Author's response to reviews

Title: Coping and back problems: Part 1. Analysis of multiple data sources on an entire cross-sectional cohort of Swedish military recruits.

Authors:

Charlotte Leboeuf-Yde (chyd@shf.fyns-amt.dk)
Kristian Larsen (kristian.larsen@svf.au.dk)
Ingvar Ahlstrand (Ingvar.ahlstrand@piktverket.se)
Ernest Volinn (Ernest.Volinn@hsc.utah.edu)

Version: 2 Date: 16 January 2006

Author's response to reviews: see over
Response to BMCreview

Title: Coping and back problems: Part 1. Analysis of multiple data sources on an entire cross-sectional cohort of Swedish military recruits.

Response to Alex Burdorf

Major compulsory revisions

Comment 1: “The content of the coping variable should be presented, otherwise the results cannot be interpreted (this is a black-box model approach).”

Response: As fully explained in the text, we are well aware of this being a “black-box model”. However, we argue that 1. the overall definition of coping being “coping with stress during war-like situations” does give us an overall idea of what we are dealing with, 2. the results are interesting – even if we are dealing with a “black box”, because we have found a variable that has such a strong levelling effect on others in multi variate analysis, 3. the results are interesting also because the pattern of associations is so similar in back problems and ill health, 4. the results are worthwhile publishing because we are not satisfied with the “black box” but go on to testing the coping variable in part 2, where we are using a well-known coping questionnaire (SOC -13), as the reviewer knows, because he is also reviewing part 2.

In order to make it perfectly clear to the readers that we understand the problem with the “black box”, we have emphasized this in the text (p.14 last para): “Interesting as our results may be, they cannot really be taken seriously, until the coping variable is properly defined and described but, unfortunately, the Swedish Army is not able to divulge its interview questions.” In addition, in our conclusions, we now always describe coping as “coping, as defined by the Swedish Army”, to make it clear to everybody that the results are not necessarily transferable to any definition of coping. We have also added a comment under our conclusions: “The “black box” of coping, as defined by the Swedish Army, must be defined and described.

Comment 2: “Although the authors point at the cross-sectional design in the discussion, the text used words as predictor etc incorrectly.”

Response: We are well aware of the difference between mere associations and risk factors. In order to make this clear, we have now specified in the text that we are working with “potential predictor variables” and that we will call these “predictors” in the text (p.6, 2nd para). In our discussion we also spell it out that we are not testing for a causal relationship (p.14, 3rd para, 2nd sentence). We are therefore satisfied that there should be no misunderstandings in relation to this.

Comment 3: “In the introduction the onset, aggravation and consequences of back pain are mixed up. Coping has been demonstrated to affect chronicity and consequences but not onset of LBP.”

Response: There is no confusion at all in our minds about onset, aggravation and consequences of back pain, and we do not believe that we have in any way mixed them up. We have, however, attempted to simplify the Background text, that may have been a bit too fancy in style, and hope that the clarity of our minds comes through this time. We have had the text read by three other
(naïve) people, who none of them thought that we had mixed up the onset, aggravation and consequences of back pain, so hopefully, this time we succeeded!

Comment 4: The terminology is sometimes rather sloppy, eg psychosocial factors include intellectual capacity and education on page 1, and are defined differently on page 2. (in my opinion capacity and education are NOT psychosocial factors).

Response: The terms psychological and sociological factors have now been kept strictly apart. We can find no other “sloppy” terminology but have been very careful about being consistent throughout the text.

Comment 5: “The statistical analysis is very troublesome. the 5 determinants ahve pre-defined cut-off pointes and a polynomial analysis was only used to evaluate whether this seems reasonabale, since there is no formal evaluation whether fit increases significantly etc, I cannot see the use of it.”

Response: Statistical analysis can sometimes be very troublesome, because different people may choose different approaches. However, we have used the “recipe” as described in our statistical reference. In order to improve the understanding of why we selected this method, we have added a whole paragraph already in the Background, on our statistical approach. Obviously, in order to write about it at such length, we consider our statistical approach to be a strength and not a weakness. (p.5, 2nd para)

It is not correct that all our 5 determinants had predefined cut points. Height, weight, intellectual capacity, and coping did not. We could have chosen to analyze these as bivariate variables or we could have decided to use their mean values. Instead we decided to subclassify them as shown in Tables 3 and 4. Our analytic strategy assured us that this method gave the best “view” of the associations.

Comment 6: “The results are not well presented:
- confidence intervals are lacking
- number of cases across levels of determinants are missing
- ill health cases et are unknown (the definition is also a concern given the association with coping: I expect that ill health and coping have a strong common underlying concept by definition)”

Response: This comment deals with two issues: a. the presentation of results and b. the definition of a variable plus its interpretation.

In relation to a. confidence intervals are lacking because, as described in the text (p.10, 3rd para) they are meaninglessly small in such a large sample. The number of participants throughout the analyses has been added both in the text (p.10, Results, 1st para) and in the tables 3 and 4.

In relation to b. ill health cases are unknown. This study is not about specific diseases but about back problems (in general) and ill health (in general). Yes, we agree that ill health and coping may be but the expression of the same condition, in many people. This has been added to the text (p.13, 3rd sentence from the bottom of the page) and again, on p.14, 3rd para, 2nd last sentence.

Comment 7: “An OR of 134 is very troublesome when we do not know the content of both the determinant and the outcome”
Response: An OR of 134 is very troublesome indeed, but we are confident that we have covered our tracks by explaining this in the text. It would be particularly troublesome if we claimed that coping causes ill health. We do not claim this. We note that there are some variables that are positively associated with ill health and back problems but that these associations level out when we control for coping. If somebody wants to investigate this further, it would be necessary to a) define coping b) look at specific diseases, to see if this pattern arises again and if so, for which diseases.

Comment 8: “The rationale about differences in importance of psychosocial factors and occupational exposure strikes me as odd, when one selects a study population without any occupational exposure! The fact that LBP occurs at early age, does not say anything about association MMH and LBP (that depends on the increase of LBP among exposed).”

We agree that this argument is premature and that it can be misconstrued to look as if we believe that occupational exposure has no deleterious effect on the spinal structures. We have removed the offending text, with the exception of mentioning that we are dealing with a young pre-work cohort, which "enabled us to avoid the confounding element of occupational exposure" (p.14, 1st para), which obviously is a strength of the study, as it ensured homogeneity of the cohort in terms of this aspect.

Response to Michael Höfler

Major compulsory revisions

Comment 1: “The Term “risk factor” should be used as in Kramer HC et al “Coming to terms with terms of the risk” (Archives of General Psychiatry, 1997) that is, as a factor that precedes the outcome and can be shown to be associated with an increased risk of the outcome. It is difficult to establish temporal directions in cross-sectional studies as the present one. How can you know that e.g. obesity precedes back problems, and not vice versa. Therefore, your paper does not contribute much to the distinction between risk factors and indicators.”

Response: As we have previously written, in our response to Alex Burdorf (comment 2), we are well aware of the definitions of cause and association (risk factors and risk indicators). It is obviously true that one cannot know the direction of an association, or even if there is a direction. However, if a variable is commonly suspected in the scientific /clinical community for being a risk factor, such as many people suspect obesity to be a cause of low back pain, and if a positive association is detected between this suspected risk factor and the outcome variable BUT if this association disappears when challenged with an extraneous variable (e.g. coping) then you can be fairly sure that the original association is not a causal one. In our amended manuscript we have therefore emphasized that we are particularly looking at coping and its ability to affewct the other potential predictors.

We have gone through the text with a very fine comb to ensure that there is no hint of text that can make anybody suspect that we are as stupid as to believe that we are looking for cause of disease!
Comment 2: “Although the authors cannot establish temporality they even make conclusions on causality. They claim that back problems and ill health may have a similar causal mechanism because they have some shared risk factors. First, “ill health” is a summary of various diseases with different etiologies (to a smaller degree this is probably also the case for “back problems”). Second, the authors’ findings can have various other explanations like common biases or common risk factors that affect this factors as well as ilnes/back problems. Avoid causal language throughout the entire paper. Finally, the AURs of the final models were 0.66 for BP which corresponds with 32% explanatory power (0.66-05)/0.5 (AUR is 0.5 if the model has no explanatory value but 78% for ill health.”

Response: In our discussion section we attempt to bring in some novel thinking in the stagnant waters of back pain research. It is curious that various associations with both back problems and ill health behave similarly when subjected to a statistical challenge with a third variable (coping). This similarity of pattern plus the fact that back problems behaves like the “little brother” of ill health, is certainly worth pondering about. We do nothing more than introduce a line of thought, which the reader can accept or reject, and it certainly requires extensive research in a more direct manner. We are starting at the outer contours of this problem but the issue needs to be concretized and narrowed down in future work, if somebody finds the concept interesting and wishes to pursue this further. We agree that causal language such as “coping causes back problems” should not be in the text and we have, as in our comments to the previous reviewer, ensured that such terminology is definitely absent from the text.

Comment 3: “The sample size is impressingly large. The prize of this, however, may be large measurement error because it is difficult to ensure proper measurement for in many individuals. Measurement error often causes the largest bias. In this respect the following questions are essential: Who conducted the intellectual capacity tests? What are the psychometric properties of these tests? Which structured interview was used to assess coping? Were weight and height self-assessed by the recruits or really measured?”

Response: Obviously, in all studies there is always the problem of measurement error. However, this data collection is not performed as part of a study, but it is information collected in a systematic, professional and well organized manner, in a modern organization, where there is a strong belief in “the right person in the right position”. The Swedish Army has a very strong interest in giving its military training only to people who are motivated and mentally and physically fit. The quality of the data collection, the continuous validity checks and the “willingness to serve” – mentality is thoroughly described in the Methods section. In the Methods section, it is also clearly stated that the intellectual capacity test is performed with a computer program (touch screen method). Its psychometric properties are also listed in the text. In the text it is explained that the structured psychological interview, although calibration takes place and validation checks performed regularly, cannot be divulged. The problems with this are thoroughly aired in the text. We have now emphasized in the new text that weight and height were measured at the medical examination. Anybody with knowledge of the Swedish work culture would understand that anything else would be impossible!

Comment 4: “Back problems (and maybe also “ill health”) might be subject to large measurement error because the individuals have an interest to complain about health
problems that may truly not exist in order to be excluded from military service. How was this possibility addressed?”

Response: This has been explained in the text. Over-reporting of ill health is always a possibility, but because the demand was much smaller than the supply, it is no more necessary to do this in Sweden. It is perhaps more likely that some young men, who would like to obtain some of the skills provided in military training or who wish to join the army on a professional basis, would underreport illness. Also this possibility is discussed in the text. The fact that our data are normally distributed in relation to intellectual capacity and coping and that this has been the case for decades, is an argument for the overall correctness of data (as explained in the text).

Comment 5: “The statistical interactions you are testing have nothing to do with causal synergy. Despite of the usual problems with bias in observational studies, interactions in terms of modified odds ratios have no logical relation with causal interactions. The only measure for which there exists such a relation is the risk difference. Moreover, your factors are correlated and you therefore cannot separate correlation from co-action. Positive correlation can produce seemingly superadditive effects.”

Response: We do not believe and have not attempted to give the impressions that we are looking at causal synergy. We have tested for statistical interactions (none was found), hence we did not use interaction terms in our analyses. We looked at some bivariate (2x2) associations and observed what happened when they were challenged by other variables. This could have been done through stratification as well, but the data set was better analysed this way. Yes, of course, these variables are all correlated, which we have described in the text. These correlations show the danger associated with drawing conclusions based on small study samples which are likely to consist of subsamples of the general population. For example a study sample with relatively many short boys with high intellectual capacity (à la Woody Allen) would be completely different from a study sample with more tall guys with high intellectual capacity (à la Bill Clinton). Their coping capacity, for sure, would be very different.

Comment 6: “Pseudo R-square is known to have very bad mathematical properties. The results are strongly depending on the scaling. Only use AURs which have a straightforward interpretation as the probability that a true case (e.g. with back) problems has a higher model-based probability as a true non-case.”

Response: We are fully aware of the problems with the pseudo R-square but decided (reluctantly) to include it because there is a tradition in back pain research to do so. However, we feel much better about leaving it out and we use only the area under the ROC now.

Minor essential revisions

Comment 1: “Don’t say “outcome factors”. Factors are explanatory variables while outcomes are dependent variables.”

Response: Thank you. It has been changed.
Comment 2: “What a confounder is depends on the temporal relationship between the variables. For instance, age might be a confounder for the effect of coping on back problems but it is not likely that coping is a confounder for the effect of age. Don’t use this term.”

Response: The standard definition of a “confounder” in epidemiologic text books is the one we had in mind.

Another term could be “modifying factor”, although this, strictly speaking is not the same. A completely neutral term is “extraneous factor”, which means that one does not attempt to interpret whether there is confounding or modification. Although we mention the possibility of confounding, we have decided to use the term “confounder/effect modifier” on p.10, 2nd para.

Comment 3: “You need to be more explicit to define “ill health”. What medical conditions are sufficient to be excluded from military service? In Germany, for instance, you receive a degree from very good to bad health and depending on the need of the recruits the threshold to be included rised recently.”

Response: We believe that the explanation of “ill health” is fairly explicit, in terms of the needs of this study. We were not interested in looking for causes for various diseases but to look at the “rejects” for any medical reason. Even if we had wished to identify specific disease this would have been impossible because usually, once the medical practitioner found a medical problem of sufficient magnitude to interfere with military service, the search for other serious disease would stop. Therefore, there was often only one “reject diagnosis”, even if there would have been, medically speaking, more than one.

Comment 4: “Report confidence intervals of odds ratios. These contain important information about the power”.

Response: As explained in the text of the manuscript, confidence intervals in a study sample of more than 40,000 provide no useful information, because they are so SMALL.

Response to Eugene J Carragee

General

Comment: “This is an important paper and I enjoyed the chance to read the manuscript. There are number of issues that can be resolved with some revision. Since I believe this can be a landmark study it will need to be better supported regarding lost subjects, the screening that apparently some subjects received and others did not, and the proper statistical analysis/presentation. Finally I hope the authors can put together a table of the raw data including the breakdown from the numbers considered for enrolment, through the screening process and the final cohort that is accepted for military service with the rates of outcomes and dependent variables at each stage. I look forward to seeing the revised manuscript.”

Major compulsory revisions
Comment 1: “Abstract: The discussion of risk factors versus risk indicators does not seem germane to this study as it is a simple cross sectional study design and cannot discriminate between these issues.”

Response: We agree to having been too carried away here, and the offending text has been rigorously removed.

Comment 2: “Methods: The 2% of the study population rejected, some for severe illness, is an important loss. It is not clear why these were excluded and if the data still exits for this group. How many were rejected for military service at this first screening due to back problems.”

Response: In most epidemiologic studies there would be a target population (in our case all the young men who were summoned for the examination). A random selection would typically take place to capture a sufficiently large and representative sample from this background population. In our case, this was not necessary. After this random selection has taken place, only a proportion of potential study subjects will accept to participate. A response rate of 70% or less is not uncommon. In our case we captured 98% of the total POPULATION, and the few subjects (2%) who were excluded were indeed severely ill (institutionalized) or in prison. This fact might not have noted by the reviewer in the text, but if the reviewer reads the methods section again, he will see, that this is very clearly stated. Furthermore, there was no “first screening” in which back problems could be detected. As these 2% never got there, there is no data on them. In other words, our study sample consisted of 98% of the total underlying target population. Two percent of these do not represent an “important loss”. Hopefully, this misunderstanding is now cleared.

Comment 3: “It is hard to tell the impact of the progressive screening (in which not all the recruits get the same battery of tests). To be clear did all the enrolled subjects take all of the tests?”

Response: This has been better explained in the text. No, not all subjects took all of the tests but the large majority did. Those who did not take all the tests were those who it was obvious would be rejected. All the others took all the tests. The number of participants in the screening procedure are listed in Table 2. The numbers kept for the various analyses are now shown in Tables 3 and 4.

Comment 4: “It would be helpful to better understand the threshold for making diagnoses on the part of the screening doctors. Were these apparently equal at all centers? What were the instructions given to the screening doctors in making diagnoses? This reviewer has performed thousands of screening examinations of military personnel for duty, the purpose is sometimes to make a comprehensive list of all medical issues in the soldiers history. In other cases the purpose is to list only those which in the opinion of the military doctor is likely to impact performance of duty. In that case the “severity” of the disease is built into the screening.”

Response: The Swedish Army employs military practitioners, whose career is in the army. They are instructed in how to deal with these examination procedures and, as explained in the text, any deviant statistics is checked at Central Office, and discussed with the person responsible for the “different” reports (i.e. “calibration” took place). As explained in the text, the validity of data is considered high, because the system is highly professional and systematic. This systematic approach is a very strong national trait in Sweden, I might add.
Obviously, the contact between the young man and the medical practitioner is in relation to military service. However, the entire medical history (as explained in the text) is taken into consideration. The grading system used, is used in the light of the ability of the young man to perform military service. It is correct, that the cutpoint for “illness” is related to military service, as explained in the text.

Comment 5: “It would be helpful to understand the ability of recruits to avoid service by feigning ill-health or exaggerating mental or copying deficiency. A young man hoping to avoid service may purposefully present as unable to cope with most past health issues…”

Response: It would always be possible to feign illness in certain areas. However, this is not really necessary, as the demand is much smaller than the supply. The Swedish Army is only interested in “good” subjects, both in relation to physical health and mental status/attitude. This is explained in the text. Underreporting of disease is also a possibility.

Comment 6: “Is it known how many subjects determined to have ill-health had back pain disorders?”

Response: The presence of a disorder is always recorded. However, the severe cases, those we classified as “ill health” because they were considered unsuitable for medical service, can only be counted as number of persons. The reason for this is that, often, all reasons for ill health in one individual were not noted down. After the first obvious reason, no further probing was often necessary. Therefore, we can only say how many of the people with back problems who were also diagnosed with “ill health” but we cannot see how many of the persons with back problems had back problems severe enough to be classified as “ill health”. Please, see the 3rd para in the first section of Results, where this is explained.

Comment 7: “Results: The findings of only 11% of subjects with back problems (as opposed to any history of back pain which may be expected to be higher) seems to indicate that a higher degree of severity was built into the screening process than simply any back problems with not expected functional impact.”

Response: Yes, this is correct and it is a good point to include in the text, which we have done in the first section of the Results.

Comment 8: “Some to the ICF 10 diagnoses used in the study should perhaps be reconsidered. It seems the authors or really interested in the risk factors for back pain / troubles association is subjects without significant structural disease. It is puzzling to see concrete structural diagnoses included in the diagnoses for analysis (scoliosis, Scheuermann’s disease, etc). I believe the face facility of the study would be improved if these diagnoses were excluded. By far most scoliosis and kyphosis patients are without serious symptoms and are simply examination observations.”

Response: Clinicians like to provide a diagnosis, and so they did. When a young man complained about back problems, the clinician obviously selected the item in the list of conditions that he thought best could explain the problem, whether this was correct or not. However, this had no consequence for the young man, unless his functional ability appeared to be sufficiently severely
compromised to give him the grade of “reject”. Also this is explained in the text (Methods, Outcome variables, 1st bullet).

Comment 9: “Page 9, last paragraph -- Is it correct that approximately 35% of the recruit population was rejected for military service based on health reasons. This seems to be a very high number. This paragraph seems somewhat hard to understand. If the occurrence of these outcomes is really this high, it may be that the reporting of Odds Ratios is inappropriate. As pointed out in an article in “Statistics in medicine “ by Helen Kraemer (2004, “Reconsidering the odds ratio as a measure of 2 x 2 association in the population”), the odds ratio is a baffling figure for statisticians and nonstatisticians alike. It is easily misinterpreted, unless it is used as an approximation for the relative risk of a factor. This latter interpretation is allowed only if the prevalence of the outcome under consideration is rare. As about 30-40% of the subjects have the outcomes in question, the OR may be inappropriate statistic in this setting.”

Response: In the old manuscript Result section, under “Study subjects”, 2nd para, first sentence, we stated “In all 28% were classified as having “Ill health” (IH), i.e. to be unsuitable for military service”. However, the army does not include all those who “pass”, because they do not need them. We can only hope that this statement makes sense on a second reading.

Also in this paragraph, it is clearly stated that 11% had “back problems”. We therefore do not understand the argument that about 30-40% of the subjects had the outcomes in question. There were 28% and 11% who had the outcomes.

The odds ratio is a simple statistic which is used widely in the back research domaine. It is the correct measure to use in cross-sectional studies and is easily comprehensible. We are using the odds ratio because there is no other alternative really, when you want to do a multi variable analysis. The odds ratio can be used to compare results with those obtained in other studies, too, which is an additional plus. With help of the odds ratio we could show that the variable “coping” has a nasty effect on the other “predictors”, but the other “predictors” can’t do much about rocking the “coping” variable.

Comment 10: “The authors state in the same paragraph that there were definite associations between all five predictor variables, then the authors state at the bottom of the page there were no significant interactions between any of the predictor variables. Please clarify, this seems hard to follow.”

Response: We talk about “associations” and “interactions”, as dealt with in the epidemiologic literature. We can only refer to any standard text book on this issue, in case there is confusion as to their definitions. If not, we do not understand the reviewer’s concern.

Comment on quality of written English: “Not suitable for publication unless extensively edited.”

Response: We shall of course be happy to comply with this, providing that specific deficiencies are listed.