



# Faculty of Public Health

of the Royal Colleges of Physicians of the United Kingdom

Working to improve the public's health

## **PART A EXAMINATION FOR MEMBERSHIP OF THE FACULTY OF PUBLIC HEALTH**

*Of the Royal Colleges of Physicians of the United Kingdom*

**JANUARY 2008**

**EXAMINATION QUESTIONS WITH KEY POINTS AND  
EXAMINERS' COMMENTS**

**N.B. Please note that these are key points, not model  
answers**

## Paper IA

### Question 1

List and define measures which may be used to describe relative and absolute differences in rates of disease between two groups of subjects in prospective studies.

### KEY POINTS

*Most of all of the following would be required for a pass:*

Prospective studies: cohort or randomised controlled trials

If  $p_1$  = proportion of subjects acquiring disease (or outcome) in group 1 (exposed) (similarly  $p_2$  for unexposed)

Absolute risk reduction or risk difference =  $p_1 - p_2$  (also, sometimes, called attributable risk)

Relative risk (risk ratio) =  $p_1/p_2$   
(or transpose  $p_1, p_2$ )

To take account of varying follow up times in cohort studies, use  
Rate ratio = no. events in group 1 divided by the number of person years at risk in group 1 / same in group 2.

Relative risk reduction =  $1 - RR$  (if  $RR < 1$ )

NNT =  $1/\text{absolute risk difference}$  – this is number needed to treat for one person to benefit – cannot be negative.

Can also have “Number Needed to Harm” (NNH) – to describe adverse events.

Odds Ratio =  $[p_1(1-p_2)]/[p_2(1-p_1)]$  may be hard to interpret directly; a more appropriate measure for case-control studies.

*The following are additional points which might improve the answer to “good” or “excellent”:*

Hazard ratio – comparison of time to event data/survival analysis – in RCT

Population attributable risk and population attributable risk fraction  
 $PARF = (\text{risk in whole study population} - \text{risk in unexposed}) / \text{risk in whole study population}$ . Need to recognise assumptions made.

Confidence intervals should be calculated for any statistic listed, though very complicated to interpret for NNT when result is not significant.

Adjusted odds ratio is the measure of effect directly estimated by logistic regression models, irrespective of type of study. Does not always approximate well to the relative risk, particularly, if common outcome in control group (20-30%) tends to overestimate the effect, therefore, may not be particularly good measure of risk in

randomised control trial when event rates are high and quantitative rather than qualitative interpretation important.

Odds ratio may also be a measure of summary effect estimated by some statistical techniques for meta-analyses.

Magnitude of NNT and absolute risk differences are dependent on/take into account the underlying disease incidence rate. Absolute risk differences are more useful for health service planning; has clearer interpretation in estimating treatment benefits for a population. NNT: useful for quantifying benefits of intervention in RCT, considered may be easier for clinicians and patients to understand. High risk patients have smaller values of NNT and low risk patients larger NNT for similar relative risks.

### **EXAMINERS' COMMENTS**

This question was concerned with defining measures of relative and absolute differences in prospective studies. The underlined words are those (all of them in the original question) that are key to answering this question well.

On the whole this question was not well answered with many candidates not answering the question that they were asked. Common mistakes were to list and define methods for measuring disease occurrence (e.g. prevalence, incidence, SMR even in some cases perinatal mortality, infant mortality etc.) without seeming to recognise that the question was concerned with differences between two groups; to list, define and describe strengths and weakness of study types used in epidemiology (e.g. ecological study, cross-sectional study, case control study, cohort study, randomised controlled study); to attempt to write everything one possibly could about epidemiology. Few of the answered were clearly and well structured. Only a minority of candidates defined numbers needed to treat (or to harm).

A good answer was one that started with a statement that prospective studies were: (i) prospective cohort studies, in which two comparison groups would be defined on the basis of an observed exposure (e.g. current smoker versus non-smoker; user of HRT versus non-user; high fat diet versus low fat diet); (ii) randomised controlled trials, in which two comparison groups were generated on the basis of random allocation to intervention (e.g. HRT, dietary advice, screening) or control (e.g. standard treatment, placebo). By definition both study types follow participants prospectively and assess outcome some time after measurement of exposure or random allocation to groups. Good answers then briefly described what risk, rates and odds were and how these would be estimated in each group before clearly defining relative measures of difference (i.e. risk ratio, hazard ratio, odds ratio, relative risk reduction) and absolute measures of difference (risk difference, rate difference, number needed to treat). Good candidates would note that odds ratios are not a necessary nor ideal measure in a prospective studies but are sometimes used in these studies; odds differences are rarely presented.

Our main feedback to candidates with respect to this question would be to re-emphasise the absolute essential requirement of reading the question carefully before starting to answer and making sure that you answer what is being asked, and for questions that are not broken down into smaller parts having a clear structure to answer the question.

## Question 2

The following table was presented in a paper reporting a meta-analysis of six trials of treatment for chronic renal insufficiency.

Mortality ratios for death from renal causes in six prospective randomised studies of protein restriction in chronic renal insufficiency

Study Number	Mortality ratio (treatment/control)	95% Confidence interval
1	0.46	0.19 to 1.13
2	0.29	0.04 to 2.11
3	0.61	0.34 to 1.09
4	0.37	0.12 to 1.17
5	0.28	0.08 to 0.95
6	1.09	0.40 to 3.02
<b>Pooled estimate</b> (fixed effect assumption)	<b>0.54</b>	<b>0.37 to 0.79</b>

Chi square = 4.05, df = 5, p = 0.54 for test of heterogeneity between mortality ratios.

Answer each of the following (20% of marks for each part of the question):

- What is a meta-analysis?
- What is meant by the "fixed effect assumption"?
- The authors stated "The hypothesis that the effect of treatment in the different trials is heterogeneous is rejected." How can this conclusion be justified from the above table?
- What consequences would heterogeneity of the mortality ratios have had for the interpretation of the meta-analysis?
- On the basis of the figures presented in the table what overall conclusion might be drawn and what reservations would you attach to it?

### KEY POINTS

- The definition should make clear that meta-analysis involves the combination and re-analysis of data from a number of studies of similar aim and design; the data are not merely aggregated, the analysis recognises the structure (grouping) inherent from what are now being considered replicates within a single large study. The purpose is to seek more conclusive findings than provided by the studies individually. Some writers distinguish between formal and informal meta-analysis. That described above is formal. Informal meta-analysis entails arriving at a judgement (according to stated criteria) of the overall import of evidence from diverse sources without combining the original data from further statistical analysis.
- The "fixed effect" pooled estimate assumes that all the trials/studies in the meta-analysis are measuring a single underlying parameter (here, the mortality ratio), and that any differences between the results of individual studies have arisen by the play of chance, rather than by variations in the

underlying “true” effect. This contrasts with the “random effects” method which assumes that there is a distribution of “true effects”, due to differences in factors such as location, study populations, survey methods or underlying disease prevalence. The confidence interval around a random effects pooled estimate is wider than around a fixed effect estimate.

- c) Candidates should show that they know what is meant by heterogeneity in this context. The authors performed a significance test for heterogeneity which led to a chi squared statistic. The interpretation of this statistic is that the probability of making a type 1 error is 54%. Thus, if, the 5% level of statistical significance were being used, there would be no reason to reject the null hypothesis. Often a more conservative significance level of 10% or 20% is used to assess heterogeneity, but even on these criteria the heterogeneity is statistically “non-significant”.
- d) Interpretation would be difficult if there were substantial heterogeneity. Heterogeneity would suggest that the studies (perhaps by virtue of slightly different objectives, study populations, treatment regimes, etc.) really might not be considered as replicates within a single larger study (meta-analysis) and that estimation of a combined estimate could be meaningless. It is important to investigate the possible reasons for heterogeneity in the studies. Methods are available to examine relationships between study factors and treatment effect. Tests for heterogeneity are conservative and have low power to detect heterogeneity between studies i.e. The probability of making a type II error is high. It might be noted that some heterogeneity is almost inevitable in a meta-analysis. It happens in this instance that the number of subjects/trials involved is such that the significance test did not have the power to detect it. The issue is not so much whether there is statistical evidence of heterogeneity but whether such heterogeneity as there is matters; this rests on judgment.
- e) The meta-analysis gives a combined estimate of the mortality ratio as 0.54. the 95% confidence interval is 0.37 to 0.79. Thus, there appears to be evidence that protein restriction produces a substantial reduction in death from chronic renal insufficiency.

This conclusion must be tempered by our lack of knowledge about several issues. Without reading the whole paper we cannot know how the authors selected the studies for meta-analysis. It is possible that these studies are a biased selection from among a range of eligible studies. Also, we need to know how similar the studies were (design, case and control selection, the exact treatment regimen, the duration of follow-up, measures of quality of life etc.). From the table alone no conclusion should be drawn that would lead to a change of clinical practice.

## **EXAMINERS ' COMMENTS**

On the whole this question was poorly answered, which is disappointing given the importance of meta-analyses in health technology assessment and public health and epidemiology research.

- (a) Many candidates defined a systematic review in answer to this question and failed to understand that a meta-analysis is not a systematic review but is “A statistical analysis which combines the results of several independent studies considered by the analyst to be ‘combinable’” (Huque 1988). This answer also

required participants to note that this statistical analysis required results from original studies to be in the same form or a form that could be combined with other studies following appropriate statistical manipulation. Excellent answers noted that the meta-analysis would in more cases result in estimates that were more precise than those in any of the individual studies and that meta-analysis methods usually gave more weight to larger studies (or those with more outcome cases / greater precision in their original results). It was not necessary to mention or describe systematic reviews in this answer. However, we feel it is worth noting some factual errors that were made by several candidates in this question: (i) A systematic review does not have to have a meta-analysis; likewise meta-analysis can be performed in studies where there has not been a systematic review (e.g. increasingly used to combine results from several independent primary studies from large collaborative groups in genetic association studies); (ii) Neither meta-analyses nor systematic reviews are used only for randomised controlled trials (they can both be used with all quantitative studies and several methods now exist for systematic reviews and 'pooling' of qualitative studies); finally contrary to what many candidates wrote (iii) a meta-analysis does not have to have a forest plot to be a valid meta-analysis. As with all analyses the investigators (sometimes with input from reviewers and editors) can choose to display results as figures, in tables or a combination of these. Public health practitioners must be able to understand results (such as the ones in this question) however they are presented.

- (b) This was particularly poorly answered. The definition of a fixed effect assumption is provided in the key answers. A substantial minority of candidates simply did not answer this question. Many of those who did answer the question demonstrated that they did not understand what the fixed effect assumption is. For example, several said it was a fixed outcome e.g. 'death'. Some of the better answers mostly stated that this was the approach used when there was no heterogeneity between studies. These answers could not obtain full marks since they do not clearly answer the question i.e. they do not actually describe the underlying assumption of the fixed effect model.
- (c) This was very variably answered, with about half of the candidates giving a clear and correct answer (though only a tiny minority of these seemed aware that for heterogeneity tests, because of their generally low power, a p-value of 0.1 or 0.2 is a more appropriate threshold to suggest departure from the null than one of 0.05). But there were a substantial number of candidates who did not understand heterogeneity and were very confused in trying to interpret a main effect (pooled estimate) with a 95%CI that excluded the null value together with a p-value that they assumed related to this pooled estimate that provided evidence in favour of the null effect. This question could only be answered correctly if the candidate understood that the test for heterogeneity was testing whether the results for each individual study were statistically consistent with each other.
- (d) Many candidates simply answered this question by stating 'the meta-analysis would be invalid (or in some answers 'unreliable'), without giving any justification for this statement. A good answer needed to relate the heterogeneity (if it existed) to the fixed effect analysis which assumes that each study is measuring the same effect (ie. no heterogeneity). Few candidates discussed the general low power of heterogeneity tests and few

discussed additional analyses that one might undertake in situations where there was heterogeneity. Several candidates did write statements such as 'A random effects analysis should be done', but again they did not provide any justification. Adequate answers needed to define what the random effects assumption is and therefore how (if at all) this would help.

- (e) This question ought to have been well answered and it was disappointing to have so many candidates answer it in ways that suggested they really didn't understand what a meta-analysis did. The majority of candidates described the individual studies (1-6) and stated that their main reservation was that all but one of these studies were 'not statistically significant' (or words to that effect) and that one of them had a result in the different direction to the other studies. One of the reasons that it is absolutely essential not to over-emphasise  $p \leq 0.05$  (an arbitrary threshold), in this way, is that if one fails to look at point estimates and 95%CI and give these appropriate emphasis public health and clinical policy would fail. The correct interpretation of studies 1-5 is that in each study treatment has a strong protective effect, but the 95%CI are all very wide and demonstrate high levels of uncertainty (because of the small numbers). Therefore the appropriate thing is, not to dismiss these findings, but to consider whether it would be reasonable (in terms of clinical and statistical heterogeneity) to combine them, which is what the researchers actually did. For study 6 the point estimate suggests increased risk in this study sample for the intervention, but we know from the test for heterogeneity that this study result is actually statistically consistent with the other study results, all be it that since they all have wide confidence intervals there is very limited power to pick up differences. In short the main conclusion should have been based on the pooled estimate but many students failed to emphasise this. It was also disappointing that many candidates did not note as their key reservations the fact that all we have here are results and before accepting a conclusion one would want to know more about the quality of the studies used in the meta-analysis and whether all relevant studies had been included. An important minority of candidates misinterpreted 'mortality ratio' and assumed these were SMRs (this is despite the fact that the column heading includes 'treatment/control'); some candidates stated that the null value for the mortality ratio was zero (it is 1).

### Question 3

How would you evaluate the effectiveness of a local health promotion campaign to promote breast-feeding in the population?

#### KEY POINTS

- Definition of effectiveness and a broad based answer about evaluation of a health promotion campaign including structure, process and outcome.
- Evaluation framework (e.g. Donabedian or other relevant) - then clearly used relating to breastfeeding.
- Brief mention of the possible types of evaluation (ie quantitative, qualitative, cost effectiveness).
- Use of comparison areas, baseline measures, comparison trends/practices.
- Awareness of possible confounders affecting how much can be attributed to local campaign versus some other intervention/campaign.
- Evaluation against original aims and objectives of the campaign.
- Demonstrating particular knowledge of breastfeeding measures, information sources and professionals involved, mothers views etc.
- Subgroup analysis – awareness of need to look at specific groups such as ethnic minorities, young mums, deprived areas within locality.
- Awareness of other potential benefits of the campaign – professional practice, improved facilities, workplace etc., disincentivising formula milk etc. even if main outcome of the evaluation is inconclusive against “harder” measures.
- Practical aspects – who involved, reporting, using lessons learned etc.
- Potential conflict of interest if campaign lead carries out evaluation and evaluation needs different perspectives etc.

#### EXAMINERS' COMMENTS

Some candidates included irrelevant information in the answer such as detailed epidemiology of breastfeeding, detailed evidence on health benefits of breastfeeding or the detailed requirements of surveys. None was relevant to how an evaluation of a local programme may be carried out.

Some candidates used generic terms and methods of evaluation without relating to the question. Some wrote about how they would set up the campaign from scratch – again this was not the question.

Good answers demonstrated the practical aspects of carrying out an evaluation and an understanding of the limitations of both quantitative and qualitative data and information on breastfeeding.

#### Question 4

An increasing proportion of the population now travels overseas for recreation, business and family reasons.

- a) What are the health protection implications of this travel? (60% of marks)
- b) How are these implications best addressed? (40% of marks)

#### KEY POINTS

##### a) Health protection implications:

- Reason for increased travel and travel may have health implication (sex industry, health tourism, escaping regime)
- Comment on migrant health (and first generation may not be covered or immune)
- Comment on global variation in prevalence of infectious/other diseases
- Risk to host population from travellers (e.g. disease, economy) but also positive aspects

Mention of some if not all of the following:

- Accidents
- Risk taking behaviour (STIs, drugs, alcohol)
- UV light exposure
- Crime

Mention of

- Communicable disease risks with specific examples (foodborne, malaria, hepatitis and others)

Good answers included mention of:

- Variation in quality and access to healthcare
- Variation in quality of public health – surveillance/ notification/ contact tracing etc
- Health care demand
- Other difficulties – language, out of pocket payments
- Wider implications –e.g. spread – pandemic/emerging diseases, environmental impact of travel
- Specific hazards of unusual methods/ locations of travel – ebola, rabies, snake bites
- Other health risks– DVT, mental health,
- Costs - health insurance/difficulties and costs of repatriation

##### b) Addressing

**Individual measures**– immunisation, hygiene, cover up, condoms, malaria prophylaxis

**Health service** – professional awareness, contact tracing methods, specialist infection control units etc, access to primary care

**National** – port health controls, alerts to travellers (foreign office/health department etc)

**Surveillance and international collaboration** - (WHO International regulations etc.)

**Specialist resources**

Advice to travellers (Travel agents etc)

Other e.g. Brief mention on environmental measures/ climate change

**EXAMINERS' COMMENTS**

Many candidates gave mixed answers rather than answering each part of the answer in an organised manner. Some candidates limited their answer to communicable disease aspects whereas wider health protection considerations were required for a good answer. Some candidates did not give a balanced approach to control by considering exclusively individual control aspects or wider aspects (such as trade restrictions). The answer required included a range of measures from individual prevention measures (hygiene, immunisation, safe sex), professional awareness, facilities, to wider aspects such as surveillance, port health and international collaboration.

Many answers were however well considered and demonstrated a good approach to the question and contained many additional considerations to those in the key points, and these were reflected in the marks.

## Question 5

Define the key features, uses and limitations of cancer registration systems in public health research and practice.

### KEY POINTS

(Points marked \* may indicate an excellent answer, if they are in addition to the basic answers)

#### Key features

- Identifies individuals with particular cancer e.g. according to ICD10 coding
- Longitudinal data – incidence
- Based on geographical population
- Sources of data – clinical, laboratory, oncology or diagnostic radiology, deaths
- Agreed case definition
- Defined minimum data set
- \* Needs a system to maintain quality
- \* Needs a robust data collection, analysis and storage system
- \* May be based on voluntary or statutory reporting

#### Uses of cancer registration system

- National system of identifying incidence of cancer in a country – within population sub-groups and between cancer sites
- Health needs assessment, and service evaluation and planning
- Trends in incidence and survival
- Survival studies – within sub-groups and between cancer sites
- Comparison between regions and internationally
- Audit and evaluation of the effectiveness of screening programmes
- \* Research dataset – sampling frame for case control studies, cluster analysis

#### Limitations of cancer registration system

- Developing and using standard case definitions can be difficult
- Variable quantity and quality of data – missing data, duplicates
- Case ascertainment – if voluntary there may be under-reporting of cases
- Linkage to other data systems can be difficult – although with IT developments it is becoming much easier e.g. linking hospital admissions or activity data with cancer registries data
- \* Time bias data – often information is given some time after the event
- \* In some situations it can be difficult to estimate the denominator to give true rates – e.g. in communities with a sizeable 'floating' or 'hidden' population (e.g. migrant workers)
- \* Expensive and/or resource intensive to set-up, run and maintain a system

## **EXAMINERS' COMMENTS**

Few candidates mentioned the term 'denominator', much less defined it by reference to census data. Hence few candidates wrote about incidence rate. Many did not include 'time trends' in their answers. Both of these were major omissions.

Many candidates described the collection of 'risk factors' and their usefulness in epidemiological analysis. In general, cancer registries (CR) do not record information on risk factors.

Many candidates put down in their answers the uses of CR include generating data on incidence and prevalence. In reality, it is extremely difficult to document prevalence data on cancer.

It is also very difficult to evaluate treatment options, although the UK CR does collect information on treatments, as this is not possible to control for confounders. Treatment is better evaluated by clinical trials than by interpreting CR data.

Many candidates exaggerated the usefulness of the CR data as a source that identifies cancer clusters. This is usually not the case (although in the key points, this is mentioned, but in the context of a 'research dataset'). Most cancer clusters are identified from empirical evidence. Data in the CR may take one or two years to complete and edit, and this time lag between the publication of the data and the onset data is usually too much for any publicity associated with the usual 'clusters' that alert the public and health authorities.

All in all, we found that many candidates have little actual understanding of how a cancer registry is run. And the lack of understanding of terms like incidence and prevalence is especially disappointing for candidates at membership exam level.

## Question 6

Systematic information on ethnicity is collected by many health services.

- a) What is the rationale behind recording information on ethnicity? (50% of marks)
- b) What epidemiological concerns should you be aware of when considering data on ethnicity? (50% of marks)

## KEY POINTS

(Points marked \* may indicate an excellent answer, if they are in addition to the basic answers)

### Rationale

Candidates should give evidence of an understanding of the issues around ethnicity and health in the health service.

- In order to fully understand the extent of the problem, and which groups are affected.
- Ethnic groups may have different health needs (e.g. high birth rate; care of the elderly).
- Ethnic groups may have different health behaviours (e.g. diet and exercise).
- Ethnic minority groups are widely believed to have poorer access to health care.
- Ethnic minority groups generally have worse health outcomes than the indigenous population.
- This may be due to specific factors for particular diseases (e.g. South Asians and heart disease, Afro-Caribbean and hypertension).
- But may be partly due to the experience of discrimination leading to poorer educational, employment and life opportunities (e.g. reduced mental well-being).
- \* Evaluation of interventions to reduce inequalities in health

### Epidemiological concerns

Candidates should be able to show that they are aware of the problems of recording and using data on ethnicity.

- Problems of definition. Ethnicity is not the same as racial origin, skin colour, or country of birth. It is now common to ask individuals to identify their own ethnic group from a standard list (e.g. that defined by the national census in many countries).
- There is a tendency to lump heterogeneous categories together (e.g. Asian, Black) which may be unhelpful.
- Focusing on ethnic group may miss other deprived groups that cross ethnic boundaries in the population (e.g. children living in poverty).
- Inadequate control of confounding factors (e.g. social class, social disadvantage, poorer education and lower income).
- Selective analysis. Ethnicity data tends to be used most when it can help to show that ethnicity gives rise to health 'problems'. However systematic unbiased analysis may often show better health outcomes for certain categories.
- Inappropriate use of analyses. Epidemiological analyses often focus on differences, rather than trying to understand the underlying processes.

## **EXAMINERS' COMMENTS**

This question was generally well answered, however some candidates did not focus on the term 'rationale' and produced answers (usually outlining the uses of data on ethnicity) that were not relevant to the question asked.

On the 'epidemiological concern' section, most candidates addressed the problems of data quality, definitions and coverage. Only few candidates (probably those with a strong research background) gave good answers on confounding, analytic epidemiology, and causal relation.

## Question 7

The head of your health organisation asks your opinion on the findings of a published economic evaluation. Outline the main steps to critically assess the evaluation and to judge whether it can be applied in your local situation.

### KEY POINTS

A well answered question will use a framework such as the one below to show a structured approach to the question.

The main steps will assess

- whether the methodology was appropriate,
- the validity of the results,
- whether they will apply to your local setting.

1. Was there a well-defined question, with a comparison of alternatives?  
Economic evaluations are efficiency evaluations and should address the **relative** value of a course of action compared with doing other things with the same resources.
2. Was a comprehensive description of other options given?  
A clear statement of the primary objective of each alternative treatment, programme or service is needed, for a number of reasons. The reader should be able to judge the applicability of the programme to local settings; and to replicate the programme, if desired. It is also needed to assess whether the most appropriate methodology has been used (cost benefit, cost effectiveness and cost utility), and whether all important costs and consequences have been included.
3. Was there evidence of the effectiveness of the treatment or programme?  
This should have been established before an economic evaluation was considered. There is little point in implementing an ineffective intervention efficiently! Was the evidence robust? What potential is there for significant bias?
4. Were all important costs and consequences identified?  
In order to identify costs and consequences, all relevant view points have to be identified and clearly stated, e.g. patient, carers, provider, community, social etc. The range of view points and costs and consequences has to be wide enough to address the research question at hand, but it may not be possible to list all, or to estimate the wider ripple effect that a health programme might introduce.

Categories of cost include:

- Health care resources, both variable and fixed
- Patient and family resources
- Impacts on wider economy (for health programmes, this is most likely to be in the social care, voluntary or public sector).

Categories of consequence include:

- Changes in health state (health state preference scores, willingness to pay)

- Other value – e.g. reduction in anxiety
- Resources saved (e.g. savings made by prevention programmes averting more expensive treatment of subsequent conditions)

5. Once identified, were costs and consequences measured accurately and in appropriate physical or natural units?

Examples of units of measurement might include hours of nursing time, number of examinations performed, lost work days, number of prescriptions etc. All the components identified in step 4 should be measured. Some costs will be shared by one or more programmes. The basis of allocating costs between programmes should be clear.

6. Were costs and consequences valued credibly?

Costs are usually valued in local currency and based on prevailing prices. Future costs are valued to a specified base year to remove effects of inflation. Be aware that average costs may be very different to the actual costs. For example, the cost of an average hospital day or hospital bed may be very different to the actual costs for the specific condition being evaluated. It is more accurate to distinguish 'hotel' costs of the hospital day or bed, from treatment costs. 'Free' goods, such as volunteer labour should be costed to their full market value.

The techniques for valuing 'consequences' varies according to the type of economic evaluation chosen. Valuing health states involves adjusting the duration of time spent in a particular health state with a factor that increases or decreases the value of the time spent in that state. The adjustment factor varies according to whose preference is used (patient, provider, tax-payer etc.). Techniques vary between evaluation method; in cost utility analysis, adjustment factors are constructed using rating scales, 'standard gamble' or 'time trade off' techniques; in cost benefit analysis, consequences are set out in monetary terms and constructed through techniques such as 'willingness to pay' and 'human capital approach'.

Cost effectiveness analysis does not include values and these may need to be estimated separately.

7. Were costs and consequences adjusted for differential timing?

Future costs and consequences should be discounted to reflect 'time preference'. The choice of discount rate should be appropriate and justified.

8. Was an incremental analysis of outcomes and costs of alternatives performed?

This will be necessary if the economic evaluation compares alternative strategies to detect or treat condition(s). The evaluation should show the additional cost of strategy 1 over strategy 2, compared to the additional benefit/yield of strategy 1 over strategy 2. If only average costs for the two strategies are compared, the higher incremental costs (of the extra cases detected/treated) is masked, being spread throughout the total number of cases detected/treated.

9. Was a sensitivity analysis performed?

Critical assumptions/uncertainties should be identified (e.g. discount rates, costs, compliance rates), and the analysis reworked using different assumptions to test whether this substantially alters the results and

conclusions. If cost/consequence data are stochastic, differences can be presented with confidence intervals and estimates of statistical precision.

10. Did the presentation and discussion of study results include all issues of concern to users?  
Various technical and value judgements should be explicit, and preferably matched to a range of results, so that the study user can make their own judgement about the values employed.  
Methodological limitations should be clearly presented, such as the impact of equity on the determination of values (willingness to pay may be distorted by ability to pay).
11. Were the conclusions justified by the evidence presented?
12. Can the results be applied to the local population?  
Are the populations studied similar to the local population? Age structures of populations vary from country to country and may affect incidence of disease, lifestyle factors etc.  
How does your setting compare to that in the study? Variations in practice and availability of alternative treatments can affect the relative cost effectiveness of treatments. Differences in relative prices, e.g. between diagnostics and treatment between countries, may alter the direction of the conclusion.  
Will the service or treatment be acceptable to the local population?  
Will there be political or other issues that may impact on implementation?

Data may be available in the evaluation to allow local modelling to adapt the results from one setting to another.

### **EXAMINERS' COMMENTS**

A few candidates answered this question well and were clearly knowledgeable about the important components of an economic evaluation.

Credit was given to clearly structured answers, even if these did not follow the format adopted in the key points.

Unfortunately, many candidates did not answer the question, choosing instead to describe all they knew about different types of economic approach. This is a common failing that is repeated each year. Candidates need to demonstrate that they have applied their knowledge, rather than make random statements of fact, in order to gain full credit.

A number of candidates confused this question with a critical appraisal of a RCT, discussing such issues as sampling, bias and intention to treat analyses.

A small number of candidates wrote vague answers and did not provide any detail of the factors that they would consider when assessing the quality of the evaluation.

## Question 8

What is the scope for improving health in and through the workplace?

### KEY POINTS

A good answer will take account of the following factors. An excellent answer will include some discussion of underlying theoretical concepts of mediation between work stress and ill-health - the job demand-control model, and the effort-reward imbalance model; and discussion of how a workplace intervention could be initiated and evaluated.

#### 1. Impact of work on health

A. Positive:

Beneficial effects of work:

- provide material benefit and income
- reduce health inequalities
- create a meaningful role and social status
- offer improved opportunities
- can be important for recovery and rehabilitation.

Unemployment leads to poorer health and shorter life. Increased smoking at onset of unemployment; associated with increased alcohol intake, weight gain, drugs misuse, sexual risk taking, reduced well being with increased self harm, depression and anxiety.

B. Negative:

Direct risks to health through occupational illnesses or risks.

- Musculo-skeletal conditions. } greatest cause of sickness absence
- Mental ill-health and stress } and incapacity benefits.
- Occupational hazards and risks from poor work systems.
- Exacerbation of chronic illness or disability by workplace exposures.
- Physical inactivity has health consequences – including obesity, coronary heart disease and cancer.

Relationship between psychosocial work environment (lack of job control, boring, repetitive and monotonous work and imbalance between effort and reward, 'unfairness' at work) and long term conditions, such as coronary heart disease

#### 2. Impact of health on work

- Poor health can be a barrier to employment through stigma and discrimination.
- Sickness absence costs
- The cost of making reasonable adjustments to keep an employee who develops a health condition or disability (although this is typically lower than the cost of recruiting and training a new employee).
- Alcohol misuse increases absenteeism, unemployment and premature death.

#### 3. Scope to improve health through workplace initiatives

Work place initiatives give the opportunity to target people of working age but these can have wider community benefits through impact on families of lifestyle choices, mental wellbeing, and financial remuneration.

### **3.1 Occupational related risk or illness**

Competent occupational health support and advice is currently not widely available in the workplace. Need to identify incentives for businesses to provide such services, or access to them.

Key roles of occupational health services can include;

- Pre-placement / pre-employment health screening
- Advice on fitness to work / adaptations needed to accommodate disability
- Advice on rehabilitation following illness
- Support and advice services, including counselling
- Support to wider corporate responsibilities to assess risks and hazards in the workplace.
- Role in surveillance or monitoring of risks, hazards or health status.
- Support to corporate plans to prevent or limit risk, e.g. through immunisation programmes, provision of training, emergency planning etc.

Corporate practices to reduce risk (e.g. for musculoskeletal conditions):

- Risk assessment procedures e.g. assessment of manual handling component of work and targeted training
- Work planning to take account of repetitive or boring tasks
- Early reporting and response systems for symptoms,
- Flexible working arrangements
- Early rehabilitation

### **3.2. Pre-existing illness or disability**

Work can be of positive therapeutic benefit for people with both mental and physical health problems. Research recognises that the chances of return to work diminish with increasing sickness absence.

People do not have to be fully fit or recovered before being reintroduced to the work place. Employers are encouraged to make reasonable adjustments and view work as a positive component of rehabilitation. Temporary job modifications and personal services may be essential to encourage people back to work.

Corporate level interventions include

- Recruitment procedures / policies. This includes working with recruitment partners and intermediaries who can provide varying levels of support to current and prospective employees.
- Development of policies for rehabilitation, sickness absence, flexible working, diversity etc, with attention to Disability, Human Rights and Health and Safety legislation.
- Organisational culture development, including initiatives to improve awareness and reduce stigma.
- Occupational health support – innovative services form links with primary care and social care providers, to assist people staying in, or returning to work following health problems.

Individual level interventions (line manager to person) include

- Clear induction processes and assessment of job/person fit
- Sound management processes of appraisal, work planning, training, and assessment
- Early detection of problems and discussion of supports / adjustments
- Assertive engagement and return to work planning with people off sick
- Return plan with monitoring and support – increasing role of ‘intermediary’ such as specialist employment adviser.

### **3.3. Health promotion in the work place.**

#### *Tobacco smoke and smoking cessation*

Develop a smoke-free policy; provide smoking cessation.

Ensure designated smoking areas are effectively screened, so that smoke does not adversely affect other employees at work.

#### *Alcohol and other substance misuse*

Develop policy and code of conduct for alcohol and substance misuse in the workplace.

#### *Physical activity*

Encourage employees to walk or cycle to work.

Involve employees in organising a workplace activity programme.

Make the stairwells more attractive and use signage to encourage use of stairs rather than lifts.

Provide information on the benefits of physical activity.

Consider negotiating discounted membership of a local gym for employees.

#### *Healthy eating*

Provide healthy choices in canteens and vending machines; remove salt from tables

#### *Engage families/communities*

Develop family friendly policies, flexible working hours, support breast feeding.

Be sensitive to cultural needs.

The work place focuses on people of working age but can have wider community benefits if health benefits reach into families and other social networks.

#### *General health advice*

Encourage employers to support provision of advice and interventions for behavioural change from lay / peer groups e.g. health trainers.

#### *Service provision*

Flexible health service provision in partnership with employers. Role for primary care, working collaboratively with occupational health services.

## **4. Implementing and evaluating a workplace initiative**

Issues to address:

- Employee job satisfaction
- How work is organised and carried out
- Physical working conditions
- Employee consultation and involvement
- The organisation's policies, procedures and rules.

Potential benefits:

- Improves productivity and performance
- Reduces absenteeism and other costs associated with ill health
- Fewer injuries, accidents, and insurance and compensation claims
- Improved employee morale and staff retention
- Employees more receptive to and better able to cope with change
- Enhanced business reputation and corporate responsibility

### **EXAMINERS' COMMENTS**

This was a very straightforward question and most candidates wrote adequate, if somewhat superficial answers. Few candidates discussed the positive role of work as part of a 'recovery' model or in the context of rehabilitation or maintaining those with long-term conditions.

A small number of candidates performed well, discussing the topic broadly and drawing on work/stress theory.

Poorer answers were too narrowly focused, e.g. on vague health promotion messages; or were incomplete, e.g. only describing the process of strategy development.

## Question 9

What are programme budgets? (30% of marks)

How can the strengths and limitations of programme budgets be taken into account when establishing and evaluating public health programmes? (70% of marks)

### KEY POINTS

#### What are programme budgets? (3 marks)

1. Definition: Programme budgets are funded budgets dedicated to implementing programmes of various activities for usually a year but this could be longer.
2. Nature of programmes: The programmes have detailed parts and sub-parts which themselves have detailed budgetary allocations. The parts and subparts are important in their own right.
3. Nature of financial budgets: Budget comprises all spending allowable in revenue and capital monetary terms in a given time period on a programme, across sectors and budget headings.
4. Purpose: They are useful for several purposes. One is to show accountability through detailing how much money was spent for such a topic or purpose. They then allow an audit of funding expenditure that can check whether funding was spent for its intended purpose.
5. Examples: They can apply in the healthcare sector and other sectors such as government, research granting bodies, and international bodies. For example, the WHO uses programme budgets for its global health programmes.

#### How can they be used for public health purposes? (7 marks)

1. Uses: Uses include public accountability on health funding expenditure, a major focus on public health management, marginal economic analysis for investment and disinvestment purposes, clarity of current funding and future funding needs for priority setting, motivation of staff with a known budget and delegated responsibility, and financial audit of budgetary allocation and planned expenditure.
2. Public health management: health programmes and programme budgeting allow for public health management. This includes:
  - advocacy and management roles
  - knowledge and action
  - managerial capacity and infrastructure
  - networking to create partnerships across organisations & disciplines
  - broad involvement of people and skills
  - infrastructure and curricula for education
  - evidence based policy and practice
  - an outcome-based focus
  - a national agenda for health and health services research.
3. Marginal economic analysis: Expert group identifies proposals for investment and disinvestment in a health programme such as those for CHD or mental health. Economists then analyse the health impact of marginal changes in investment in these areas and offers proposed priority list of investments and disinvestments. One

can influence relative priority between programmes at the margins but not in total. One can also change priorities within a programme by balancing spend within a programme.

4. Drawbacks: 1) The major drawback is the difficulty in creating accurate budgets that cover all aspects of a health programme. Thus a programme budget for cervical screening would have to cover expenditure in primary care for nurses and GPs time, community clinics for screening and well women, GUM clinics, laboratory activity, and health promotion activity. 2) Another process is needed for deciding how to allocate budgets across health programmes if there is to be a big shake-up of funding.

5. References: Anything from David Hunter and Richard Alderslade. They are still both active in pushing this agenda globally. Anything on marginal economic analysis. Reference to current DoH Commissioning Directorate, National Programme Budget project board, and NHS Health Knowledge Team and their project on programme budgeting, or references to similar developments from other countries.

### **EXAMINERS' COMMENTS**

There was a very low pass rate for this question. Answers were often rambling and poorly structured. Some clearly demonstrated a complete lack of knowledge about the subject. There was limited use of anything other than lay terminology. Little mention was made of health economics concepts which were relevant to this question such as marginal analysis or more general everyday terms such as ring fencing and top slicing. Some answers were wrongly emphatic that programme budgets do not cover capital costs or staff time.

## Question 10

You have been asked to prepare an annual report on the health of the local population for which you are responsible (population size 500,000). Outline how you would manage the process (70% of marks) and assess the subsequent impact of the report in the year following the publication (30% of marks).

### KEY POINTS

#### A. Manage the process of production (7 marks)

Establish the nature of the task

- who is asking for the report, why are they asking and who are the intended audience
- is there any suggested or expected content
- is the task part of your job description, is it an official requirement
- do standards exist for the production of the report
- does the report have to link with previous reports

Assess resources available

- Authors – think widely – health, local government and voluntary organisations
- Editorial team to monitor process, agree house style and edit
- Type and sources of information and data analysis
- Design and layout
- Any surveys needed? Or qualitative work?

Establish timescale

- Obtaining data – writing content – getting agreement on key messages and recommendations – editing – publication

Agree content

- Comprehensive or topic based
- Consistent with previous reports

Sources of information

- Nationally and locally available data
- Self-assessed health, Census data, lifestyle and health belief data, illness incidence and prevalence, mortality, births and fertility, health service use, local health projects, disability, user opinions, health outcomes
- Establish what information will be used

Authors write contributions

- In knowledge of style template

Agree key messages and recommendations

- Progress on health targets and health programmes
- Points of concern
- Agree recommendations with agencies concerned and discuss funding issues
- Write forward

Prepublication

- Ensure consistency, check facts, copy edit

Consider method of publication and communications strategy

- Online, hardcopy – how many, distribution list, method of distribution, any need for translation
- Presentation to specific people and organisations

### **B. Monitoring impact (3 marks)**

Survey a sample of the intended audience

- Means of survey
  - Telephone or postal using semi-structured questionnaire
  - Informal anecdotal evidence
  - Include survey as part of report to be sent in after specified time
- Did the report reach the intended audience?
- Was the content regarded as clear and unambiguous?
- Was it readable?
- Did they enjoy reading it?
- Were the recommendations and key messages regarded as evidence based?
- Did they take any actions because of the report?

Review of key messages and recommendations after a specified time, usually a year

- Have proposals for change been implemented
- Were they regarded as relevant with adequate justification
- Were they regarded as appropriate and achievable
- Have the recommendations resulted in a reduction in inequalities

Bad fail – 2 or less of the key points

Borderline pass – five of the key points with exposition and with one from Section B

Good pass – more than six of the key points with fairly full exposition.

### **EXAMINERS' COMMENTS**

The examiners expected to see the basic points covered in this question which was a topic to which every entrant in this examination would have been exposed. However, comprehensive answers were elusive. Some candidates focused too much on management theory without conveying a sense of the practical and necessary stages in the production of a public health report. At times the disgorgement of this theory had only a loose connection with the task in hand. This approach did not instil confidence that the candidate would be a competent member of a team involved in the production of such a report. Other candidates discussed the content of the report in a degree of detail that was not required and omitted discussion of the other aspects of public health report production. With reference to monitoring impact very few answers talked about the expected impact of the report on inequalities. This question was relatively straightforward but very few candidates answered to a good standard.

## Paper IIA

You are a public health advisor to a group responsible for planning maternity services for a population of 250,000 people. A letter has been sent to you from the director of a national diabetes charity drawing your attention to a recently published paper in the BMJ and asking why your local hospitals do not screen pregnant women for diabetes and have clear guidance for strict control of high blood sugars in pregnant women with diabetes. The paper to which the director refers is:

Macintosh MCM et al. Perinatal mortality and congenital anomalies in babies of women with type 1 or type 2 diabetes in England, Wales and Northern Ireland: population based study. BMJ 2006;333:177-180.

1. Write a critical appraisal of this paper. (40% of marks)
2. In the discussion the authors compare their results to other studies that have found higher perinatal mortality rates and congenital anomaly rates in women with pre-existing type 2 diabetes compared to those with type 1 diabetes, whereas in this current study there was no such difference. The authors state *"However, even this study was underpowered to detect a significant increase of less than 25% for perinatal mortality and of less than 80% for congenital anomaly, if such increases were present."* Explain what the authors mean by this statement. (10% of marks)
3. The authors found that only a minority of women with diabetes had good periconceptional glycaemic control. What type of study would you undertake to determine whether better periconceptional glycaemic control would improve perinatal mortality in women with diabetes? Justify your answer and briefly describe how you would undertake this study. (30% of marks)
4. Write a letter replying to the director of the national diabetes charity. This should include comments on the BMJ paper and its relevance, but also cover local policy and issues. (20% of marks)

### KEY POINTS

*Examiners are asked to note that although the paper is based in a UK setting, candidates may be answering from the context of a different health care system.*

1. Write a critical appraisal of this paper. (40% of marks)
  - Important question – rates of type 1 & type 2 diabetes increasing – particularly the latter (in relation to obesity) in women of reproductive age. Pre-existing diabetes in pregnancy is a known risk factor for perinatal mortality and morbidity and over decades there have been international (e.g. St Vincents Declaration) and national initiatives to improve antenatal and obstetric care for women with diabetes in order to reduce adverse outcomes in mother and child. The authors note that this study was conducted as the NSF for diabetes (which includes guidance for management of diabetes in pregnancy) was released and therefore these results will act as a baseline for monitoring the impact of the NSF.
  - This is a population based prospective cohort study with the objective of providing perinatal mortality and congenital anomaly rates for babies born to

women with pre-existing type 1 or type 2 diabetes in England, Wales and Northern Ireland.

- Data from CEMACH used with all health professionals in UK required to take part and therefore results are likely to be generalisable to the UK, at least. Since perinatal mortality in relation to pre-gestational diabetes may be related to health service provision the results are not necessarily generalisable to other countries.
- Similar data collection system (questionnaire) used across the whole of UK; though no discussion of whether there was any evidence that reporting of cases varied by participating unit or what % of questionnaire responses were missing and whether this varied by unit.
- We know from the results that 68% of women with diabetes had a record of glycaemic control but we don't know whether the % with this record varied by unit / area of the country. If it does this may affect generalisability of results in relation to glycaemic control.
- Pre-gestational diabetes clearly defined as having been diagnosed at least 1 year before the woman's estimated delivery date
- Standard definition of perinatal mortality and congenital anomalies collected in standardised way in all centres and confirmed with post-mortem findings, genetic results or correspondence
- For perinatal mortality results were compared with those from CEMACH for all women so comparing with standard UK population
- Main results for congenital anomalies given as standardised prevalence ratio i.e. for each outcome the observed number of women with type 1 diabetes is divided by the expected number for women in a similar age group (standardised for age) from EUROCAT so here the UK women with diabetes are being compared with pregnant women in general from across Europe. – The authors definitely don't say that they only used UK values from EUROCAT. If there were large variations between countries in Europe then using this group as the standard for comparison may exaggerate or underestimate the 'effect' of diabetes in UK pregnant women.
- The authors note the association of deprivation with type 2 (compared to type 1) diabetes and also higher rates of type 2 diabetes in women from minority ethnic groups. Perinatal mortality is related to socioeconomic position and congenital anomalies are more common in some minority ethnic groups, but these are not controlled for in the final analyses that compare rates by type of diabetes or rates in women with diabetes to the rest of the population. Other factors that could potentially confound / explain these associations are also not explored. E.g. Obesity is a strong risk factor for type 2 diabetes and an independent risk factor of perinatal mortality and so may explain some of the increased perinatal mortality in women with type 2 diabetes. Determining whether control for glycaemic control reduced any of the associations between diabetes and adverse outcomes would also help to understand whether this was important in explaining the increased rates. It is impossible to do some of these adjustments in this study because information on e.g. obesity is not collected in women without diabetes. However, the authors could have considered examining whether rates of perinatal mortality and congenital anomalies varied by level of glycaemic control (perhaps didn't have sufficient power).
- Ultimately this is a descriptive study highlighting increased risk of perinatal mortality and congenital anomalies in women with pre-existing diabetes but it doesn't take things forward in terms of understanding the key reasons for this.

### **Excellent candidates**

Might point out that the relative differences between women with diabetes compared to the standard populations of UK or European pregnant women are highlighted with no mention of absolute differences.

2. In the discussion the authors compare their results to other studies that have found higher perinatal mortality rates and congenital anomaly rates in women with pre-existing type 2 diabetes compared to those with type 1 diabetes, whereas in this current study there was no such difference. The authors state *“However, even this study was underpowered to detect a significant increase of less than 25% for perinatal mortality and of less than 80% for congenital anomaly, if such increases were present.”* Explain what the authors mean by this statement. (10% of marks)

The authors are acknowledging that even with a study that includes all pregnant women with diabetes in the UK the small numbers who experience perinatal mortality or congenital anomalies are too small to be able to detect what might be considered important differences between women with type 1 and type 2 diabetes. They do not state what they consider to be adequate (e.g. 80%, 85%, 90%) or an appropriate level of statistical significance (0.05, 0.01) but they indicate that they only have adequate power at a given statistical significance level to be able to detect a difference of 25% or more between type 1 and type 2 diabetes for perinatal mortality and 80% or more between type 1 and type 2 diabetes for congenital anomalies if such differences exist.

### **Excellent candidates**

Might comment on whether they think being able to detect a difference of less than 25% for perinatal mortality is important or not (ditto for 80%) i.e. whether small differences would influence policy recommendations /clinical practice

3. The authors found that only a minority of women with diabetes had good periconceptional glycaemic control. What type of study would you undertake to determine whether better periconceptional glycaemic control would improve perinatal mortality in women with diabetes? Justify your answer and briefly describe how you would undertake this study. (30% of marks)

- Randomised controlled trial would provide the best evidence of effectiveness BUT might be considered unethical to consider randomising women of reproductive age with diabetes to different levels of glycaemic control and might be impossible because would need such massive numbers (see in this study there are just 75 cases of perinatal death (160 total still birth, perinatal and neonatal death) for all women with diabetes in UK. An RCT in this particular group would need to include many countries to provide a timely answer.
- A retrospective cohort study built into CEMACH continued surveillance of women with diabetes but with better collection of preconceptual glycaemic control –e.g. through retrospective record linkage to primary care notes would be possible. Might need to collect data over several years to have sufficient numbers to determine precise association of glycaemic control with perinatal mortality.
- If candidate says they would do an RCT they must discuss issues of sample size, with clear need for very large multi-centre trial and how that would be co-ordinated and standardised. They should describe ethical issues and

related to that just what the intervention would be. Also discuss how randomisation would be concealed, outcome assessment blinded and analyses conducted that made sure all those entered into the trial could be accounted for and what intention to treat analyses would involve.

- If candidate says they would do a cohort study or case control. Need to discuss numbers required and how they would obtain these. How they would ensure standardised collection of data / avoid measurement error in exposure and outcome. What confounding factors they would consider and how these would be assessed. What analyses would be undertaken.

#### 4. Letter to the director of the national diabetes charity (20% of marks)

##### *Introduction*

- Acknowledge the directors concern about the negative consequences for mothers and babies of diabetes during pregnancy and child birth and the importance of close glycaemic monitoring and control.

##### *BMJ paper and relevance*

- Acknowledge importance of the study- highlighting increased risk of perinatal mortality and congenital anomalies in women with pre-existing diabetes; particularly given increasing rates of type 1 & type 2 diabetes (in relation to obesity) in women of reproductive age.
- Indicate study results may be generalisable to the whole of the UK, given use of national data collection/CEMACH data; but possibly will be different in other countries and may also vary regionally within the UK. Need to take account of local population characteristics and circumstances and quality of services.

##### *Local policies re local monitoring and clinical management*

- Progress in local implementation of clinical guidelines or service frameworks for diabetes and standard for diabetes management during pregnancy- that are evidence based (including systematic review evidence; good practice experience)

Could highlight:

- Collection of data on numbers (and prevalence) of women with pre-existing diabetes in antenatal care locally, and comment on trends including the role of social deprivation and ethnicity.
- Definition of standard practice for care of women with diabetes in the area
- Whether there any specific features of the population in the area (e.g. large minority ethnic group) that would suggest you ought to be screening for diabetes in pregnant women
- System for quality assurance: -reference to any recent audits of monitoring and treatment for women with diabetes who are of reproductive age in primary or secondary care in your area

**Excellent candidates will be able to convey key points above in a clear and accessible style/ language**

## **EXAMINERS' COMMENTS**

Candidates' answers on interpretation of power (the study being 'underpowered') were overall weak. Candidates were able to provide a basic definition of 'power' but found it more difficult to interpret its application to the study.

A minority of candidates did not have sufficient command of basic definition of different study designs and therefore were unable to provide satisfactory answers to what type of study would be appropriate to determine whether better periconceptional glycaemic control would involve perinatal mortality in women with diabetes. Nearly all candidates failed to address the scale of studies required, given the small numbers of perinatal deaths involved.

Candidates need to practice preparing letters; answers to question 4 on replying to the director of a national diabetes charity were generally disappointing. Few candidates were able to provide comprehensive answers: interpretation of relevant points from the BMJ study, application to the local population, and reference to local policies, audits and reviews. Candidates should be aware that they must not sign their names to the letter, i.e. identify themselves.

## Paper IIB

You are a member of a cardiovascular prevention group which gives advice on planning health services for an area with a population of 300,000 people.

Atrial fibrillation (AF), an irregular disturbance of heart rhythm, is a recognised risk factor for cerebrovascular disease, and the associated risk of stroke can be reduced by anticoagulant treatment.

The cardiologist in your advisory group draws attention to a recent review article stating that *“atrial fibrillation fulfils many of the criteria for a screening programme”*. She suggests that all persons over 65 years of age should have an electrocardiogram (ECG) performed to detect or exclude AF.

Q1. List the criteria for judging whether it is appropriate to screen for atrial fibrillation. Outline the information that you would seek to obtain, within one month, to inform the decision whether to introduce AF screening in your area. [20% of marks]

The ability of primary care health workers to diagnose atrial fibrillation from ECGs was assessed in a study of patients invited for screening. (Br Med J 2007;335:380-382). The “gold standard” diagnosis was determined by two consultant cardiologists. General practitioners correctly detected 79 out of 99 AF cases and misinterpreted 114 out of 1355 cases of normal (sinus) rhythm as AF. Practice nurses detected 76 of the 99 AF cases but misinterpreted 203 of the 1355 cases of normal rhythm as AF.

Q2. Calculate and compare the positive predictive values of a diagnosis of AF, as made by a general practitioner and a practice nurse. Comment on possible implications for a screening programme based on ECGs in primary care. [20% of marks]

Q3. Suggest a suitable statistical test for evaluating the level of concordance between the two cardiologists. [5% marks]

[Continues overleaf]

Fifty general practices were included in a cluster randomised trial of screening for AF, comparing opportunistic and systematic approaches to screening within 25 intervention practices, and with routine clinical practice in 25 practices assigned to no intervention (Br Med J 2007;335:383-386). The table below shows the numbers of new cases of AF detected during the 12-month trial period, as determined by inspection of the GP notes.

Group	Patients recruited	AF known at start of trial period	AF diagnosed during the trial period	Missing case notes
Control practices (normal clinical care)	4936	389	47	34
Intervention practices				
Total	9866	679	149	50
Opportunistic screening	4933	340	75	18
Systematic screening	4933	339	74	32

Q4. Compare the rates of detection of atrial fibrillation in control and intervention practices, and among patients randomised to opportunistic or systematic screening within the intervention practices. Comment on possible explanations for these results.

[25% of marks]

Q5. Write a 500-word report for consideration by your chief medical officer / director of public health, discussing the relative merits of opportunistic screening and systematic screening for AF to prevent cerebrovascular disease among the elderly population in your area. [30% of marks]

### KEY POINTS

Q1. *List the criteria for judging whether it is appropriate to screen for atrial fibrillation. Outline the information that you would seek to obtain, within one month, to inform the decision whether to introduce AF screening in your area.* [20% marks]

Candidates will be expected to cite the Wilson-Junger criteria or derivations from them:

- Characteristics of the disease (stroke is the preventable disease outcome, AF is the pre-symptomatic risk factor)
  - a) Is the disease an important public health problem? – CVD is a major cause of death and disability in most populations. Seek supporting evidence from national or local mortality statistics, hospital admission data, stroke registries if available, national or local disability surveys.
  - b) Is there an identifiable latent or early symptomatic stage of disease? – AF may persist unrecognised for years before embolic complications occur. It can be detected with high sensitivity during this time by ECG, and is not uncommon among the elderly in most populations (c.5% of over 65's in developed countries) . Information required on the prevalence of unrecognised AF from surveys in the epidemiological

literature, projected to the local population with due allowance for age structure, and relative prevalences of rheumatic and ischaemic heart disease.

c) Is the natural history of the disease known? – Partially. Certainly there is sufficient evidence of embolic complications to warrant preventive anticoagulant therapy, as stated in the question. Seek confirmatory information from medical literature (particularly review articles on AF) on relative and absolute risks of stroke associated with AF.

- Characteristics of the test:

d) Is the technique to be used for screening effective? – Validity (sensitivity and specificity) and reproducibility. Seek information from published literature and screening programmes elsewhere.

e) Are the tests acceptable to the population? – ECG is safe and non-invasive. Unlikely to need further evidence on acceptability prior to the committee meeting.

f) Is the cost of screening acceptable? – What is important here is opportunity cost. Information required to assess this locally will include estimates of time (GP or nurse) and equipment costs (ECG machines) for the test, plus time for diagnostic follow-up and costs of long-term treatment (including regular monitoring of anticoagulant levels).

- Treatment

g) Is there an effective treatment for early disease? – Yes, as stated in the question.

h) Is effective treatment available and does management of cases in the early stages have a favourable impact on prognosis? Seek confirmatory information from published RCTs and meta-analyses, particularly to estimate numbers needed to treat and the balance sheet of risks and benefits from long-term anticoagulant therapy.

- Service provision

i) Is there a strategy for determining which patients should and should not be treated? – Anticoagulants are more hazardous (and may be contraindicated) in patients at risk of bleeding. Need information on existing local policies, clinical guidelines, and procedures for quality assurance.

j) Are facilities for further diagnosis and treatment available? – The local situation should be checked before the committee meeting. Implications for clinical workload in primary and secondary care, for doctors and nurses, and for haematology labs. Management of any organisational changes would need to be addressed.

*Q2. Calculate and compare the positive predictive values of a diagnosis of AF, as made by a general practitioner and a practice nurse. Comment on possible implications for a screening programme based on ECGs in primary care. [20% of marks]*

[Suggested allocation: 5% for correct concept of PPV, 5% for correct calculations, 10% for implications.]

		Cardiologist:		Total	
		AF	No AF		
GP:	AF	79	114	193	PPV (GP) = 79/193 = 40.9%
	No AF	20	1241	1261	
	Total	99	1355	1454	

		Cardiologist:		Total	
		AF	No AF		
Nurse:	AF	76	203	279	PPV (nurse) = 76/279 = 27.2%
	No AF	23	1152	1175	
	Total	99	1355	1454	

Both GPs and nurses achieved similar (suboptimal) levels of sensitivity, but the specificity and therefore the positive predictive value was higher for the GPs.

Implications for the screening programme:

- Costs arising from the screening programme include diagnostic follow-up costs, which are scaled in proportion to the reciprocal of the PPV (ie number of referrals per treatable abnormality detected).
- GPs perform better but their time is more expensive. The lower PPV for nurses may be improved by seeking a second (GP) opinion on the abnormal ECGs. This would not reduce sensitivity greatly, but would achieve the higher (c.40%) PPV with GPs time.
- Neither GPs nor nurses offer the level of performance required of a diagnostic test. Consideration might be given to referring all ECGs (or all ECGs considered abnormal by the primary care team) for specialist interpretation centrally, to improve PPV.
- Whoever reads the ECGs requires specialised training and periodic updates.

*Q3. Suggest a suitable statistical test for evaluating the level of concordance between the two cardiologists. [5% marks]*

Kappa statistic. This is a measure of reproducibility, not validity, since there is no "gold standard".

$Kappa = (Obs - Exp) / (1 - Exp)$ , where:

Obs = Proportion of paired observations that are concordant

Exp = Expected proportion of concordant pairs, given the overall prevalence of AF

*Q4. Compare the rates of detection of atrial fibrillation in control and intervention practices, and among patients randomised to opportunistic or systematic screening within the intervention practices. Comment on possible explanations for these results. [25% of marks]*

[Suggested allocation: 5% for each incidence comparison, 5% for calculating 95%CI or sensitivity analysis for missing notes, 10% for commenting on possible explanations]

The key outcome is AF detected during the trial period. Thus, the appropriate denominator excludes AF known at the start. The simplest way to deal with missing case notes is by removing them from the denominator and from the numerator.

A more sophisticated "sensitivity" analysis would consider the possibilities that all the missing notes had AF, or that they all had no AF.

Control practices:

- a) missing notes excluded  $\text{Incidence} = 47/(4936-389-34) = 1.04\%$
- b) missing notes included as AF  $\text{Incidence} = 81/(4936-389) = 1.78\%$
- c) missing notes included as no AF  $\text{Incidence} = 47/(4936-389) = 1.03\%$

Intervention practices (total)

- a) missing notes excluded  $\text{Incidence} = 149/(9866-679-50) = 1.63\%$
- b) missing notes included as AF  $\text{Incidence} = 199/(9866-679) = 2.17\%$
- c) missing notes included as no AF  $\text{Incidence} = 149/(9866-679) = 1.62\%$

Candidates may be tempted to test this comparison for significance as a difference between proportions but, strictly, that would be inappropriate because of the cluster design (randomising practices, not patients).

Possible reasons for the difference between control and intervention practices:

- a) Practices differ in their background rates of AF. This can be evaluated by the prevalence of known AF at the start of the trial:  $389/(4936-34) = 7.9\%$  in control practices,  $679/(9866-50) = 6.9\%$  in intervention practices. In the same direction as the observed trial effect, but not of the same (relative) magnitude.
- b) Involvement in the trial led to greater vigilance for AF during normal clinical care in the intervention practices.
- c) Screening is effective in detecting additional cases of AF that would not be picked up in normal clinical care.
- d) Some of the AF cases recorded in the screening practices may be false positives (see Q2).

Comparison of patients individually randomised to opportunistic or systematic screening within the intervention practices can, however, be tested as a difference between proportions, for example, with missing notes excluded:

Intervention practices (opportunistic)  $\text{Incidence} = 75/(4933-340-18) = 1.639\%$

Intervention practices (systematic)  $\text{Incidence} = 74/(4933-339-32) = 1.622\%$

Difference in proportions  $= 0.01639 - 0.01622 = 0.00017$  (or 0.02%)

Variance of difference  $= 0.01639*(1-0.01639)/4575$

$+ 0.01622*(1-0.01622)/4562$

$= 0.000007 = 0.00265^2$

95%CI for difference  $= 0.00017-1.96*0.00265$  to  $0.00017+1.96*0.00265$

$= -0.005$  to  $+0.005$

Thus, the difference in annual incidence of new AF between the two modes of screening is unlikely to be greater than 0.5% in either direction.

Possible explanations for this lack of difference:

- a) Most elderly patients attend for medical care at least once in the year, so little to be gained from systematic invitations.
- b) Some cases of AF may have been diagnosed in normal clinical care before screening took place (up to two-thirds, judging from the control practices), so the potential additional impact of different modes of screening is limited.

*Q5. Write a 500-word report for consideration by your chief medical officer / director of public health, discussing the relative merits of opportunistic screening and systematic screening for AF of the elderly population to prevent cerebrovascular disease. [30% of marks]*

Advantages of opportunistic screening:

- a) Minimises inconvenience for patients
- b) High uptake at the time of medical consultation
- c) Preferentially screens those already attending with chronic conditions (including hypertension and diabetes, at high risk of CVD)

Disadvantages of opportunistic screening:

- a) Relies on patients attending health care system.
- b) Limited coverage of healthy and housebound groups.
- c) Screening tests have to be timetabled within routine clinic time.

Advantages of systematic screening:

- a) In theory, at least, comprehensive coverage
- b) Screening tests can be timetabled efficiently
- c) Easier to standardise recordings (and, perhaps, interpretation of ECGs)

Disadvantages of systematic screening:

- a) Costs of setting up patient registers, sending invitations and reminders.
- b) Additional clinic visit for patients
- c) Problems achieving full coverage, including housebound patients.

The report should present these general principles in relation to the results of the AF screening trial in Q4:

- a) Screening for AF seems to be worthwhile, in terms of detecting additional AF cases that would not be found in the course of normal clinical care, but...
- b) The mode of screening does not influence the yield greatly.
- c) The choice between opportunistic and systematic approaches to screening may therefore be made on grounds of cost-effectiveness, rather than effectiveness.
- d) The cost-effectiveness of both approaches may be improved by strategies to increase sensitivity, specificity and (particularly) predictive value of ECG interpretation in the screening situation (Q2).
- e) The decision whether or not to introduce screening at all is a more complex cost-benefit assessment, balancing costs of screening, diagnosis, treatment and follow-up against the benefits (monetary and QALY) of preventing cases of stroke.

A good report will be appropriately styled, clearly structured, and balanced in its assessment of the options.

## **EXAMINERS' COMMENTS**

### Question 1:

Generally answered well although some candidates did not mention internationally recognised screening criteria. Some candidates failed to outline the information that they would collect, or did not relate their comments to AF.

### Question 2:

Some candidates were unable to define or calculate PPVs.

### Question 4:

Most candidates correctly noted that the intervention practices identified more cases than control practices and that there was no significant difference between the two intervention groups. Few candidates addressed the issue of the missing case notes. Higher scores were given to candidates who conducted a sensitivity analysis or calculated confidence intervals.

### Question 5:

This question was generally poorly answered. Many reports were poorly structured and omitted key information. Some did not address the subject in the question. Some candidates did not allocate sufficient time to this part of the paper which carried 30% of the total marks.

**(end of document)**