Author's response to reviews

Title: Are self-report of disability pension and long-term sickness absence accurate? Comparisons of self-reported interview data with national register data in a Swedish twin cohort

Authors:

Pia Svedberg (pia.svedberg@ki.se)
Annina Ropponen (annina.ropponen@uef.fi)
Paul Lichtenstein (paul.lichtenstein@ki.se)
Kristina Alexanderson (kristina.alexanderson@ki.se)

Version: 2 Date: 7 December 2010

Author's response to reviews: see over
December 7, 2010

Editorial Office
BMC Public Health regarding MS: 1785698121447743

Dear Professor, Editor

Enclosed is a revised version of the manuscript entitled “Are self-report of disability pension and long-term sickness absence accurate? Comparisons of self-reported interview data with national register data in a Swedish twin cohort” (MS: 1785698121447743) by Pia Svedberg, Annina Ropponen, Paul Lichtenstein, and Kristina Alexanderson. We thank you for this opportunity to revise our manuscript and respond to the comments given by the reviewers, which we have used in the revision.

Below is a response to each of the reviewers’ concerns point-by-point.

Changes made in the manuscript text are highlighted/coloured in yellow.

The name of the ethics committee is the Regional Ethical Review Board of Stockholm, Sweden and the number of the decision on the application for this project is (2007/524-31). This information is given in the manuscript, at page 8.

Yours sincerely,

Pia Svedberg
Corresponding author, on behalf of all co-authors

Division of Insurance Medicine
Department of Clinical Neuroscience
Karolinska Institutet
SE-171 77 Stockholm
Sweden
E-mail: pia.svedberg@ki.se
Reviewer 1.
Reviewer's report:

Major Compulsory Revisions
none

Minor Essential Revisions
1. Although addressed in the introduction, it is not entirely clear why MZ/DZ status would affect reporting of sickleave. It seems more a convenience sample that allows for a validity study of much larger size than ever before and, thus, the possibility to analyse important determinants of reporting validity, such as education, age and sex. Unfortunately, education and age were not investigated. This would add to our understanding.

As pointed out by the reviewer, this study is based on a larger sample than previous results of other studies of validity of self-reported DP and LTSA compared to administrative/register data. To the best of our knowledge, there were no previous published reports of the validity of self-reported DP at the time when this study was planned and designed. Nor are there any such studies on the validity of self-reported data of DP or LTSA using a genetically informative sample. Our additional aim of this study was therefore to evaluate the influence of familial factors on self-reporting behaviour given our unique study setting. Familial effects of reporting behaviour is largely unknown for most outcomes and exposures, but is assumed not to exist. However, genetic effects are obviously present for many behaviours and traits as reported in numerous publications in the field of behaviour genetics. Genetic influences are indicated if the MZ twin pairs more often are concordant for reporting behaviour compared to DZ twin pairs, while equal concordances between zygosity groups indicate that it would be possible to exclude genetic influences on response style. We have now revised the manuscript and made this rational more clear in the introduction section, pages 4 and 5.

Also, as suggested, results of post hoc kappa analysis of education and age as determinants of reporting validity for DP and LTSA has been added in the results section, pages 8-9. We found no difference in Kappa values between two age groups (cut off 55 years) or educational level for LTSA or across age groups for DP. Estimates differed slightly between the two groups of education. A higher $k$ value was found for the group with less than 13 years in school (0.77), than for the group with $>13$ years in school (i.e. college and university) (0.68). However, both estimates are within the range of substantial agreement. Results are also mentioned in the discussion section, page 12, in relation to the findings of the previous study conducted in Norway.

2. The introduction refers to validity studies on sickness absence, with much short periods of absence than investigated in the current study. This should be mentioned.
Thank you for pointing this out. Previous validity studies have evaluated both short- and long-term sickness absence, but as pointed out, mostly shorter periods of sick-leave. We have rewritten this to make it clearer in the introduction and discussion sections.

3. It is not clear what time frames were used in the comparison of the questionnaire and register. Was date of filling out the questionnaire linked to the register and information retrieved for that date (or prior month etc)

   The date of the telephone interview for each individual was linked to the DP/LTSA register and information about DP and LTSA was retrieved for that specific date, something we considered a strength of the present study. Hence, this is now also mentioned in the discussion section. This procedure is described in the method section, page 6, paragraphs 1 and 2.

4. I would prefer a clear statement in the text with reported and registered prevalence of DP and LTSA

   As asked for, clear statements of the prevalence of self-reported and register-based DP and LTSA are now reported in the text. This was also pointed out by reviewers 3 and 4. For LTSA the prevalence of self-report was 3.3% while prevalence based on the official records was 6.5%. For DP corresponding prevalence were 8.9% (self-report) vs. 11.2% (records) (results section, pages 8-9).

5. The discussion should make clear that the LTSA was asked as point prevalence, whereas the debate on validity of self-reports is also on the acceptable period of recall. It should be noted that the current study has no recall bias and, thus, may have a better validity than previous reports who used 3 to 6 months recall periods.

   Thank you for this comment. We have now taken this into consideration in the discussion section, pages 11 and 13.

Discretionary Revisions

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.
Reviewer 2.
Reviewer's report:
This paper tests the degree of agreement between self-report, disability pension and long-term sickness absence in a large twin survey (SALT) in Sweden compared to the official Swedish registers. The paper is very clearly written and easy to follow, but there are two major concerns and a couple of minor:

**Major concerns**

1. The results are only applicable to the Swedish registers. The whole manuscript is written as if it is a more or less universal finding that self-report of DP and LTSA would also apply to other settings. This might be true for other Nordic countries, but the manuscript should include some discussion of this. In any case, the results are unlikely to be of interest for researchers outside the Nordic countries where such resources rarely are available.

   In most countries this type of high quality register data on DP and LTSA is not available. In studies of DP and sickness absence (SA) one therefore usually have to depend on self reports on DP, LTSA, and SA. This is true for many studies also in the Nordic countries, where permission to link to register data has not been obtained. Therefore, many researchers have found it important to study the self-reported data of SA. In this study, we also do this for DP and LTSA, using a much larger study sample than has been done in previous studies, and also using a twin setting. This means that our study, on the contrary to what is suggested by the reviewer, probably will be much used also in other countries, where this type of register information is more difficult to obtain, e.g. in the USA.

   Most societies have some sort of sickness compensation/insurance system and not only the Nordic countries even though the social insurance systems in the Nordic countries are more similar to each other than to systems in for example England and the Netherlands. From a research perspective it is even more important to find out whether self-reports are valid or not if register (national, administrative, or employer) data are not at hand resulting in that one have to rely on self-reported data in the field of sickness absence research. Hence, we consider the possible limitation is in the field of interest (sickness absence research or twin studies) rather than about similar register data being present or not in a specific country.

2. It is not clear why this is a separate paper; the authors are undoubtedly in a process of estimating the relative importance of genetic and environmental factors on disability pension and long-term sickness absence, and they have in this connection access to both self-reported data and register data. So, it seems that the whole paper would be better suited for a few paragraphs in a paper on these topics, where the results of the twin analyses were reported, and where the register data and self-reported data were used showing that the results were robust towards which source of information is used.

   Analyses of the relative importance of genetic and environmental influences on DP and LTSA were beyond the scope of the present study. The aim was to replicate earlier findings of validity of DP and LTSA, using a
larger sample, high quality population-based national register data, with the ability of comparing the exact date of LTSA/DP self-report with register data, something not done before. An additional aim was to study potential influence of familial factors on reporting behaviour, an aspect not previously investigated in the context of sickness absence, nor much reported for other outcomes or exposures. In twin models when estimating the relative importance of genetic and environmental factors for self-reported traits or diseases we assume no differential effects of reporting behaviour between MZ and DZ twins i.e. we expect robust results. Still there are few such validity results presented in the literature. Mostly, only the presence of the specific disease or disorder of interest has been validated, while evaluation of participants reporting absence of disease is not validated. We believe that reports of genetic liability to DP and LTSA are better suited and communicated in separate manuscripts.

Research on DP and SA is an emerging research area, in much need of methodological and theoretical development. This study is within the realm of that field.

Minor comments:
3. It is stated on page 7 that it is assumed that national register data are correct, but it would be good with some backup of this statement.

Thank you for pointing this out. The national registers of the Social Insurance Agency are not research registers, but administrative registers kept by an authority to have good information about benefits paid out after legal decisions about that. No benefits can be paid out without being registered as such. People expecting benefit who have not gotten them would react. That it is not a research register but an administrative register, based on monetary transactions, is one reason for researchers in general to regard them as being of high quality. Moreover, in Sweden, as opposed to some other countries, there is only one public social insurance system covering all people above 16 years of age – that means that all who have had that type benefit are registered within the same agency.

4. Table 2 is very hard to read and interpret. It gives basically just the raw numbers and no summary statistics. Also, the few key results reported in the text actually point in the direction that MZ twins are more concordant for misreporting than DZ, but there is no formal statistical testing of this.

Summary and Chi-square test statistics are given in the text (result section, pages 8-9), as noted frequencies are given in table 2. We agree that results point in the direction that MZ twins are more concordant than DZ for misreporting, however, this observation was not statistically significant. We
find it important to give also the raw numbers, in an research area that is developing.

**Level of interest:** An article of limited interest

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Reviewer 3.**

**Reviewer's report:**
This is a well-written and well conducted study, investigating methodological issues concerning the use of self-reported data contra register data. The study is of great interest and the findings are useful for future research on sickness absence and disability pension. However, I have some comments and questions that might be useful to consider:

**Minor Essential Revisions**
1. As I understand it, the paper describes two different aims of the study; 1. To consider the agreement between self-reported data on long-term sickness absence (LTSA) and disability pension (DP) and 2. To investigate the genetic/environmental influence on reporting behavior on LTSA and DP.

First, I do not feel that the title reflects or prepare the reader for the second aim.

The primary aim was to evaluate agreement between self-reported and register data on DP and LTSA, hence replicate earlier findings, in a larger cohort than previously and with the possibility to compare actual date of self-report and register data. An additional aim was to evaluate familial effects on reporting behaviour. By including "... twin cohort" in the title we believe that we have prepared the reader by stating that we use a genetically informative sample. We believe that the title would be difficult to comprehend if more details are given.

Secondly, I have some objections with the second aim, at least as it is currently presented in the paper. I understand and appreciate that the second aim is interesting in a broader context, in other words - to investigate nature-nurture influences of reporting behavior in itself. However, I am not sure that it is of equal interest to focus (to this extent) on the finding that we can trust self-reported data on LTSA and DP equally in monozygote and heterozygote twin pairs. And further, the finding that genes do not seem to influence reporting behavior on these specific issues. Perhaps these findings would be more justified if presented in a broader context of genetic/environmental influence on reporting patterns in a separate paper?

Thank you for this comment. We realize that not all readers within the field of sickness absence research might be interested in the specific issue.
of familial influences of reporting behaviour. However, we felt that it would be
difficult to divide this manuscript in two, but instead would read better when
results are taken together. Moreover, we were interested in the possibility of
investigating influences on reporting behaviour, something that is not
possible to investigate in other than twin- or family settings. We think that this
adds new insights into the issues of validity of self-reported sickness
absence data as compared to register information. However, since the
primary focus of this study was on validity and secondly on familial effects of
self-reports, we have now revised the introduction somewhat to clarify that
the latter was an additional aim.

2. You present a hypothesis for aim 2, but not for aim 1. Either you should
include hypotheses for both aims, or not include hypotheses all together. If you choose to
postulate hypothesis for the aims, the empirical rational for the
hypotheses should be given in the introduction.

We have now revised the manuscript accordingly and excluded the
hypothesis for the second aim, introduction section (page 5).

3. I had a really hard time understanding how many participants that were actually included
in the ‘agreement’ analyses of LTSA and DP. The flow-chart was of great help, however
when reading the ‘participants’ section in the ‘Method’ chapter, a lot of space is used to
describe the total SALT population, and the reader could easily get the impression that you
have used n=12129 in (all) the analyses. It might be better to reduce the overall information
on the total study population in SALT, and include more information on the number of
participants used in the analyses in this specific study. It would also help to include n in
table 1.

Thank you for pointing this out. The complete sample of 31,122
individuals was included in the agreement analyses. In order to describe the
“source population” and the “study sample” more clearly, these two
subheadings has been added and the description of the two somewhat
rephrased (method section, pages 5-6). Also, n has been included in Table
1.

4. It would be interesting to have more descriptive data of the sample that was actually
included, for instance age, education and reason (diagnoses) for LTSA/DP. If you have such
data, it would also be interesting to investigate if the agreement between self-reported data
and register data differs between age groups, educational groups and diagnostic groups.

We conducted post hoc analyses of age and educational groups as also
asked for by reviewer 1 (see response above). Kappa results have been
added in the result section pages 8-9. Concerning diagnoses we agree that
this would be interesting to evaluate, however, diagnoses for LTSA are not
available in the national registry, also, the diagnoses causing the LTSA or DP were not asked for in the SALT survey.

5. The discussed weakness of the study by not quantifying ‘long-term’ for the respondents on page 11, should be moved to the paragraph of methodological considerations (limitations and strengths) in the end of the manuscript.

   This weakness is now spelled out in the section of limitations and strengths, page 13, second paragraph.

6. Make sure there is a consistent use of the term ‘sickness absence’ (if that is what you want to use). Now there is a mixture of ‘sickness absence’ and ‘sick leave’ both in the text and in the table texts.

   Done. Sick leave has been changed to sickness absence throughout the manuscript to be consistent.

7. At least in my view, it is more logical to present LTSA before DP, both in the title and in the results.

   For this manuscript we have chosen to present DP before LTSA since few results on validity of DP has been presented so far. In fact, at the planning stage of this study and when data were analysed, no such study was published, to the best of our knowledge. Hence, accuracy of self-reported DP was our main focus.

8. In Figure 1 the ‘DP-line’ (DP in SALT, Not DP in SALT or MiDAS, and DP in MiDAS at time of…) does not add up to n=31122 but to n=31121. What happened to that one person?

   Thank you for pointing this out. This typo is now corrected and numbers should add up to 31,122 for both LTSA and DP. "DP in MiDAS at time of interview, but not self-reported in SALT interview” should read 1047 and not 1046.

   Level of interest: An article whose findings are important to those with closely related research interests

   Quality of written English: Acceptable

   Statistical review: No, the manuscript does not need to be seen by a statistician.
Reviewer 4.

Few studies have explored the validity of self-reported sick leave and disability pension benefits. This calls for more studies in the field and make the main aim of this study relevant and important. The paper is well written and the communication is clear, however I have some comments.

**General comment:**

It is difficult to see the uniqueness of the data being from a twin cohort. The focus on possible differences in reporting behavior between twin zygosity groups seemed a little bit odd.

This issue was also raised by reviewers 1 and 3. Please see our response given above and changes in the text on pages 4-5.

Even though I realize that it is not necessary for the limited aim of this study, I would prefer some more background information beyond Figure 1. It would be interesting and useful with more specific information about the actual number of individuals on DP and LTSA (self reported and registered) reported separately for MZ and DZ pairs of twins, for instance presented in contingency tables. I believe the manuscript would communicate better with this information available.

The study sample (n=31,122) is described in figure 1 while the number of pairs in each zygosity group is described in the text (page 6). Of the complete number of individuals the sample consist of 12,237 twin pairs (3247 MZ pairs, 4421 DZ pairs, 4461 opposite sexed DZ pairs, and 108 pairs with unknown zygosity). Those pairs of unknown zygosity were not included in the pair analyses. Additionally 6648 twin individuals, where information is missing for the co-twin, i.e. the co-twin did not participate in the interview study (SALT) is included in the total sample. Since agreements between self-reported and register information of DP and LTSA are the focus of this manuscript we have chosen not to report actual number of concordant or discordant pairs for groups (self-reported, register) of LTSA and DP separately, instead focus on concordance and discordance in reporting behavior. The latter numbers reported in Table 2. We believe that the suggested additional table is better placed in another manuscript focusing on genetic liability to LTSA/DP.

**Minor comments:**

Background

§ 3 “Most studies have compared self-reported number of sick-leave days during a specific time period (weekly, …) compared to employer administrative register data [7-9, 11, 16],…” Reference no. 11 has not compared to employer register data. Content of the first part of the sentence is correct though and is believed to be the most important message in this sentence. Suggest therefore to write: “(weekly, …) compared to register data [7-9, 11, 16], …”
Thank you for pointing this out. Previous studies have sometimes been able to compare self-reports to employer (or) administrative register data, we have therefore changed the sentence on page 3 to read "employer OR administrative register data".

Results
§ 3 under “Twin pair correspondence”:
“Chi-square test yield no statistically significant differences between zygosity groups for DP or LTSA (p>0.05), …” Reporting the exact p-value would be preferable.

Done. P-value for DP is (p=0.31) and LTSA (p=0.64), reported on page 9, results section.

And in the same paragraph:
“The number of concordant female twin pairs in which both members of the pair reported incorrectly being on LTSA as compared to the male pairs were 8/2 in MZ twins and 4/1 in DZ twins…” This whole sentence is unclear and may be due to the lack of more background information.

In order to clarify that these are the total number of pairs in which both members reported incorrectly the sentence has been somewhat rephrased (page 9, last paragraph, result section). The total number of pair reports - concordant or discordant - is given in Table 2.

Discussion
§ 4, Last sentence: “Part-time DP benefits can also be granted in Sweden which might explain the underreporting of DP in this study and suggests that in countries with only full-time DP benefits, estimates of agreement might be as high or even higher, such as in Norwegian reports [11,18].”
This statement is based on wrong information. It is possible to receive part-time DP benefits also in Norway (50 percent or more).

Thank you for pointing this out. The sentence has been changed accordingly to “Part-time DP benefits can also be granted in Sweden which might explain the underreporting of DP in this study and suggests that in countries with only full-time DP benefits, estimates of agreement might be higher” (Page 12).

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.
In conclusion, we appreciate much the comments given by the reviewers and have had good help from them in revising our manuscript.

The final version of the submitted manuscript has been seen and approved by all authors, who have all contributed substantially.