BMJ

Accuracy of diagnosing atrial fibrillation on electrocardiogram by primary care practitioners and interpretative diagnostic software: analysis of data from screening for atrial fibrillation in the elderly (SAFE) trial

Jonathan Mant, reader,[1] David A Fitzmaurice, professor of primary care,[1] F D Richard Hobbs, professor and head of department,[1] Sue Jowett, research fellow,[2] Ellen T Murray, research fellow,[1] Roger Holder, head of statistics,[1] Michael Davies, consultant cardiologist,[3] Gregory Y H Lip professor of cardiovascular medicine[4]

[1]Department of Primary Care and General Practice, University of Birmingham, Birmingham B15 2TT

[2]Health Economics Facility, Health Services Management Centre, University of Birmingham, Birmingham B15 2RT

[3]Department of Cardiology, Queen Elizabeth Hospital, Edgbaston, Birmingham B15 2TH

[4]University Department of Medicine, City Hospital, Birmingham B18 7QH

Correspondence to: D A Fitzmaurice d.a. fitzmaurice@bham.ac.uk

ABSTRACT

Objective To assess the accuracy of general practitioners, practice nurses, and interpretative software in the use of different types of electrocardiogram to diagnose atrial fibrillation.

Design Prospective comparison with reference standard of assessment of electrocardiograms by two independent specialists.

Setting 49 general practices in central England.

Participants 2595 patients aged 65 or over screened for atrial fibrillation as part of the screening for atrial fibrillation in the elderly (SAFE) study; 49 general practitioners and 49 practice nurses.

Interventions All electrocardiograms were read with the Biolog interpretative software, and a random sample of 12 lead, limb lead, and single lead thoracic placement electrocardiograms were assessed by general practitioners and practice nurses independently of each other and of the Biolog assessment.

Main outcome measures Sensitivity, specificity, and positive and negative predictive values.

Results General practitioners detected 79 out of 99 cases of atrial fibrillation on a 12 lead electrocardiogram (sensitivity 80%, 95% confidence interval 71% to 87%) and misinterpreted 114 out of 1355 cases of sinus rhythm as atrial fibrillation (specificity 92%, 90% to 93%). Practice nurses detected a similar proportion of cases of atrial fibrillation (sensitivity 77%, 67% to 85%), but had a lower specificity (85%, 83% to 87%). The interpretative software was significantly more accurate, with a specificity of 99%, but missed 36 of 215 cases of atrial fibrillation (sensitivity 83%). Combining general practitioners' interpretation with the interpretative software led to a sensitivity of 92% and a specificity of 91%. Use of limb lead or single lead thoracic placement electrocardiograms resulted in some loss of specificity.

Conclusions Many primary care professionals cannot accurately detect atrial fibrillation on an electrocardiogram, and interpretative software is not sufficiently accurate to circumvent this problem, even when combined with interpretation by a general practitioner. Diagnosis of atrial fibrillation in the community needs to factor in the reading of electrocardiograms by appropriately trained people.

INTRODUCTION

Atrial fibrillation is an important risk factor for stroke and is present in about 5% of people over the age of 65.[1 2] It can be diagnosed by a simple investigation—electrocardiography—and treatment with anticoagulation can substantially reduce the risk of stroke.[3] Many electrocardiograms are now generated and read in primary care, whether by a general practitioner, a practice nurse, or interpretative software. However, little research has been done into the extent to which the type of reader affects the accuracy with which atrial fibrillation is detected by electrocardiography. A systematic review of studies of interpretation of electrocardiograms found that physicians of all specialties made frequent errors when interpreting electrocardiograms, but none of the 41 studies identified focused specifically on atrial fibrillation.[4] An evaluation of a computer software algorithm found that it could detect atrial fibrillation with a sensitivity of 91% and a specificity of 99%.[5] A study in one general practice found that the general practitioner could detect atrial fibrillation accurately (sensitivity 100%, specificity 98%), but this result cannot be generalised to all primary care physicians.[6]

A subsidiary question to the reliability of interpretation is whether a full 12 lead electrocardiogram is needed if the purpose of the investigation is simply to diagnose atrial fibrillation. The potential advantage of doing limited electrocardiograms, such as single chest lead or just the limb leads, is that they are simpler, quicker procedures that need less undressing of the patient.

The aim of this study was to assess the accuracy with which general practitioners, practice nurses, and interpretative software diagnose atrial fibrillation on 12 lead electrocardiograms, single lead thoracic placement electrocardiograms, and limb lead recordings.
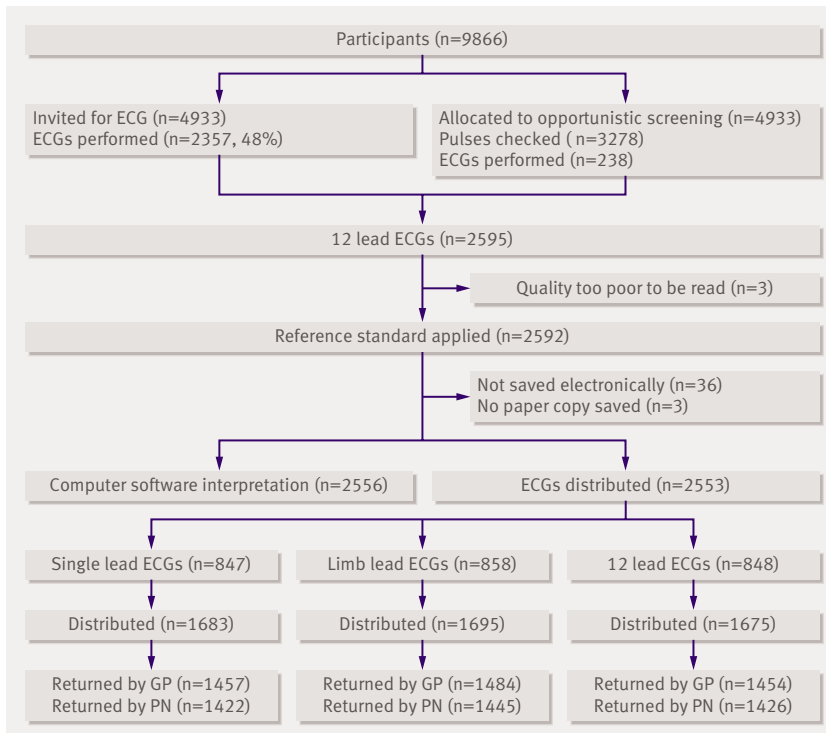
Fig 1 | Flowchart of generation and reading of electrocardiograms. ECG=electrocardiogram; GP=general practitioner; PN=practice nurse. *Each ECG was sent to two practices, except for one batch that was sent to only one practice

METHODS

We did the study as a prospective sub-study within the screening for atrial fibrillation in the elderly (SAFE) randomised controlled trial of different methods of screening for atrial fibrillation in primary care funded by the Health Technology Assessment Programme.[7] This involved 50 practices in central England, which were randomly allocated as 25 intervention practices, in which a screening programme was initiated, and 25 control practices. One general practitioner and one practice nurse from each practice was involved in the study. The training of practitioners in the intervention practices included a one hour session on how to interpret an electrocardiogram and, in particular, how to detect atrial fibrillation. Practitioners in the control practices received no training.

Generation of electrocardiograms

We selected a random sample of 9866 people aged 65 or over from the 25 SAFE "intervention" practices. We invited a random half of these people to attend the practice for an electrocardiogram and invited the remaining half only if opportunistic screening identified them as having an irregular pulse. This generated 2595 12 lead electrocardiograms, including 238 from opportunistic screening during 2001-3. Digital machines (Biolog, manufactured by Numed, Sheffield) were used to do all the electrocardiograms, and the data were sent electronically to the study centre.

At the end of the SAFE study, approximately three years after the initial training session, we printed out all the electrocardiograms as 12 lead electrocardiograms, a random third of them as single thoracic placement electrocardiograms, and a random third as limb lead electrocardiograms (fig 1). We assembled at random 25 batches of approximately 100 electrocardiograms, comprising a third each of 12 lead, single thoracic placement, and limb lead electrocardiograms, and distributed them to the 49 practices (one practice had elected not to participate in this sub-study), where they were read by one general practitioner and one practice nurse in each practice. We sent each batch (except one) to one intervention practice and one control practice.

Reading of electrocardiograms

We interpreted all the electrocardiograms (as a 12 lead) with the Biolog interpretative software. We asked each participant to indicate on a form whether or not atrial fibrillation or atrial flutter was present in each case. Practitioners were blinded to patients' identities, the diagnoses made by the specialists, and the diagnoses generated by the interpretative software.

Reference standard

Two consultant cardiologists, blinded to the software interpretation and that of the primary care practitioners, read all the 12 lead electrocardiograms independently of each other. If the cardiologists disagreed, then a third consultant cardiologist arbitrated.

Statistical methods

We treated sensitivity, specificity, and positive and negative predictive values as binomial proportions and calculated exact 95% confidence intervals accordingly. We used logistic regression to examine variation in sensitivity and specificity both with the type of electrocardiogram (single lead, limb lead, and 12 lead) and between control and intervention practices. For comparison of sensitivity and specificity between types of electrocardiogram and between general practitioners and nurses, we did both matched and unmatched analyses. We give corresponding pairs of P values where appropriate, with the unmatched P value in parentheses.

RESULTS

We judged three of the electrocardiograms to be of insufficient quality to be read by the cardiologists. Two cardiologists read the remainder. For seven (0.27%) electrocardiograms, the cardiologists disagreed on the presence of atrial fibrillation, and a third cardiologist made the decision. Forty two (86%) primary care physicians and 41 (84%) nurses returned the results of their interpretation.

Tables 1, 2, and 3 show the results for interpretation of the different types of electrocardiogram, and table 4 provides the 95% confidence intervals for these results. The prevalence of atrial fibrillation was 8.4%. Interpretative software was the most accurate method of reading electrocardiograms but did not give a rhythm diagnosis in 109 (4.3%) and missed 26 (12%) cases of

Table 1 | Detection of atrial fibrillation on 12 lead electrocardiogram

	Reference standard		
	Atrial fibrillation	Not atrial fibrillation	Total
General practitioner			
Atrial fibrillation	79	114	193
Not atrial fibrillation	20	1239	1259
Uncertain	0	2	2
Total	99	1355	1454
Primary care nurse			
Atrial fibrillation	74	198	272
Not atrial fibrillation	22	1127	1149
Uncertain	0	5	5
Total	96	1330	1426
Interpretative software			
Atrial fibrillation	179	21	200
Not atrial fibrillation	26	2221	2247
Uncertain	10	99	109
Total	215	2341	2556
General practitioner and interpretative software			
Both positive	71	3	74
Only software positive	12	7	19
Only general practitioner positive	8	111	119
Both negative	8	1234	1242
Total	99	1355	1454

atrial fibrillation—or 36 (17%) if we include the cases in which no rhythm diagnosis was made. Ten per cent of computer diagnoses of atrial fibrillation were incorrect. The combined sensitivity of general practitioner and interpretative software was 92%, and the specificity was 91%.

General practitioners and practice nurses detected similar proportions of cases of atrial fibrillation (80% v 77% on 12 lead electrocardiogram), but diagnosis by general practitioners was more specific. Nevertheless, a diagnosis of atrial fibrillation by a general practitioner was still more likely to be wrong than right (positive predictive value 40.9%). Use of 12 lead, limb lead, or single thoracic placement electrocardiograms made little difference to the ability of primary care

Table 2 | Detection of atrial fibrillation on limb lead electrocardiogram

	Reference standard		
	Atrial fibrillation	Not atrial fibrillation	Total
General practitioner			
Atrial fibrillation	104	156	260
Not atrial fibrillation	22	1194	1216
Uncertain	0	8	8
Total	126	1358	1484
Primary care nurse			
Atrial fibrillation	85	220	305
Not atrial fibrillation	31	1095	1126
Uncertain	2	12	14
Total	118	1327	1445

Table 3 | Detection of atrial fibrillation on single lead thoracic placement electrocardiogram

	Reference standard		
	Atrial fibrillation	Not atrial fibrillation	Total
General practitioner			
Atrial fibrillation	112	180	292
Not atrial fibrillation	19	1141	1160
Uncertain	1	4	5
Total	132	1325	1457
Primary care nurse			
Atrial fibrillation	92	222	314
Not atrial fibrillation	42	1060	1102
Uncertain	0	6	6
Total	134	1288	1422

practitioners to correctly diagnose the presence of atrial fibrillation (P=0.52, matched (0.52, unmatched) for general practitioners; P=0.08 (0.35) for nurses). A significant difference occurred when diagnosing the absence of atrial fibrillation, however; 12 lead electrocardiograms gave a better outcome for general practitioners (P<0.001 (<0.001); P=0.12 (0.23) for nurses).

No significant difference existed between the performance of general practitioners from intervention and control practices. Control general practitioners showed a sensitivity of 84.0% and a specificity of 88.1%, and intervention general practitioners showed a sensitivity of 81.3% and a specificity of 88.9% (P=0.57 for sensitivity; P=0.19 for specificity). However, practice nurses from intervention practices interpreted electrocardiograms more accurately than did those from control practices: sensitivity 76.5% versus 68.9%; specificity 88.9% versus 78.9% (P=0.11 for sensitivity; P<0.001 for specificity).

The ability of individual general practitioners and practice nurses to diagnose atrial fibrillation accurately on an electrocardiogram varied widely. Figures 2 and 3

Table 4 | Summary statistics of accuracy of interpretation of electrocardiograms by reader and type

Reader and electrocardiogram type	Sensitivity (95% CI)	Specificity (95% CI)
Interpretative software:		
12 lead	83.3 (78.3 to 88.2)	99.1 (98.7 to 99.5)
General practitioner:		
12 lead	79.8 (70.5 to 87.2)	91.6 (90.1 to 93.1)
Limb lead	82.5 (74.8 to 88.7)	88.5 (86.9 to 90.2)
Chest lead	84.8 (78.7 to 91.0)	86.4 (84.6 to 88.3)
Practice nurse:		
12 lead	77.1 (67.4 to 85.0)	85.1 (83.0 to 86.9)
Limb lead	72.0 (63.9 to 80.1)	83.4 (81.4 to 85.4)
Chest lead	68.7 (60.1 to 76.4)	82.8 (80.7 to 84.8)
General practitioner and interpretative software*:		
12 lead	91.9 (86.6 to 97.3)	91.1 (89.6 to 92.6)

For the purposes of calculations for this table, "uncertain" diagnoses are counted as missed diagnoses for sensitivity and as not atrial fibrillation for specificity (see tables 1-3).
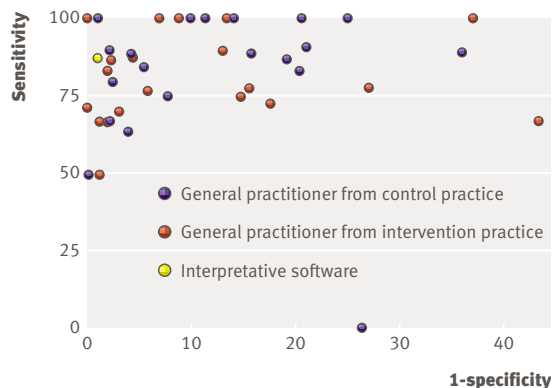*Defined as positive result if either (or both) is positive.

Fig 2 | Accuracy of diagnosis of atrial fibrillation by 42 general practitioners

show the sensitivity and rate of false positives when the results from the different types of electrocardiogram are combined. The sensitivity of individual doctors varied from 50% to 100% and that of practice nurses from 0% to 100%; the standard deviations of individual sensitivities were 31% for general practitioners and 37% for nurses. The false positive rate of general practitioners varied from 0% to 44% and that of practice nurses from 0% to 61%, and the respective standard deviations were 13% and 17%. Two general practitioners and two practice nurses performed better than the computer software. Most of the outlying poorly performing practice nurses were from control practices.

DISCUSSION

In this study, general practitioners were unable to diagnose atrial fibrillation accurately on an electrocardiogram. Twenty per cent of cases of atrial fibrillation were missed, and the probability that a positive diagnosis was correct was only 41%. Changing from the 12 lead to simpler electrocardiograms resulted in further loss of specificity.

Our results are substantially different from those reported by Somerville and colleagues, who found
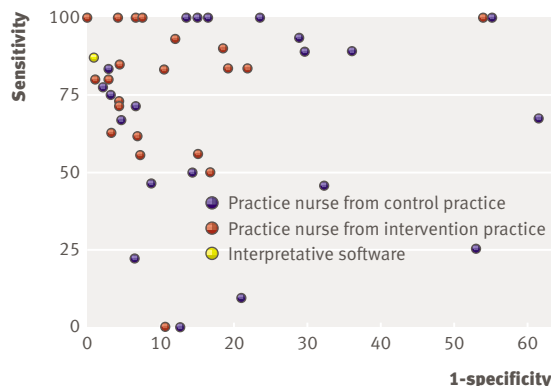

Fig 3 | Accuracy of diagnosis of atrial fibrillation by 41 practice nurses

that a general practitioner could detect atrial fibrillation accurately.[6] Some doctors in our study performed as well as the one in that study, but our results suggest that such performance is atypical among general practitioners. Practice nurses were less accurate than the doctors, and interpretative software was more accurate than the doctors. The performance of the interpretative software was similar to that reported by Poon and colleagues in an analysis of 4297 electrocardiograms in a secondary care setting, where the prevalence of atrial fibrillation was 6%.[5]

The generally lower P values on the matched analyses reflect the variation in performance between the raters. This variation was greater for practice nurses than it was for general practitioners, reflecting a greater range of ability.

Strengths and weaknesses of study

The electrocardiograms being read for this study were generated as part of a screening programme and so reflect the sort of electrocardiograms that primary care practitioners would need to read if screening for atrial fibrillation. They had an appropriate prevalence of atrial fibrillation, so our estimates of predictive value are directly applicable to this screening context. Previous studies of the interpretive ability of general practitioners have tended to focus on a few practitioners or a few electrocardiograms.[6 8 9] A strength of this study is the large number of practitioners and electrocardiograms involved. Previous studies have tended to use a single cardiologist as a "reference standard" for detecting atrial fibrillation on an electrocardiogram.[6 10] In this study, we used two consultant cardiologists, with a third arbitrating as necessary. In fact, the agreement between the cardiologists was very high (over 99%), confirming that diagnosis of atrial fibrillation can be made reliably through the reading of an electrocardiogram by a physician with relevant training and experience.

The primary care practitioners in this study were recruited from practices active in research and had volunteered to take part in a trial of screening for atrial fibrillation. As such, one might anticipate their ability to detect atrial fibrillation on an electrocardiogram to be better than the average practitioner. On the other hand, they were asked to read the electrocardiograms in artificial circumstances—a primary care practitioner would not normally be sent 100 electrocardiograms to read—so it may be that the electrocardiograms were not read as carefully as they would have been in a clinical situation.

The response rate was reasonably high; 84-86% of participants returned their electrocardiograms. Nonrespondents may have been less accurate at interpreting the electrocardiograms, which would strengthen the general conclusion of this study.

The circumstances were also artificial in that the primary care practitioners did not have access to the other clinical information (symptoms, signs, and software interpretation) that they would usually have when interpreting an electrocardiogram. However,

WHAT IS ALREADY KNOWN ON THIS TOPIC

Electrocardiography is recognised as the standard investigation for diagnosing atrial fibrillation

Little evidence exists as to whether primary care practitioners can reliably diagnose atrial fibrillation on an electrocardiogram

WHAT THIS STUDY ADDS

Although a few primary care practitioners could diagnose the presence or absence of atrial fibrillation accurately on an electrocardiogram, most could not

The accurate identification of atrial fibrillation in the community requires that electrocardiograms are read by appropriately trained people

symptoms and signs are of limited use. The most common symptom of atrial fibrillation is palpitations, but these are present in only half of people with atrial fibrillation.[11] Although palpation of an irregular pulse is reasonably sensitive (93-100%), it has a very low positive predictive value (8-23%).[12] If the prevalence of atrial fibrillation was 20%, as might be expected in people with an irregular pulse, this would raise the predictive value of a positive reading of the electrocardiogram by a general practitioner to 66% (assuming constant sensitivity and specificity of interpretation of electrocardiograms). Therefore, although the high sensitivity of pulse palpation might compensate for the low sensitivity of detection of atrial fibrillation on the electrocardiogram by primary care practitioners, the practitioners will still not be able to adequately discriminate on the electrocardiogram between those people with an irregular pulse who do and do not have atrial fibrillation.

Combining the results of the interpretative software with interpretation by a general practitioner led to some improvement in sensitivity, but at a cost of lower specificity. Therefore, although the lack of access of the study practitioners to the character of the pulse and the software interpretation might have led to an underestimate of sensitivity of interpretation of the electrocardiogram in the real world setting, we may have overestimated specificity. Furthermore, the cardiologists achieved their very high agreement without access to any clinical data or the electrocardiogram software.

Implications

In most cases in which an arrhythmia is diagnosed by a general practitioner, a specialist opinion is recommended.[13] If more than half of diagnoses of atrial fibrillation in primary care are incorrect, this might lead to a lot of unnecessary referrals. Conversely, if a screening programme relied on reading electrocardiograms in primary care, about 20% of cases of atrial fibrillation would be missed and therefore be untreated. Computer software performed much better, but still had an error rate sufficiently high to mean that decisions on treatment cannot be based on diagnosis by computer alone, even when combined with interpretation by a general practitioner. Therefore, strategies to identify atrial fibrillation in the community,

whether through population screening or for diagnosis of patients with symptoms, need to take into account how and by whom the electrocardiogram will be interpreted.

If primary care practitioners are to detect atrial fibrillation on an electrocardiogram reliably, they need appropriate training, accreditation, or both. We could not test in this study whether training would improve the accuracy with which primary care practitioners could read electrocardiograms. The training that was provided to practitioners in intervention practices was done to support the SAFE study, rather than specifically to test the efficacy of training. The training was given three years before we asked practitioners to read the electrocardiograms, and because of changes in personnel not all the intervention practitioners who read the electrocardiograms had received training. Despite this, the training did seem to have a sustainable effect on practice nurses, although not to a sufficient level to enable the electrocardiograms to be read safely. Alternatively, electrocardiograms generated in primary care will need to be sent to a specialist for accurate interpretation. The electronic transfer of 2595 electrocardiograms to a central storage facility and onward to specialists was an efficient mechanism in the SAFE study and could be replicated in clinical practice.

In England and Wales, general practitioners are now required to set up registers of patients with atrial fibrillation as part of the quality and outcomes framework.[14] The recent guideline for atrial fibrillation from the National Institute for Health and Clinical Excellence recommends that electrocardiography is used to diagnose atrial fibrillation, but it makes no recommendations about the reading of the electrocardiogram.[11] This study suggests that quality control of interpretation of electrocardiograms is an important part of diagnosing atrial fibrillation in primary care.

1 Lip GY, Boos CJ. Antithrombotic treatment in atrial fibrillation. *Heart* 2006;92:155-61.
2 Sudlow S, Thomson R, Thwaites B, Rodgers H, Kenny RA. Prevalence of atrial fibrillation and eligibility for anticoagulants in the community. *Lancet* 1998;352:1167-71.
3 Lip GYH, Edwards SJ. Stroke prevention with aspirin, warfarin and ximelagatran in patients with non-valvular atrial fibrillation: a systematic review and meta-analysis. *Thromb Res* 2006;118:321-33.

4   Salerno SM, Alguire PC, Waxman HS. Competency in interpretation of 12-lead electrocardiograms: summary and appraisal of published evidence. *Ann Intern Med* 2003;138:751-60.

5   Poon K, Okin P, Kligfield P. Diagnostic performance of a computer based ECG rhythm algorithm. *J Electrocardiol* 2005;38:235-8.

6   Somerville S, Somerville J, Croft P, Lewis M. Atrial fibrillation: a comparison of methods to identify cases in general practice. *Br J Gen Pract* 2000;50:727-9.

7   Hobbs FDR, Fitzmaurice DA, Mant J, Murray E, Jowett S, Bryan S, et al. A randomised controlled trial and cost-effectiveness study of systematic screening (targeted and total population screening) versus routine practice for the detection of atrial fibrillation in people aged 65 and over: the SAFE study. *Health Technol Assess* 2005;9:40.

8   Houghton AR, Sparrow NJ, Toms E, Cowley AJ. Should general practitioners use the electrocardiogram to select patients with suspected heart failure for echocardiography? *Int J Cardiol* 1997;62:31-6.

9   Sur DKC, Kaye L, Mikus M, Goad J, Morena A. Accuracy of electrocardiogram reading by family practice residents. *Fam Med* 2000;32:315-9.

10 Wheeldon NM, Tayler DI, Anagnostou E, Cook D, Wales C, Oakley GDG. Screening for atrial fibrillation in primary care. *Heart* 1998;79:50-5.

11 National Collaborating Centre for Chronic Conditions. *Atrial fibrillation: national clinical guideline for management in primary and secondary care*. London: Royal College of Physicians, 2006.

12 Sudlow M, Rodgers H, Kenny RA, Thomson R. Identification of patients with atrial fibrillation in general practice: a study of screening methods. *BMJ* 1998;317:327-8.

13 Department of Health. *National service framework for coronary heart disease. Chapter 8: Arrhythmias and sudden cardiac death*. London: DH, 2005.

14 BMA/NHS Employers. Revisions to the GMS contract 2006/07. www.nhsemployers.org/primary/primary-886.cfm.