PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Cost-effectiveness of a European preventive cardiology programme in primary care: A Markov Modelling Approach
AUTHORS	Mistry, Hema ; Morris, Stephen; Dyer, Matthew; Kotseva, Kornelia; Wood, David; Buxton, Martin

VERSION 1 - REVIEW

REVIEWER	Borislava Mihaylova DPhil University Research Lecturer University of Oxford, UK
REVIEW RETURNED	07-Mar-2012

THE STUDY	Thank you for the opportunity to review the manuscript "Cost-effectiveness of a European preventive cardiology programme in primary care: a Markov modelling approach". I have read the manuscript with great interest and here are my main points: 1) The manuscript presents some commendable evidence for lack of comparability between the INT and UC groups. As this is the main driver of the cost-effectiveness results and is not looked into in any detail in the EUROACTION results Lancet paper, it needs to be presented in much more detail, for example with a table of all important characteristics at baseline, including country and the estimated 10-year vascular disease risk at baseline as per D'Agostino et al. This table/differences would need a substantive discussion. 2) The adjusted analysis is central to the results/discussion of additional effectiveness, costs and finally cost-effectiveness of intervention. Yet, the actual OLS regressions, adjusting for group allocation, age, gender , etc. and the quality of these regression models are not presented. In general, is regression the most appropriate approach in this situation? 3) Given the international character of the study, it is crucial to enable interpretation as to the main drivers of costs and cost differences. Therefore a presentation of resource use (categories) across countries and intervention, and separately for the duration of the study and the extrapolation period will be informative (i.e. Nr visits, average duration, Nr CVD events/hospital admissions for vascular disease etc.). The current presentation makes it difficult to interpret the separate contribution of the programme costs and cardiovascular/other events to the incremental costs in the unadjusted and adjusted analysis. The link between health outcomes and costs should be presented in more detail. Some more minor comments: • Participants (abstract) Please rephrase to clarify that there were more participants randomised into this comparison in

	EUROACTION; the current cost-effectiveness analysis is based on those randomised participants completing 1 year follow-up. • Uncertainty: (pg 11) How was the bootstrap procedure developed? Was it based on sampling from predicted costs and effects within 11 years by (study and the Markov model;and ignoring uncertainty in Markov model)? How was the probabilistic analysis developed: was it for the 10-year post-study extrapolation period only? • I am not sure why prediction of baseline risk factors in the usual care group was needed: was it only to test the validity of D'Agostino model • Partners: The EUROACTION study did report effects on partners as well which were not taken into account in the present analysis. If done, the estimated cost-effectiveness of intervention might improve.

REVIEWER	Jeffrey A. Bakal, PhD., P.Stat. Senior Biostatistician Canadian VIGOUR Centre Faculty of Medicine and Dentistry University of Alberta Edmonton, Canada -no competing interests
REVIEW RETURNED	09-May-2012

RESULTS & CONCLUSIONS	Given the differences in Baseline conditions, is it worthwhile as a sensitivity analysis to to an age-sex matched subgroup analysis. I am a little concerned that the risk adjustment may be washing over the effect. This may also help identify if there is a general subgroup of patients who may benefit from this.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

1) The manuscript presents some commendable evidence for lack of comparability between the INT and UC groups. As this is the main driver of the cost-effectiveness results and is not looked into in any detail in the EUROACTION results Lancet paper, it needs to be presented in much more detail, for example with a table of all important characteristics at baseline, including country and the estimated 10-year vascular disease risk at baseline as per D'Agostino et al. This table/differences would need a substantive discussion.

Our response: We have taken on board comments from the reviewer and have added a new table into the paper showing the baseline characteristics between the intervention group and the UC subsample (see Table 2). We have also added in the following text into the start of the results section: "The baseline characteristics for the intervention group as a whole and the usual care subsample who were seen at baseline are shown in Table 2. There were significant differences in the distribution between countries. Mean total and HDL cholesterol levels were significantly higher for the intervention compared with the UC group. Whilst no statistically significant differences were observed for other baseline characteristics, but the 10-year CVD risk at baseline [5] was numerically higher for the UC group than the intervention arm."
These baseline differences between the two groups is further emphasised in the second paragraph in the discussion.

2) The adjusted analysis is central to the results/discussion of additional effectiveness, costs and

finally cost-effectiveness of intervention. Yet, the actual OLS regressions, adjusting for group allocation, age, gender , etc. and the quality of these regression models are not presented. In general, is regression the most appropriate approach in this situation?

Our response: We used OLS regressions to adjust for group allocation, country and characteristics. We have included one regression model for costs and QALYs (assuming that the duration of effect of intervention beyond the end of trial doesn't last more than 0 years) as an example in the appendix (see Table A3). However, we haven't added all the other regression models due to space constraints, but the readers can get a feel for what the regression model looks like. We have added in the results section the following text:
"(an example of the various regression models is shown in the Appendix)"

3) Given the international character of the study, it is crucial to enable interpretation as to the main drivers of costs and cost differences. Therefore a presentation of resource use (categories) across countries and intervention, and separately for the duration of the study and the extrapolation period will be informative (i.e. Nr visits, average duration, Nr CVD events/hospital admissions for vascular disease etc.). The current presentation makes it difficult to interpret the separate contribution of the programme costs and cardiovascular/other events to the incremental costs in the unadjusted and adjusted analysis. The link between health outcomes and costs should be presented in more detail.

Our response: We have taken into account the reviewers comments and included two new figures. Figure 1a shows the observed costs split by type of cost for the intervention and usual care arm and the accompanying text has been added to the methods section.
"Figure 1a shows that the 1-year observed costs (split by type of cost) for the intervention group was significantly more than the usual care group for all countries. This higher cost was explained by the EUROACTION intervention programme costs and contacts with EUROACTION staff, whilst neither arms experienced significantly high cost cardiac interventions or cardiac medications."
Figure 1b, further emphasises the mean costs for two groups in the model and the following text has been added to the results section:
"Figure 1b further emphasises that the observed additional costs of the EUROACTION intervention programme and staff costs were not offset by the estimated reduced costs of cardiac interventions in the subsequent years."

Some more minor comments:
• Participants (abstract) Please rephrase to clarify that there were more participants randomised into this comparison in EUROACTION; the current cost-effectiveness analysis is based on those randomised participants completing 1 year follow-up.

Our response: The reviewer is right in saying that the current cost-effectiveness analysis is based on randomised participants completing 1 year follow-up. We have revised the wording in the abstract to reflect this:
"1,019 patients who were randomised to the EUROACTION intervention programme and 1,005 patients to usual care and who completed the one-year follow-up".

• Uncertainty: (pg 11) How was the bootstrap procedure developed? Was it based on sampling from predicted costs and effects within 11 years by (study and the Markov model; and ignoring uncertainty in Markov model)? How was the probabilistic analysis developed: was it for the 10-year post-study extrapolation period only?

Our response: We have revised the text to say:
"We represented uncertainty due to sampling variation in both the unadjusted and adjusted cost-

effectiveness ratios using non-parametric bootstrapping. In the unadjusted analyses we sampled individuals in our model with replacement and used their costs and outcomes over the 11-year period to compute replications of the incremental cost per QALY gained. We repeated this approach in the adjusted analyses, also adding the regressions to control for confounding factors. In each case, we generated 10,000 bootstrap replications of the cost-effectiveness ratios and used these to construct 95% confidence intervals around the point estimate of cost-effectiveness."

• I am not sure why prediction of baseline risk factors in the usual care group was needed: was it only to test the validity of D'Agostino model

Our response: The prediction of baseline risk factors as measured by the D'Agostino model in the usual care group (for the sub-sample only) was done not only to test the validity of the D'Agostino model, but also to show a comparison with the intervention arm, now shown in Table 2. Therefore, we have felt it was appropriate to leave this part in the paper.

• Partners: The EUROACTION study did report effects on partners as well which were not taken into account in the present analysis. If done, the estimated cost-effectiveness of intervention might improve.

Our response: Our analysis was restricted to patients only and did not include partners. This was because in the usual group, we had a sub-sample of patients who had baseline assessments, but none of their partners had baseline assessments. So for the usual care arm we did not have any baseline data for partners and therefore it was not appropriate to include them in the analysis as there was not enough information to estimate the cost-effectiveness of the intervention.
The following paragraph has been added to the text in the discussion section:
"Our cost-effectiveness analysis did not include partners. If partners were included it might improve the cost-effectiveness, but we have no good measure of the effect on partners to know how substantial the impact on the incremental cost-effectiveness ratio might be".

Reviewer 2:
Given the differences in Baseline conditions, is it worthwhile as a sensitivity analysis to an age-sex matched subgroup analysis. I am a little concerned that the risk adjustment may be washing over the effect. This may also help identify if there is a general subgroup of patients who may benefit from this.

Our response: We conducted sensitivity analyses to see whether age-sex matched subgroup analysis would have any effect if the duration of the intervention lasted beyond the 1-year trial period.
The following paragraph has been added to the text in the results section and a new table has been added to the appendix (Table A5):
"Due to baseline differences, we conducted age-sex matched subgroup analyses and the adjusted results confirmed that the intervention remained dominated, even when an optimistic timeframe was considered (an example of age-sex matched subgroup analysis is shown in the Appendix)."
To further emphasise that the intervention is dominated by usual care, new text has been added to the results section:
"Although there is considerable uncertainty around those point estimates with the 95% confidence intervals ranging from acceptably cost-effective to highly dominated, but the probability of being cost-effective are very low, as shown in the adjusted CEACs in Figure 2b".

Minor amendments to paper made by authors
• We have also updated the author list of institutions on the first page of the paper.
• We have removed the word HEARTSCORE from page 6 as this was not part of the EuroAction

programme.
• We have also had to renumber the tables and figures to take into account the extra analysis.

VERSION 2 – REVIEW

REVIEWER	Borislava Mihaylova University Research Lecturer University of Oxford United Kingdom
REVIEW RETURNED	23-Jun-2012

THE STUDY	Thank you. I reviewed with interest the revised version of the manuscript and here are my comments: (1) I have serious concerns as to the ability of the analytical framework presented to evaluate cost-effectiveness of the preventive cardiology programme in primary care in the EUROACTION study: a. The EUROACTION study reported improvements (although not always statistically significant) in virtually all measured risk factors (Table 3 in Lancet 371:1999-2012). The same manuscript also suggests improvement from baseline in all of these risk factors in the intervention group alone from baseline to 1-yer (see Tables 2 and 3, Figure 3 in Lancet 371:1999-2012). b. Framingham risk equations will produce improved health outcomes for improved risk factor profile. c. Although the multivariate nature of cardiovascular risk might in some extreme conditions produce opposite results when one looks into individual factors and then into combined impact of risk profile at individual level, this is not likely and needs to be based on very clear justification as to why this might be the case (including presenting values of all used risk factors to propagate effects of intervention (lipids, blood pressure etc.) with their values at baseline and at one year in the two groups). d. The manuscript claims to adjust for differences between the intervention groups and presents impacts on QALYs that have substantial probability mass into the negative incremental benefits section (indeed with negative mean incremental QALYs). e. Given a. above, an analysis that eliminates all misbalances between a treatment and a control group is actually the one modelling impact of changes in risk factors in the same population. For example, a model can be applied only to intervention group participants and used to estimate “usual care” scenario by keeping risk factors at their values at baseline for people in the intervention group, and “intervention” scenario by propagating a scenario with risk factors at their 1-year values for the same people in the intervention group. f. It seems to me that the Markov model is fully deterministic: for the same set of risk factors it produces a single estimate of expected QALY, expected cost etc. In that respect it ignores the uncertainty in parameters, the Framingham equation parameters for example (where all parameters are estimated with some uncertainty, ideally represented by the variance-covariance matrix). Thus, the uncertainty in QALYs, costs and cost-effectiveness will be underestimated by accounting only for the sampling uncertainty in the risk profiles. g. The manuscript outlines a framework that uses individual participant study data at the level of risk profile and then attaches

	average/expected survival, QALY, cost over 11-years based on a model (at which level uncertainty in these is already underestimated) and then runs series of regressions to adjust for differences at baseline. Specifying OLS regressions on observed individual participant outcomes, costs etc is likely challenging (and should be very strongly supported by evidence that these models were well specified and able to produce reliable and unbiased estimates for treatment allocation; the single model added into the appendix does not include any diagnostic/model fit information.), but specifying these on modelled data might compound even further the issues as to the reliability of estimated parameters on treatment allocation. Minor issue: Some of the characteristics in Table 2 are available for all "usual care" participants included in current analyses and should be fully presented (e.g. country, gender, age etc.)

VERSION 2 – AUTHOR RESPONSE

Response to reviewer's comments (Reviewer: Borislava Mihaylova)

(1) I have serious concerns as to the ability of the analytical framework presented to evaluate cost-effectiveness of the preventive cardiology programme in primary care in the EUROACTION study:

a. The EUROACTION study reported improvements (although not always statistically significant) in virtually all measured risk factors (Table 3 in Lancet 371:1999-2012). The same manuscript also suggests improvement from baseline in all of these risk factors in the intervention group alone from baseline to 1-year (see Tables 2 and 3, Figure 3 in Lancet 371:1999-2012).

Our response: Yes, we agree with this point made by the reviewer.

b. Framingham risk equations will produce improved health outcomes for improved risk factor profile.

Our response: Yes, we agree with this point made by the reviewer.

c. Although the multivariate nature of cardiovascular risk might in some extreme conditions produce opposite results when one looks into individual factors and then into combined impact of risk profile at individual level, this is not likely and needs to be based on very clear justification as to why this might be the case (including presenting values of all used risk factors to propagate effects of intervention (lipids, blood pressure etc.) with their values at baseline and at one year in the two groups).

Our response: We do not accept this interpretation of our data as there are important differences between outcomes in the EUROACTION trial and those from the health economics analysis which is a function of the variables used. In the EUROACTION trial we obtained impressive differences in diet and physical activity between intervention and usual care but no significant difference in prevalence of smoking. The only lifestyle variable in the Framingham function is smoking. There is no measure of diet or physical activity. So our clinically important outcomes in terms of proportions of patients achieving the diet and physical activity targets could not be taken into account in the health economics analysis. Proportions of patients achieving the blood pressure and lipid targets were outcomes for EUROACTION and these variables are also included in the Framingham function but there is an important caveat to the analysis and interpretation of the lipid results. By the play of chance in this cluster trial the proportions of patients at target for lipids at baseline was significantly different between intervention and usual care and in favour of intervention. Therefore, it was not possible to draw any conclusions from one year outcomes, comparing intervention with usual care,

given this imbalance at baseline. We compared intervention with usual care for changes over time from baseline to one year and this showed statistically significant, but clinically small, improvements in proportions achieving lipid targets. This secondary analysis of changes over time was limited in terms of power by the fact that we only collected baseline data on a random sub-sample (25%) of the usual care patients. So the health economics analysis, based on Framingham, was necessarily limited to smoking, blood pressure, lipids, diabetes and anti-hypertensive medications together with age and sex. We had no effect on smoking and neither diet nor physical activity could be taken into account.

d. The manuscript claims to adjust for differences between the intervention groups and presents impacts on QALYs that have substantial probability mass into the negative incremental benefits section (indeed with negative mean incremental QALYs).

Our response: As already explained above the two groups were not identical to start with as they were not randomised at patient level. The unadjusted analyses showed that for the intervention group the mean QALYs were greater than for the UC group. However, we have tried to account for this imbalance between the two groups by adjusting for baseline differences which is common practice [Manca et al, Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health Economics 2005; 14:487-496] and this has resulted in the intervention group gaining fewer mean QALYs than the UC group, hence we have negative mean QALYs.

e. Given a. above, an analysis that eliminates all misbalances between a treatment and a control group is actually the one modelling impact of changes in risk factors in the same population. For example, a model can be applied only to intervention group participants and used to estimate “usual care” scenario by keeping risk factors at their values at baseline for people in the intervention group, and “intervention” scenario by propagating a scenario with risk factors at their 1-year values for the same people in the intervention group.

Our response: We do not understand the point being made by the reviewer but maybe our response to point (c) above helps to explain the apparent discordance between our principal clinical results and the health economics outcomes.

f. It seems to me that the Markov model is fully deterministic: for the same set of risk factors it produces a single estimate of expected QALY, expected cost etc. In that respect it ignores the uncertainty in parameters, the Framingham equation parameters for example (where all parameters are estimated with some uncertainty, ideally represented by the variance-covariance matrix). Thus, the uncertainty in QALYs, costs and cost-effectiveness will be underestimated by accounting only for the sampling uncertainty in the risk profiles.

Our response: We agree with the reviewer that the Markov model is deterministic and hence we may have underestimated some of the sampling uncertainty in the risk profiles. However, we have tried to take into account of this uncertainty in the expected costs, QALYs and cost-effectiveness ratios for both the unadjusted and adjusted analyses by the use of non-parametric bootstrapping.

g. The manuscript outlines a framework that uses individual participant study data at the level of risk profile and then attaches average/expected survival, QALY, cost over 11-years based on a model (at which level uncertainty in these is already underestimated) and then runs series of regressions to adjust for differences at baseline. Specifying OLS regressions on observed individual participant outcomes, costs etc is likely challenging (and should be very strongly supported by evidence that these models were well specified and able to produce reliable and unbiased estimates for treatment allocation; the single model added into the appendix does not include any diagnostic/model fit information.), but specifying these on modelled data might compound even further the issues as to the

reliability of estimated parameters on treatment allocation.

Our response: For both costs and QALYS, we tested for normality (histograms, kernel-density plots and Shapiro-Wilk test), heteroscedasticity (Cameron and Trivedi decomposition test) and model fit (link test and Ramsey RESET test). For normality the graphs were adequately normal; for heteroscedasticity the p value for the test was not significant (costs: p value = 0.967; QALYs: p value = 0.835), hence data was not heteroscedastic; and while the p value for both the link and Ramsey RESET test for the model fit were lower than 0.05 suggesting we could improved the model fit by adding a few more variables which were not collected in the dataset.

Minor issue: Some of the characteristics in Table 2 are available for all "usual care" participants included in current analyses and should be fully presented (e.g. country, gender, age etc.)

Our response: Table 2 has been updated.

Cost-effectiveness of a European preventive cardiology programme in primary care: a Markov modelling approach

Hema Mistry, Stephen Morris, Matthew Dyer, et al.

BMJ Open 2012 2:
doi: 10.1136/bmjopen-2012-001029