, Ades AE, De Angelis D, McGarrigle CA, Mercer CH, Tookey PA, Fenton K, Gill ON. Estimates of human immunodeficiency virus prevalence and proportion diagnosed based on Bayesian multiparameter synthesis of surveillance data. J R Stat Soc A 2008;171(3):541-580. Available at http://www.rss.org.uk
6. Flowers J. Technical Briefing 2: Statistical process control methods in public health intelligence. York: APHO; 2007. Available at http://www.apho.org.uk/resource/item.aspx?RID=39445
7. Baron-Cohen S, Scott FJ, Allison C, Williams J, Bolton P, Matthews FE, Brayne C. Prevalence of autism-spectrum conditions: UK school-based population study. Br J Psychiatry 2009;194(6):500-509.
8. Cormack RM. Interval Estimation for Mark-Recapture Studies of Closed Populations. Biometrics 1992;48(2):567-576.
9. European Monitoring Centre for Drugs and Drug Addiction (EMCDDA). Methodological Guidelines to Estimate the Prevalence of Problem Drug Use on the Local Level. Lisbon: EMCDDA; 1999.
10. Eastern Region Public Health Observatory. Comparison of smoking prevalence source data in COPD modelled prevalence estimates. ERPHO; 2010. Available at http://www.erpho.org.uk/viewResource.aspx?id=21180
11. Taylor M. What is sensitivity analysis? Haywood Medical Communications; 2009. Available at http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/What_is_sens_analy.pdf
12. Zou KH, O'Malley AJ, Mauri L. Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. Circulation 2007;115:654-657. Available at http://circ.ahajournals.org/cgi/content/full/115/5/654
13. Box GEP, Draper NR. Empirical Model-Building and Response Surfaces. Wiley; 1987:424.
14. Martin D, Wright JA. Disease prevalence in the English population: A comparison of primary care registers and prevalence models. Soc Sci & Med 2009;68(2):266-274. Available at http://dx.doi.org/10.1016/j.socscimed.2008.10.021
15. Nacul L, Soljak M, Samarasundera E, Hopkinson NS, Lacerda E, Indulkar T, Flowers J, Walford H and Majeed A. COPD in England: a comparison of expected, model-based prevalence and observed prevalence from general practice data. J Public Health fdq031 first published online June 3, 2010 doi:10.1093/pubmed/fdq031. Available at http://jpubhealth.oxfordjournals.org/content/early/2010/06/03/pubmed.fdq031
16. Unal B, Capewell S and Critchley JA. Coronary heart disease policy models: a systematic review. BMC Public Health 2006;6:213. Available at http://www.biomedcentral.com/1471-2458/6/213

All links accessed 19 January 2011.

About the Association of Public Health Observatories (APHO)

The Association of Public Health Observatories (APHO) represents and co-ordinates a network of 12 public health observatories (PHOs) working across the five nations of England, Scotland, Wales, Northern Ireland and the Republic of Ireland.

APHO facilitates joint working across the PHOs to produce information, data and intelligence on people's health and health care for practitioners, policy makers and the public.

APHO is the largest concentration of public health intelligence expertise in the UK and Republic of Ireland, with over 150 public health intelligence professionals.

APHO helps commissioners to ensure that they get the information they need and our websites provide a regular stream of products and tools, training and technical support.

We work with partners to improve the quality and accessibility of the data and intelligence available to decision makers.

We are constantly developing and learning new and better ways of analysing health intelligence data. We use these new methods to improve the quality of our own work, and share them with others.

Updates and more material, including methods and tools to support our technical briefing series, are available through our website at http://www.apho.org.uk

For further information contact:
Association of Public Health Observatories
Innovation Centre, York Science Park,
Heslington, York YO10 5DG

Telephone: 01904 567658
http://www.apho.org.uk