Genetic epidemiology and primary care

Blair H Smith, Graham CM Watt, Harry Campbell and Aziz Sheikh

ABSTRACT

Large-scale, population-based studies of genetic epidemiology are under way or planned in several countries, including the UK. The results will have many implications for GPs and their patients. Primary care has much to contribute to this research, and basing genetic epidemiology studies in primary care will confer several advantages. These include enhanced public engagement, building on the personal relationships and trust that are at the core of primary care practice; methodological factors that will strengthen study design; and the potential of linkage of multiple datasets and between networks of research practices. Essential development work with primary care professionals and the public is, however, required for this to happen, and, if undertaken, this work will have the additional important benefit of increasing the uptake of new knowledge into general practice.

Keywords
epidemiology; genetics; primary care; public engagement.

***BH Smith**, MD, MEd, FRCGP, reader in general practice, Department of General Practice and Primary Care, University of Aberdeen, Aberdeen. **GCM Watt**, MD, FRCGP, FMedSci, professor of general practice, Division of Community-based Sciences, University of Glasgow, Glasgow. **H Campbell**, MD, FFPH, FRCP, professor of genetic epidemiology and public health; **A Sheikh**, MD, MRCP, FRCGP, professor of primary care research and development, Division of Community Health Sciences, University of Edinburgh, Edinburgh.*

Address for correspondence
Dr Blair H Smith, Department of General Practice and Primary Care, University of Aberdeen, Foresterhill Health Centre, Westburn Road, Aberdeen AB25 2AY.
E-mail: blairsmith@abdn.ac.uk

Submitted: 30 December 2004; **Editor's response:** 4 May 2005; **final acceptance:** 29 July 2005.

INTRODUCTION

Large studies aiming to identify and quantify genetic contributions to common complex diseases are current or planned in at least nine countries[1–3] (Table 1), and represent the next logical step in post-genomic research.[4] In the UK, Biobank, funded by the Medical Research Council, The Wellcome Trust, the Department of Health and the Scottish Executive Health Department, is planned as a cohort study of 500 000 individuals aged between 45 and 69 years.[5,6] A separate and complementary initiative is also planned in Scotland, funded by the Scottish Executive Health Department, where Generation Scotland will aim to recruit 15 000 rising to a proposed 50 000 adults in family groupings from the community.[7] These and other projects will seek to combine detailed information about genotypes and phenotypes and, in doing so, will aim to tackle questions pertinent to the aetiology, prevention, development, progression and management of common complex (multifactorial) diseases of public health importance.

There are several approaches to recruiting the large numbers of subjects and collecting the potentially highly sensitive data required for these studies (Table 1). For example, in Estonia, participants are recruited as they attend for primary medical care.[8] In contrast, the study in Iceland is based on samples and data that have been routinely collected and assumes consent to use of this information unless individuals specifically opt out.[9] This paper discusses the potential benefits to and practicalities of recruiting through primary care, and argues that while genetic epidemiology is important and relevant to primary care, primary care also offers considerable 'added value' to genetic epidemiology.

GENETICS AND PRIMARY CARE

Genetic science has advanced exponentially in recent years, including, most notably, the mapping of the human genome. This has, and will increasingly have, many implications for primary care practice, as has been highlighted in several recent reviews.[10–13] These include: increased public and professional awareness of the genetic

Table 1. Large-scale genetic epidemiology studies planned or under way.[1–3]

Study location	Study name	Sample size (proposed)	Sample characteristics	Stated objectives
Canada	CARTaGENE[61]	60 000	Representative of the population of Quebec, aged 25–74 years	To study the founder effect and ethnic heterogeneity
China	Kadoorie Project[62]	500 000	Representative sample across several regions of China	To study gene–environment interaction in the aetiology of common diseases, and to serve as a resource for future research
Estonia	Estonian Genome Project[8]	1 000 000	Volunteers, during routine primary medical care (no stated age restriction)	To create a database of health, genealogy and genome data that will comprise a large part of the Estonian population, aiming for about 71% of the population
Iceland	DeCode/Icelandic Health Sector Database[9]	275 000	The population of Iceland, unless individuals opt out. Routine data and samples	To map the genes of the Icelandic population, and to investigate 12 genetically influenced diseases
Latvia	Genome Database of the Latvian Population[63]	None stated	Volunteers (no age restriction), during routine primary medical care, aiming for a nationally representative sample	To build a genome database of the Latvian population, with multiple clinical and commercial expectations
Sweden	UmanGenomics[64]	100 000	Inhabitants of Västerbotten, northern Sweden	Database of genetic and medical information, for commercialisation and academic research
Singapore	Genome Institute of Singapore[1]	None stated	The population of Singapore	To establish a database of information on medical history, genealogical and genetic data
UK	Biobank[6]	500 000	Adults aged 45–69 years. Recruitment method to be finalised, but probably involving primary care	To study gene–environment interaction in the aetiology of common diseases, and to serve as a resource for future research
US	Not yet determined[3]	500 000	Not yet determined	To study genetic and environmental influences on common diseases. Currently a preliminary proposal only
Scotland	Generation Scotland: the Scottish Family Health Study[7]	50 000	Siblings, aged 35–55 years, recruited through primary care	Assessment of the familial aggregation and heritability of important disease-related traits, and identification of genetic loci that contribute to these traits
Multinational	GenomEUtwin[65]	850 000	Twin pairs — Finland, Sweden, Norway, Denmark, The Netherlands, Italy, UK, Australia	To characterise genetic, environmental and lifestyle components in the background of health problems

contribution to some diseases, such as cancer; increased availability of genetic tests; increased requirement for genetic literacy and relevant communication skills among primary care professionals; and an increased number of ethical dilemmas in practice, including issues around self-determination, pre-determination, and confidentiality.

Until recently, most genetic science and clinical practice has focused on relatively uncommon genetic diseases such as cystic fibrosis and phenylketonuria. These diseases are caused by highly penetrant mutations in single genes and show classical Mendelian inheritance patterns. They provide important insights into the study of genetic inheritance and the identification of genes underlying hundreds of human diseases. They often have significant impact on individuals and their families, and there can also be some impact on communities and practices where there are populations with a high prevalence of conditions, such as the haemoglobinopathies. However, their overall impact on public health and daily primary care practice is relatively minor in comparison with other conditions, such as diabetes and ischaemic heart disease.

There has been recent progress in moving from a focus on these relatively simple Mendelian conditions to identifying Mendelian subsets of common complex conditions. These include breast and colorectal cancer, diabetes and osteoporosis associated with specified gene variants that have a high or relatively high penetrance.[14–16] These new 'taxonomies of disease'[17] may or may not be associated with clinically distinct patterns, but their identification informs prognosis and prevention among sufferers and family members of sufferers. At present, there are still few of these conditions, and the overall contribution of genetics to their aetiology is low: around 5% in the case of breast

How this fits in

There is major investment worldwide in genetic epidemiology research. This paper highlights the potential implications for primary care practice and professionals, and also the important contributions that they can make to this research.

and colorectal cancer.[18,19] Most clinical work in primary care, however, addresses common and complex disorders, in which gene–environment interactions are often important, and which are the subject of current and future epidemiological research.

GENETIC EPIDEMIOLOGY

Genetic epidemiology is a relatively new science.[20] Genetics has advanced by using epidemiological methods,[20,21] and this new field of epidemiology has also advanced by drawing on genetic principles (for example in Mendelian randomisation[22]). It has been argued that to have the maximum public health impact this research should be directed at modifiable risk factors, or the illumination of genetic factors that will improve the prevention or management of important conditions. It must also consider health as well as illness, seeking factors associated with longevity and absence of disease.

Genetic epidemiology studies can be grouped into those that seek to identify new genetic variants that cause disease (gene discovery studies) or those that aim to understand the importance of these variants in terms of the frequency or size of effect (gene characterisation studies). The former studies are often conducted in special groups of people or populations with particularly high disease incidence or risk, whereas the latter are typically conducted in representative groups with careful sampling so that results are generalisable to the wider population. A range of traditional and more specialised study designs are employed,[23] with increasing recognition of the complexity of study design and requirements.[24,25] These may be family studies, based on family units ranging from twins or groups of siblings through to extended pedigrees, or population-based studies such as cohort and case-control studies.

Box 1. Clinical outcomes of genetic epidemiology.[26]

- Improved understanding of the aetiological gene–environment interactions for many of the major conditions (for example, heart disease, diabetes and cancers).
- Increased genetic testing, from pre-conception onwards.
- Describing new taxonomies of pathophysiology and disease, based on molecular, classifications rather than signs and symptoms, with distinct information about prognosis and treatment.[17]
- Leading to targeted prevention and prognosis ('dia-prognosis')[12] based on genetic or molecular factors.
- Linking with pharmacogenetics, towards targeted therapeutic drug strategies.[29]
- Determining the heritability and famililial aggregation of diseases or intermediate phenotypes, and thus directing further gene discovery studies.[66–68]

CLINICAL OUTCOMES OF GENETIC EPIDEMIOLOGY

Although the ultimate goal of developing specific intervention strategies for genetic primary care conditions ('personalised medicine') is still a distant prospect,[26–30] there are outcomes of genetic epidemiology that are of more immediate interest, including the scientific gain of improved understanding of the biology of our species and the diseases we experience. Other outcomes have clinical relevance to primary care, and are summarised in Box 1.

For example, recent research has found that variants of the dopamine D_2 receptor determine responses to nicotine replacement therapy.[31] Women with the variant T allele, which had a prevalence of 41%, were significantly more likely to be successful in giving up smoking up to 8 years post-intervention with the support of nicotine patches; however, in men this allele did not seem to affect the response.[32] Similarly, polymorphisms in the cytochrome P450 CYP2C9 gene have been found to be associated with a fourfold increased risk of major haemorrhage when taking warfarin.[33] Such findings will enhance our ability to target pharmacological interventions, optimising effectiveness and cost-effectiveness, while minimising the risk of serious adverse reactions. This will require increased communication between GPs and their patients, who will need to discuss probabilities of response or adverse reactions based on the results of genetic tests, and agree decisions about prescribing.[28]

Genetic aetiological factors are not generally amenable to direct intervention, other than by pre-natal selection. This is largely because there is an inevitable lag between identifying the genetic basis of diseases and the development of new treatments or prevention strategies. For the present this will produce a dilemma in being able to identify increased risk when no action to reduce risk is yet possible. However, an immediate and important outcome for primary care will be the distinction of environmental from genetic risk factors for important diseases. We are more able to manipulate our environment, through drugs, lifestyle and behaviour, than our genes.

Furthermore, we can potentially identify genetic factors that distinguish true positive from false positive high risks of disease,[30] allowing a greater quantification and characterisation of risk, and more targeted disease prevention. For example, the great majority of individuals with hypertension do not go on to develop premature cardiovascular events: if we could identify a genetic factor that interacts cumulatively with this known risk factor, we could focus our antihypertensive efforts on those who have both, with consequent saving in time, drug costs and adverse drug reactions. The risk of breast cancer already offers an example of such targeting. A woman's family history of breast cancer is an imperfect predictor of her own future development of the disease. This is because of small family sizes, under-reporting of a family history of breast cancer (especially in second degree relatives), and incomplete penetrance of cancer-causing gene mutations.[34] Genetic screening, for example for mismatch repair gene mutations, may represent a better approach to targeting interventions to prevent cancer.[14,35] Discovering the BRCA1[36] and BRCA2[37] gene mutations and their importance in predicting breast cancer has led to the offer of prophylactic oophorectomy for women with a relevant family history and one of these mutations, a procedure that appears to lead to a gain of approximately 4 years of life.[38]

CONTRIBUTIONS OF PRIMARY CARE TO GENETIC EPIDEMIOLOGY RESEARCH

Early genetic epidemiological studies were generally based on small, highly selected samples such as those with multiple affected cases.[21] For example, the Breast Cancer Linkage Consortium studied such families with multiple cases of breast cancer.[39] These are useful for gene discovery studies, where the genetic contribution is specific or strong, or the trait under study is pronounced; that is, where there is a high 'signal-to-noise' ratio. Increasingly, however, the need for larger population studies is recognised, for three main reasons. First, association studies, which attempt to identify genes associated with specific traits, have often been underpowered, resulting in false negative results.[25,40] This is partly because the linkage between gene markers and specific genes may be relatively weak, partly because the prevalence of a gene polymorphism may be relatively rare or variable, and partly because the genetic contribution to the trait may be complex or relatively weak. Secondly, in order to examine the interaction between genetic and environmental factors, very large samples are required for most phenotypes to compensate for the background 'noise' produced by other, non-genetic factors.[41] Thirdly, the risk of a disease or trait associated with a particular genetic variant calculated from an unrepresentative population, selected perhaps for its high incidence or risk of this trait, cannot be extrapolated to the general population. Large, population-based samples are therefore required to calculate the absolute risks associated with any genetic variant. It is, we believe, important that these large studies such as Biobank relate closely to primary care, for several reasons which are considered below.

Doctor–patient relationships in primary care

There is an important difference between doing research *on* a population, and doing research *with* a population. The partnership implied by the latter is particularly important in genetic research, and GPs are among the best placed to facilitate this, acting as brokers between researchers and participants. Successful genetic epidemiology requires:

- clear public understanding of and support for the objectives of the research and the practical, medical, social, legal and ethical issues involved;
- recruitment of sufficient numbers of participants;
- consent to the use of routinely-held data for the purposes of this research, and the analysis and linkage of genetic information;
- high response rates at each stage of the study, including baseline surveys, venepuncture and long-term follow-up; and
- mechanisms to ensure continued public goodwill towards the research.

Continuing personal contact between potential subjects and informed, trusted professionals is therefore essential, together with secure informatics systems to allow recruitment and data management to proceed confidentially. General practice, based on personal relationships, with its existing and planned informatics structures,[42] seems to be the optimal environment for this to happen. Recruitment to genetic studies requires tact and sensitivity because of the social, ethical and legal issues involved. For example, the potential to uncover previously unidentified consanguinity, non-paternity, or newly-quantified disease risk must be addressed by the researchers in partnership with potential participants. A recent study found that GPs were by far the most trusted of a selection of professionals and public office-holders,[43] and this trust will be an important element in determining the choice about whether or not to participate. Mutual

loyalty and goodwill, over long periods of time, between GPs and their patients are essential for successful management of long-term illnesses; this is no less true for successful participation in longitudinal research. Basing the research in primary care provides a more favourable setting for presenting and discussing the study than is likely to be the case with studies based in other settings. However, public participation is not 'on tap' and should never be considered so. Trust, communication and relationships, just like informatics, are important components of the primary care clinical and research infrastructure, and the returns in participation rates will only match the investment of resources. This investment is an important first step in population-based genetic research, and as critical as obtaining the appropriate laboratory equipment (although probably more difficult to achieve). Equally, it is important to ensure, as far as possible, that involvement in research studies does not erode the trust that is required for clinical care. This highlights the need for ethically sound research design and governance, and fully-informed primary care professionals.

Previous family studies have benefited from close collaboration between primary care and communities, with high participation rates for several demanding types of project.[44] These studies, however, involved relatively small sample numbers, only a few locations, and did not include DNA analysis. Other developmental work in Scotland has found that almost all patients who were approached via their GP would agree to participate fully in a study such as Biobank, but would value the support of the practice during participation.[45] Primary care therefore remains largely untried in this context, and will require further careful developmental work in order to determine the most effective and satisfactory ways of working towards public engagement with genetic research, as opposed to clinical care.

Study design

Untargeted approaches to sampling, such as general advertising, or using routine databases (for example, the electoral roll, or the community health index), are possible, and may produce a willing group of participants. However, they risk a low yield and unwanted over-representation of certain characteristics, such as particular genetic diseases or personalities. Targeted approaches are more efficient,[46] and can use health-related records to identify the sample required (with careful adherence to ethical principles). These could be based on hospital disease-based cohorts such as those with congenital defects.[47] However, such samples are likely to be specific to one condition or group of conditions, restricted to one end of the disease spectrum, and present difficulty with the identification of control subjects that represent the general population.

Primary care in many countries, such as the UK and the Netherlands, offers access to the majority of the population with, increasingly, access to detailed longitudinal health records. This general population database, enhanced by primary care practitioners' knowledge of their individual patients and their families, is important for population studies because it:

- allows collection of phenotypic information such as disease endpoints (for example, ischaemic heart disease), known clinical risk factors (such as hyperlipidaemia), and prescribing information (such as statins);
- is relatively stable, allowing continued access for long-term longitudinal research, permitting continuing surveillance of the outcomes of suspected risk factors,[48] and genetic interaction with developing environmental factors; this also enhances the ability to trace members of subsequent generations;
- provides reliable information about the denominators as well as the numerators in epidemiological analysis, for descriptive epidemiology, and assessment of bias and confounding;
- provides detailed clinical and prescribing information in accessible format, for assessment of comorbidity, phenotyping and for pharmacogenetic and related studies;
- allows identification of the full illness spectrum of conditions under research, not only those attending hospital clinics;
- should ensure that the research questions addressed pertain to important clinical issues, relevant to daily practice, since GPs are in more frequent contact than any other health professionals with people who have the conditions under investigation; and
- includes, in particular, information on the many conditions that are managed only, or mainly in primary care, such as chronic pain[49] and other functional somatic syndromes.[50] Research on these conditions, if not based in primary care, would either be impractical or would provide a distorted picture by examining groups of patients that were atypical of individuals with the condition.[51]

In addition, for studies focusing on specific traits or diseases, the primary care database:

- allows identification of probands with specific diseases or traits, and some of their relatives;
- allows identification of control subjects matched for a large number of variables, maximising the power of studies and the ability to examine rarer conditions, as well as the common ones; and
- allows identification of family members, for family studies,[44] maximising the scope and power of potential studies,[52] and minimising bias due to ethnic or geographical stratification.[21]

It is important that access to this information is obtained only on the basis of the mutual understanding and consent discussed above, and that patients' trust is not exploited. To gain the most from primary care databases, the accuracy and standardisation of prospectively recorded clinical information is also crucial. This includes personal information, as well as continued morbidity recording throughout the research. The engagement of primary care professionals is therefore vital.

Added value from record linkage

If linked through networks of practices and research groups, primary care allows the generation of samples large enough for the population-based studies that are required to examine gene–environment interactions in the aetiology of complex disorders.[41] Throughout the UK, there is potential to work with the Medical Research Council General Practice Research Framework[53] to create a collaborative group of practices for genetic research. In Scotland, the intention is to work in a similar way with Scottish Practices and Professionals Involved in Research, a primary care research facilitation organisation that is funded by the Scottish Executive Health Department.[54]

Furthermore, there is the potential for linkage with other NHS data such as Hospital Episode Statistics and related datasets, with huge expansion of the research potential. In Scotland, it is possible to link routine data from any individual to anonymised personal data from numerous other clinical and related databases, including prescriptions, laboratory investigations, and hospital attendances and admissions, with, in some areas, more detailed routine information on diabetes, cardiovascular disease, asthma and strokes.

Disadvantages of primary care

This is a new area of work in primary care, and it will not be straightforward. Primary care lacks research culture and capacity,[55] and the difficulty of engaging GPs in research has been highlighted.[56] This engagement will be particularly important in genetic epidemiology research, for the reasons outlined above. Any planned research must devote the necessary resources and imagination to work productively with primary care professionals, who have many other competing calls on their time. This will include educational and financial rewards, as well as effective communication. Furthermore, the primary care infrastructure and databases, although better established than other potential research resources, are not perfect. The information held on clinical databases is only as accurate as the data that are entered. Attempts to check this accuracy have previously proven disappointing[57] and any interpretation of phenotype data drawn from routine data must therefore be cautious. Methods have been suggested for improving accuracy,[57] but would need widespread adoption before they can be applied generally. The General Medical Services contract agreed between GPs and the NHS in the UK now rewards practices for the recording of accurate clinical information, and should lead to more accurate research data as a result. This, however, only applies to certain clinical areas. Conversely, administration of practice systems in support of the new General Medical Services contract is complex, and likely to place further bureaucratic barriers between clinicians and researchers. Importantly there are currently no financial rewards inherent in the contract for participation in academic activity, including research.

Primary care is not, of course, the only source of routine data, and there are other methods of developing representative samples for genetic epidemiology study. For example, the Avon Longitudinal Study of Parents and Children (ALSPAC) was designed to examine gene–environment interactions in health and development, and has existed separately from primary care. It recruited mothers and their babies born in a specific period in a specific geographical area, using a combination of media advertising and health professionals (including primary and secondary care).[58,59] With continuing communication between the research team, participants and health professionals, a database of around 14 000 babies and their parents has been established, including a DNA bank, questionnaire information and linkage to medical and educational records. This has the potential to investigate the genetics of many conditions and outcomes,[59] and successes to date include the confirmation of a gene–environment interaction in the development of atopic eczema.[60] The ALSPAC cohort, although large, is insufficient for much of the research discussed above, and, as its success depends partly on local factors and engagement as well as an efficient

infrastructure, reproducing the cohort on a national scale would be difficult. However, there is much to learn from the methods that have led to this and others' success, and primary care could provide the infrastructure through which to develop this engagement. ALSPAC has engendered strong study loyalty, as have other community-based family studies.[44] This is something that primary care can do on a bigger scale, because of the 'cell' structure of primary care studies, with each practice a microcosm of strong relationships.

CONCLUSIONS

Genetic epidemiology has much to offer primary care, and the converse is also true. New genes and their variants associated with diseases and risk factors are being discovered, and this process will accelerate as a result of current and planned research. The next step is to characterise these genes with very large population-based samples, of the sort proposed in the studies discussed above. Basing this research in primary care is likely to bring many advantages that are not available to studies based in unselected or specialised groups of people. In particular, keeping close links with communities will foster better acceptance, trust and public understanding of the science and its potential. Involving GPs should promote this engagement and their own genetic literacy, and may thus improve the eventual uptake of new knowledge into practice. Academic primary care may have an important role to play in this context. Increasingly, as records are computerised and centralised, service GPs will need to play a third party role between researchers and patients. Patients expect their doctors to protect their confidentiality and not act as research agents. Doctors tread a dangerous path if seen in the latter role, and it is perhaps likely that patient expectations in this area will harden, not soften, with initiatives such as the new National Care Record System in England. Academic GPs can not only help to formulate research questions and designs which will work in primary care, but can also act as an intermediary between researchers and patients, for such issues as screening, and consent to being approached to take part for research on sensitive conditions. High participation rates may depend on there being sufficient provision for this intermediary role.

This is a new scientific frontier, where primary care still has an unproven track record yet seems to offer the best starting point. We must take the opportunity to explore it, but must do this sensitively and in a partnership between academic and service GPs and our patients.

Funding body

Blair H Smith was supported by a NHS R&D Primary Care Career Scientist Award, funded by the Scottish Executive Health Department

Competing interests

The authors are closely involved with planning for the Generation Scotland and UK Biobank (Scottish Regional Collaborating Centre) studies referred to in the paper

Acknowledgements

The authors are grateful to the Scottish School of Primary Care for their support in the collaboration that produced this paper.

REFERENCES

1. Austin MA, Harding S, McElroy C. Genebanks: a comparison of eight proposed international genetic databases. *Community Genet* 2003; **6:** 37–45.
2. Khoury MJ, Millikan R, Little J, Gwinn M. The emergence of epidemiology in the genomics of age. *Int J Epidemiol* 2004; **33:** 936–944.
3. Anderson MW. A 500 000-person study? Gene-environment interactions would be focus of NIH-led effort. http://www.the-scientist.com/news/20040526/04/ (accessed 3 Feb 2006).
4. Shpilberg O, Dorman JS, Ferrell RE, *et al*. The next stage: molecular epidemiology. *J Clin Epidemiol* 1997; **50:** 633–638.
5. Berger A. UK genetics database plans revealed. *BMJ* 2001; **322:** 1018.
6. Biobank. UK Biobank: Improving the health of future generations. http://www.ukbiobank.ac.uk/ (accessed 25 Jan 2006).
7. Generation Scotland. Addressing the health and wealth of Scotland. http://www.generationscotland.org (accessed 25 Jan 2006).
8. Eesti Geenivaramu. Homepage. http://www.geenivaramu.ee/ (accessed 25 Jan 2006)
9. deCode. deCode genetics. http://www.decode.is/ (accessed 25 Jan 2006).
10. Emery J, Watson E, Rose P, Andermann A. A systematic review of the literature exploring the role of primary care in genetic services. *Fam Pract* 1999; **16:** 426–445.
11. Emery J, Hayflick S. The challenge of integrating genetic medicine into primary care. *BMJ* 2001; **322:** 1027–1030.
12. Knottnerus JA. Community genetics and community medicine. *Fam Pract* 2003; **20:** 601–606.
13. Qureshi N, Modell B, Modell M. Raising the profile of genetics in primary care. *Nat Rev Genet* 2004; **5:** 783–790.
14. Mitchell RJ, Farrington SM, Dunlop MG, Campbell H. Mismatch repair genes hMLH1 and hMSH2 and colorectal cancer: a HuGE review. *Am J Epidemiol* 2002; **156:** 885–902.
15. Antoniou A, Pharoah PD, Narod S, *et al*. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet* 2003; **72:** 1117–1130.
16. Ralston SH. Genetic determinants of susceptibility to osteoporosis. *Curr Opin Pharmacol* 2003; **3:** 286–290.
17. Bell J. The new genetics in clinical practice. *BMJ* 1998; **316:** 618–620.
18. Claus EB, Risch N, Thompson WD. Genetic analysis of breast cancer in the cancer and steroid hormone study. *Am J Hum Genet* 1991; **48:** 232–242.
19. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. *Cell* 1996; **87:** 159–170.
20. Khoury MJ. Genetic epidemiology. In: Rothman KJ, Greenland S, *Modern epidemiology.* 2nd edn. Philadelphia: Lippincott-Raven, 1998: 609–622.
21. Khoury MJ, Beaty TH, Cohen BH. *Fundamentals of genetic epidemiology. A scientific foundation for using genetic information to improve health and prevent disease.* Oxford: Oxford University Press, 1993.
22. Davey Smith G, Ebrahim S. 'Mendelian randomisation': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; **32:** 1–22.
23. Khoury MJ, Little J, Burke W (eds). *Human genome epidemiology.* New York: Oxford University Press, 2004.

24. Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet* 2001; **358:** 1356–1360.
25. Colhoun H, McKeigue PM. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003; **361:** 865–872.
26. Collins FS, McKusick VA. Implications of the human genome project for medical science. *JAMA* 2001; **285:** 540–544.
27. Aitman TJ. DNA microarrays in medical practice. *BMJ* 2001; **323:** 611–615.
28. Hapgood R. The potential and limitations of personalised medicine in primary care. *Br J Gen Pract* 2003; **53:** 915–916.
29. Tucker G. Pharmacogenetics — expectations and reality. *BMJ* 2004; **329:** 4–6.
30. Watt GCM. What will the new genetic information do for us? *J Health Serv Res Policy* 2004; **9:** 186–188.
31. Johnstone EC, Yudkin PL, Hey K, *et al.* Genetic variation in dopaminergic pathways and short-term effectiveness of the nicotine patch. *Pharmacogenetics* 2004; **14:** 83–90.
32. Yudkin P, Munafo M, Hey K, *et al.* Effectiveness of nicotine patches in relation to genotype in women versus men: randomised controlled trial. *BMJ* 2004; **328:** 989–990.
33. Aithal G, Day C, Keseven P, Daly A. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and the risk of bleeding complications. *Lancet* 1998; **353:** 717–719.
34. Mitchell RJ, Brewster D, Campbell H, *et al.* Accuracy of reporting of family history of colorectal cancer. *Gut* 2004; **53:** 291–295.
35. Hampel H, Frankel WL, Martin E, *et al.* Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *N Engl J Med* 2005; **352:** 1851–1860.
36. Miki Y, Swensen J, Shattuck-Eidens D, *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRACT. *Science* 1994; **266:** 66–71.
37. Wooster R, Bignell G, Lancester J, *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature* 1995; **378:** 789–792.
38. Armstrong K, Scwartz JS, Randall T, *et al.* Hormone replacement therapy and life expectancy after prophylactic oophorectomy in women with BRCA1/2 mutations: a decision analysis. *J Clin Oncol* 2004; **22:** 978–980.
39. Easton D, Bishop D, Ford D, Croskford G. Genetic linkage analysis in familial breast and ovarian cancer results from 214 families. *Am J Hum Genet* 1993; **52:** 678–701
40. Ioannidis JP, Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG. Genetic associations in large versus small studies: an empirical assessment. *Lancet* 2003; **361:** 567–571.
41. Wright A, Carothers A, Campbell H. Gene environment interaction: the BioBank UK Study. *Pharmacogenomics J* 2002; **2:** 75–82.
42. Sullivan FM, Pell JP, Sweetland M, Morris AD. How could primary care meet the informatics needs of UK Biobank? A Scottish proposal. *Inform Prim Care* 2003; **11:** 129–135.
43. Hayward B, Mortimer E, Brunwin T. *Survey of public attitudes towards conduct in public life.* London: Committee on Standards in Public Life, 2004. www.public-standards.gov.uk/research/researchreport.pdf (accessed 25 Jan 2006).
44. Watt GCM. General practice and the epidemiology of health and disease in families [Pickles lecture]. *Br J Gen Pract* 2004; **54:** 939–944.
45. Marsden W, Duffy R, Sullivan F. *A short report on the recruitment potential of two Scottish primary care trust areas to UK Biobank.* Dundee: TayRen/Fresco, 2002.
46. Smith BH, Hannaford PC, Elliott AM, *et al.* The 'number needed to sample' in primary care research. Comparison of two primary care sampling frames for chronic back pain. *Fam Pract* 2005; **22:** 205–214.
47. Correa-Villasenor, Cragan J, Krucik J, *et al.* The Metropolitan Atlanta Congenital Defects Program: 35 years of birth defects surveillance at the Centers for Disease Control and Prevention. *Birth Defects Res A Clin Mol Teratol* 2003; **67:** 617–624.
48. Walker M, Shaper AG, Lennon L, Whincup P. Twenty year follow-up of a cohort based in general practices in 24 British towns. *J Public Health Med* 2000; 22: 479–485.
49. Smith BH. Chronic pain: a challenge for primary care. *Br J Gen Pract* 2001; **51:** 524–526.
50. Wessely C, Nimnuan C, Sharpe M. Functional somatic syndromes: one or many? *Lancet* 1999; **354:** 936–939.
51. Crombie IK, Davies HTO. Selection bias in pain research. *Pain* 1998; **74:** 1–3.
52. Campbell H. Rudan I. Interpretation of genetic association studies in complex disease. *Pharmacogenomics J* 2002; **2:** 349–360.
53. The Medical Research Council General Practice Research Framework. Homepage. http://www.mrc-gprf.ac.uk/ (accessed 25 Jan 2006).
54. The Scottish School of Primary Care. Scottish Practices and Professionals Involved in Research (SPPIRe). http://www.show.scot.nhs.uk/sspc/sppire/sppire1.asp (accessed 25 Jan 2006).
55. Department of Health. *R&D in primary care national working group report.* London: Department of Health, 1997.
56. Silagy C, Carson N. Factors affecting the level of interest in activity in primary care research among general practitioners. *Fam Pract* 1989; **6:** 173–176.
57. Ward L, Innes M. Electronic medical summaries in general practice — considering the patient's contribution. *Br J Gen Pract* 2003; **53:** 293–297.
58. Golding J, Pembrey M, Jones R, the ALSPAC Study Team. ALSPAC — the Avon Longitudinal Study of Parents and Children. I. Study methodology. *Paediatr Perinat Epidemiol* 2001; **15:** 74–87.
59. Pembrey M, ALSPAC Study Team. The Avon Longitudinal Study of Parents and Children (ALSPAC): a resource for genetic epidemiology. *Eur J Endocrinol* 2004; **151(Suppl 3):** U125-129.
60. Callard RE, Hamvas R, Chatterton C, *et al.* An interaction between the IL-4R gene and infection with atopic eczema in young children. *Clin Exp Allergy* 2002; **32:** 990–994.
61. CARTaGENE. A genetic map of Quebec. http://www.cartagene.qc.ca/ (accessed 25 Jan 2006).
62. Chen Z, Lee L, Chen J, et al. Cohort profile: The Kadoorie Study of Chronic Disease in China (KSCDC). *Int J Epidemiol* 2005; **34:** 1243–1249.
63. Putnina A. Exploring the articulation of agency: population genome project in Latvia. www.ifz.tugraz.at/index_en.php/filemanager/download/122/putnina.pdf (accessed 3 Feb 2005).
64. Uman Genomics. Homepage. http://www.umangenomics.com/ (accessed 25 Jan 2006).
65. Genom EU Twin. Studies of European volunteer twins to identify genes underlying common diseases. http://www.genomeutwin.org/ (accessed 25 Jan 2006).
66. Noon JP, Walker BR, Webb DJ, *et al.* Impaired microvascular dilatation and capillary rarefaction in young adults with a predisposition to high blood pressure. *J Clin Invest* 1997; **99:** 1873–1879.
67. Harrap SB, Fraser R, Inglis GC, *et al.* Abnormal epinephrine release in young adults with high personal and high parental blood pressures. *Circulation* 1997; **96:** 556–561.
68. Harrap SB, Cumming AD, Davies DL, *et al.* Glomerular hyperfiltration, high renin and low-extracellular volume in high blood pressure. *Hypertension* 2000; **35:** 952–957.