Reviewer’s report

Title: The effects of a multi-disciplinary, minimal smoking intervention among pregnant women and their partners: A real-life controlled intervention study

Version: 1 Date: 31 January 2008

Reviewer: Gerard J Van Breukelen

Reviewer’s report:

Major compulsory revisions
The study design, methods of analysis and results are not reported in a clear way, leading to several ambiguities and uncertainties about the correctness of the authors’ conclusions. This paper needs very careful revision on methods and results, before the conclusion and discussion can be evaluated at all. A list of unclarities is given on pages 2-5 of this report.

Level of interest
Unable to judge on this, since I am a statistician, not a public health researcher. The idea to study effects in a cohort study instead of an RCT for reasons mentioned in the Discussion, sounds interesting to me, although I do not fully agree with the reasons given (see comment on page 11). Whether the intervention itself is of sufficient interest to the audience of BMC is up to the judgement of public health researchers.

Statistical review
Is needed and done by myself, see the list of comments on pages 2-5 of this report

List of comments (p = page, l = line. If l < 0, then count from the bottom of the page)

p. 2, Methods: state clearly whether treatment assignment was by randomization or based on pre-existing groups or cohorts.

p. 2, Results:
For the baseline (i.e. at start pregnancy) smoking percentages are reported, whereas only
an odds ratio is reported for the 2nd time point (during pregnancy) and no results at all are given for the 3rd time point (after delivery). This is confusing. Given percentages and odds ratio for each time point. Partner smoking is reported unclearly too: is the OR of intervention relative to control, and if so, at what time point? Or is it at time point 2 or 3 relative to baseline and if so, for which cohort?

p. 3, l. 2-3: This conclusion is weaker than that on page 13. Further, conclusions about the effect of the intervention should be based on a comparison between the treated and control cohorts within this study. It is true that Table 3 suggests a somewhat higher quitting rate during pregnancy in Trondheim (and the PACT cohort) compared with Bergen or Norway. However, the difference in quitting rates between treated and control cohort in Table 4 is small relative to the difference of both with Norway and Bergen in Table 4. This suggests that some common influence in both PACT cohorts, e.g. selection bias, self-report bias, or a placebo effect, explains the higher quitting rates in Trondheim and the study population during pregnancy at least as much as the intervention does.

p. 4, l. -4 to p. 5, l. 4: so there are two control cohorts: one of N=1788 pregnant women and one of N=2116 women after delivery. The relevance of the second cohort to the evaluation of the intervention is unclear to me. Unfortunately, both cohorts are pooled in Table 1, column Control cohort, questionnaire 6 wks after delivery (n=3139). In contrast, table 4 appears to include the first cohort only, which is the correct method.

p. 4, l. 1 and l. 9 the nrs here imply a dropout rate of 37% and 45% in control and treated cohort, respectively. This calls for a very careful analysis of bias in effect
estimates due to selective dropout, using proper methods for including dropouts (e.g. mixed logistic regression with 3 repeated measures of smoking, and including dropouts without imputing missing values).

Further, the initial N's of 1788 and 2051 are higher than those in Table 4, column 2, which also calls for clarification.

p. 6: it is unclear from this page how the primary outcome smoking was exactly defined (e.g. smoking at least one cigarette in the last month versus weekly smoking), and whether it was defined and recorded in exactly the same way in both cohorts and also in Norway, Bergen and Trondheim. If methods differ between cohorts, this may cause differences in smoking prevalences.

p. 7: Although the two study cohorts are fairly large, a formal sample size calculation is missing. Small effects on binary outcomes require very large samples. Adjustment for confounders, which is indispensable to any nonrandomized study, further increases the total sample size needed (see the Variance Inflation Factor VIF of a treatment indicator after including covariates). Therefore, a proper sample size calculation should be included, if necessary a post-hoc one using the VIF from the data, but using the smallest clinically relevant effect size rather than the present effect size estimate. Note that using the effect size estimate from the sample will give a power of < 50% for non-significant effects and power of > 50% for significant effects, by definition and is therefore useless.

p. 7, l. 7-10: Apparently, very simple statistical methods were used that ignore both the presence of potential confounders and the nesting of women within general practices, midwive
practices or health centres (see p. 4). Ignoring the nesting can lead to underestimation of
standard errors with a higher risk of type I errors and too small confidence intervals as a
result. Ignoring confounders can seriously bias the effect estimates. Given the difference
in smoking rates at the start of pregnancy mentioned on page 2 (l. -6, -7) which, incidentally, differ from those in Table 4, the results during and after pregnancy must be
adjusted for smoking at the start of pregnancy and baseline variables on which the two
cohorts differ. Inclusion of other baseline variables may further be useful to enhance
power if these other variables are predictive of smoking during/after pregnancy. Also
include type of centre (gp versus midwive vs health centre) as covariate using dummy
coding or dichotomisation.

p. 7, l. 11-14:
- Why use UNIANOVA given that the outcome is binary ?
- Specify the logistic regression models: what covariates were included in what form (e.g.
quantitative or using dummy coding). Was a separate analysis done per time point ? If so,
why not use mixed logistic regression with the three repeated measures nested within
women nested within practices or health centres ? This would also solve or at least
reduce the problem of bias due to dropout.
- Was alfa = 5% one- or two-tailed ? Please use two-tailed tests to be consistent with
confidence intervals and to be able to detect possible adverse effects of treatment.

p. 8, l. 5: average nr of cigarettes a day: is this including or excluding nonsmokers ? If
included, this measure is a mix-up of two outcomes: smoking yes/no plus amount of
smoking among smokers. Further, it will probably be strongly skewed, rendering standard statistical methods dubious. Consider data transformation then (e.g. log
or sqrt).
p. 8, l. 9: "... augmented reduction .... in 2003" is not very clear from Table 3.
Smoking prevalence appears to drop almost linearly over the years in Norway, Bergen and Trondheim, although a trace of stronger drop from 2003 to 2004 is visible. But is it also significant? Test by e.g. linear trend analysis with an extra dummy indicator for 2004. The latter should be significantly predictive. Given the aggregated nature of these data, with heterogeneity of variance between years (since variance = N*p*(1-p), where p = proportion smokers), use either logistic regression for aggregated data or weighted least squares linear regression (the prevalences are too high for Poisson regression).
p. 8, l. -9, -8: can you explain this higher smoking prevalence at the start of pregnancy in your PACT cohort compared with Trondheim as a whole? It might indicate selection bias at baseline, with regression to the mean causing an apparent beneficial treatment or PACT cohort effect at follow-ups. Table 3 shows that the baseline difference between PACT and Trondheim has disappeared during pregnancy, but such effect could occur due to regression to the mean (see e.g. Stigler, Statistical Methods in Medical Research, 1997, Van Breukelen, J Clin Epidemiology 2006, Senn, Statistics in Medicine, 2006).
p. 9, l. 4: According to Table 4, the 40% must be 14% (i.e. the quitting rate 69.7% is 1.14 times as large as 61.3%, both % being based on the nrs of quitters within the baseline smokers subgroups of n= 475 and 462). Based on the total samples of 1729 and 1940 gives a quitting rate of 89.3% for controls and 92.8% for treated according to Table 4, and this is not a difference of 40% either.
p. 9, l. 4: OR = 1.4 must be 1.45 according to Table 4 (i.e. 322/140 divided by 291/184), at least unadjusted for covariates.
p. 9, l. 5 till bottom: these nrs are not verifiable without a clear report of the
regression
modeling procedure used (initial model, model reduction procedure, final model) and the
final results (= predictors included, with their unit of measurement or coding, B, SE, p).
The report should also clearly state which N was included into that analysis for each
cohort, which time points, and how dropouts were treated in the analysis.
p. 10 l. 3-5:
But note that the difference between treated and control PACT cohorts is small relative to
the difference between control PACT and Norway or Bergen, see comment on p. 3.
p. 11, l. 1:
Selection bias may have occurred in your study just as well as in an RCT, but it cannot be
judged from this report, since you neither say what % of the women invited to participate
refused, nor report a baseline comparison between refusers and consenters (which should
be done per cohort, i.e. treated and controls apart).
p. 13, conclusion: is too strong, see comments on pages 3, 10, 11.
table 1:
- results of the baseline measurement (at start of pregnancy) are missing
- the two control groups of p 5 lines 1-2 should be kept apart rather than pooling them
into one group of 3139 (= 1023 + 2116) women. Perhaps the group of 2116 women
having delivered before baseline should be excluded from this Table altogether since
their smoking results must have been retrospectively obtained, with potential bias.
- nr of cigarettes a day = including or excluding nonsmokers ? Also, the n's mentioned
here are very small and suggest errors of analysis or typing.
Table 3:
- how are the sample sizes and smoking rates per year per PACT cohort (control vs
treated) related to the nrs on page 4-5 and in Table 4 ? In answering take care
that the results must be consistent with each other.

Table 4:
- Sample sizes for both PACT cohorts at start of pregnancy differ from those stated on pages 4-5 (i.e. 1729 versus 1788 and 1940 versus 2051).
- why are results from Trondheim missing in this Table? Table 3 includes Trondheim.

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Quality of written English: Acceptable

Declaration of competing interests:
I declare that I have no competing interests