Author's response to reviews

Title: Development of burnout over time and the causal order of the three dimensions of burnout among male and female GPs. A three wave panel study.

Authors:

Inge Houkes (Inge.Houkes@socmed.unimaas.nl)
Yvonne Winants (Yvonne.Winants@maastrichtuniversity.nl)
Mascha Twellaar (Mascha.Twellaar@maastrichtuniversity.nl)
Petra Verdonk (Petra.Verdonk@maastrichtuniversity.nl)

Version: 3 Date: 10 February 2011

Author's response to reviews: see over
Dear Prof. Zauner,

Please find attached the second revision of our manuscript entitled *Development of burnout over time and the causal order of the three dimensions of burnout among male and female GPs. A three wave panel study* (MS 5636468632975159).

Below we have indicated how we addressed the reviewers’ comments. Furthermore, we have indicated the changes in the manuscript by means of ‘track changes’ and we checked whether the manuscript conforms to the journal style.

**Referee 2 (Olaf G Aasland)**

□ The reviewer suggests to search for and possibly include publications of Langballe, Innstrand and colleagues about burnout among physicians (as well as other professional groups). In the revised version of the manuscript we refer to the following (very interesting) studies of this research group:


Langballe EM, Innstrand ST, Aasland OG, Falkum E: The predictive value of individual factors, work-related factors, and work-home interaction on burnout in female and male physicians. Stress Health [Internet]. 2010 May 6. (Reference #40)


We discuss these studies in the Introduction section and particularly in the Discussion section of the manuscript (explaining differences between males and females in the prevalence and development of burnout).

□ The reviewer thinks that our statement about the MBI instrument is an exaggeration. We agree with the reviewer and we adjusted our formulation of this statement (see method section/Measures).

**Referee 3 (Matthew Hotopf, replacing Marcia Scazufca, who was referee 1 in the first review)**

In line with the reviewer’s suggestion, we mention in the manuscript that Dutch law does not require ethics committee approval for this sort of study (see Method section/Design, procedure and participants).

**Referee 4 (Jeremy Miles)**
Reviewer 4 had several valuable suggestions regarding the statistical aspects of the manuscript. Please find our responses to these suggestions below.

**Major compulsory revisions:**

- The reviewer mentions correctly that the phrase “we did not randomize with respect to sex” is unclear. What we meant is that we took a disproportionate sample (as regards sex) from the GP population, because we were mainly interested in studying differences between two strata (men and women). The percentage of female GPs in our sample was higher than in the Dutch GP population. We adjusted the formulation of this phrase in the manuscript (Method section/Design, procedure and participants).

- We handled missing data by means of listwise deletion in all analyses, and we used maximum likelihood as the method of estimation in the SEM. The final sample we worked with (i.e., the panel group, the respondents who participated in all three measurements) hardly contained missing data, neither for men nor for women. Hence, listwise deletion did not lead to the loss of participants. Using any other option for treating missing data (e.g., pairwise deletion, mean substitution) would not have changed the results. What we meant here was that the dropout between the three measurement points (attrition) was slightly higher for men than for women, which may have led to selection bias. Also see the discussion of this issue on page 23 of the manuscript.

- The reviewer suggests to add the Standardized Root Mean square Residual (SRMR) as a measure of model fit, and hence we added this fit measure to the manuscript (see Method section/Data analyses and Table 4).

- The reviewer correctly states that the Akaike Information Criterion (AIC) is used to compare competing models. We adjusted the text in the Method section/Data analyses and Results section accordingly. We now use AIC as a comparative fit index.

- The sentence “We did not specify the measurement models of the three burnout measures” means that we simplified the covariance structure in the LISREL analyses by assuming that the latent and observed variables were identical. In other words, we did not include measurement models in our analyses, but only structural equation models (cf. Jöreskog & Sörbom, 1993). Simultaneous consideration of all observed variables (i.e., including all measurement models and covariance structure models in one analysis) would result in unreliable parameter estimates and insufficient power (cf. De Jonge et al., 2001; Schumacker & Lomax, 1996). This procedure seems justified because the burnout measures used in this study have proven to be valid and reliable in this study and in previous studies.

- Sample sizes were 104 for the men and 108 for the women, these samples are not too large indeed. For various reasons, however, we decided to stick to the separate analyses for men and women. Messing et al. (2003, reference #8) for instance suggest that studies that aim to detect differences between the sexes should analyze men and women separately. To our knowledge, sex differences in the development of burnout have not been studied before, so our study is rather explorative in this regard. Applying multi sample analyses (MSA) would
complicate the model specification as we have no knowledge yet about what parameters are equivalent for men and women (and hence should be constrained to be equal) and what parameters are not (cf. Schumacker & Lomax, 1996). Further, this procedure seems justified because ML estimation in combination with samples sizes of 100 or more is rather robust against nonconvergence and improper solutions (Boomsma & Hoogland, 2001). All models in our analyses converged within 10 iterations. Nevertheless, when we have more information about the theoretical models for men and women, MSA could be an interesting feature to use in future research.

Stationarity: unfortunately we are not completely certain about what is meant here. Should we have tested for stationarity of the pattern of relationships? Knowing that the mean levels of burnout fluctuate over time and the time lag between measurement points is two years, we did not assume beforehand that the effects from time 1 to time 2 were different from or similar to the effects from time 2 to time 3, or time 1 to time 3 for that matter. Our study can be considered explorative with regard to this aspect. Therefore, we did not test for stationarity.

Minor essential revisions:

- Yes indeed, we used maximum likelihood (ML) as method of estimation in the LISREL analyses. We added this information to the manuscript (Method section/Data analyses).
- In accordance with the reviewer’s suggestion we added the precise p-value of the F-statistics to Table 2, and we added the effect sizes for the contrasts to Table 2 as well (also see suggestions of Andy Field, 2005 regarding this issue).

We would like to thank all reviewers for reading and commenting on the manuscript. We look forward to receiving your decision regarding this revised manuscript.

Sincerely yours,

Inge Houkes,
Also on behalf of Yvonne Winants, Mascha Twellaar and Petra Verdonk

Maastricht University
CAPHRI Research School
Faculty of Health, Medicine and Life Sciences
Department of Social Medicine
PO Box 616
6200 MD Maastricht
The Netherlands
E-mail: Inge.Houkes@maastrichtuniversity.nl
Tel: +31 43 388 2878