Reviewer's report

Title: Gastrointestinal complaints and certified long-term sickness absence in the HUSK study.

Version: 1 Date: 8 February 2011

Reviewer: Jennifer Schurman

Reviewer's report:

This article, “Gastrointestinal complaints and certified long-term sickness absence in the HUSK study,” examines the relationship of GI and other somatic and mental health complaints to absences from work lasting 2 weeks or more. The authors should be commended for undertaking a population-based longitudinal study to examine these relationships. However, as currently written, the contribution of this study to the general knowledge about the social and/or economic impact of GI complaints in adulthood, and/or the targets for potential intervention if a meaningful impact exists, is unclear. There are substantial problems with this manuscript as written that negatively impact its potential contribution to the field. The most significant problems are outlined below:

Major Compulsory Revisions

General

1. Greater precision and clarity is needed in the writing throughout this manuscript. For example, in the first sentence of the first paragraph of the Background, the authors state that GI complaints “are very common in the general population, with more than half reporting such complaints within three months.” The study cited in support of this statement is a US sample, which should be clarified, but leaves open the question of relevance to the HUSK sample (not US based). In addition, do the authors mean any level of GI complaints is “very common” or a clinically significant complaints (i.e., above a certain threshold)? Do the authors mean within three months of a certain event or within any given three month period? This is only one example of many in which the meaning of a specific sentence is unclear, the cumulative effect of which is to leave the reader unclear as to the logical flow of the authors’ argument, as well as the relevance to the current study.

Background

2. In the third paragraph of the Background, the authors state that “there is now a growing concern over increasing proportions of the work force absent from work for longer periods of time…..” However, they cite a European study here. Given that the healthcare system varies substantially (including availability and accessibility of paid time off for sick leave), the authors need to include a rationale in the Background for examining this relationship in a Norwegian sample and, in the Discussion, include implications for generalizability and future directions for study.
3. As discussed in the third paragraph of the Background, why might cross-sectional self-report of sick leave (as previously reported in the literature) be a problem in terms of understanding the relationship between GI complaints and extended sick leave? This is an important point in developing the rationale of the current study and is not adequately addressed.

4. Why is extended sick leave more important to examine than possible frequent short-term sick days? Was this selection due to convenience (availability of data) or a theoretical rationale?

5. The principal study questions need to be clarified using more precise language. Specifically, the authors appear interested in three things: 1) Does a high level of total GI complaints (number X frequency) at baseline predict one or more prolonged “sick leaves” (greater than 14 days) in the 4-year period immediately following? 2) If this predictive relationship exists, is it explained (in part or full) by a single GI complaint (e.g., nausea, stomach pain) measured at baseline, rather than the total number and frequency of all possible complaints? 3) If this predictive relationship exists, is it explained (in part or full) by associated comorbidities (e.g., anxiety, depression, other somatic complaints) measured at baseline? The language of these questions, as outlined above, needs to better operationalize the variables being examined, as well as presenting the relationships in clear, testable language. If possible, these questions should be rewritten as specific aims and/or hypotheses based on existing data.

6. All demographic variables introduced as possible moderating factors (e.g., sex, age, BMI, education level) should be discussed in the Background in terms of rationale for inclusion and hypotheses presented (if possible) regarding their potential impact on the relationship between GI complaints and extended sick leave.

Method

7. What was the rationale for tallying total number X frequency of complaints in determining a total score for GI complaints and, further, for dichotomizing this continuous scale using a cut off at the 80th percentile? Is this the 80th percentile for this group? Is this the 80th percentile based on a norm group? Is this methodology validated or was it created for this study by the authors for a reason that can be specified? If any reliability and/or validity information can be provided for this measure of GI complaints, that would be important to interpretation of results.

8. Was the intervening period of time for the information on sick leave from 1999-2003 for all participants? If so, did this leave some participants with 6 or 7 years of data versus 4 (for the participants collected at the end of the baseline period)? If so, did the authors account for this difference in time period in some fashion? Does this explain the older end of the age range having more LTSA than the younger end (i.e., more years in the study to accrue sick time)?

9. What was the rationale for serially dichotomizing extended sick leave into present/absent, 0/1(+) period of > 55 consecutive days, and 0/1(+) period of >100 consecutive days? Why did the authors decide on this approach versus counting total number of sick days as a continuous variable, or some other
approach?

10. Were any diagnoses (physical or mental health) used to exclude an extended sick leave directly related to that diagnosis given the focus of this study on relatively “unexplained” somatic or GI symptoms?

11. In the first sentence of the statistical analysis paragraph, the authors state that descriptive statistics were reported using specific tests. Descriptives would not include tests, but rather means, SDs, and/or frequencies, etc. If these were analyses designed to look at group differences on a primary variable (i.e., two groups divided into above and below the 80th percentile for GI complaints) at baseline on secondary variables, then this should be identified as such and the relevant table presenting the data should be indicated in the text. A rationale for examining baseline differences should be provided, if this is the case, and baseline differences should be statistically controlled for in the subsequent analyses to reduce noise in examining the primary relationship of interest (i.e., GI complaints on LTSA).

12. Why were odds ratios determined to be the best approach for presentation of the analyses? Why dichotomize and then use logistic regression versus hierarchical multiple regression? Why use separate models to make adjustments versus using a more stepwise approach? Please provide rationale for the use of each statistical test, in addition to rationale for decisions regarding data dichotomization.

Results

13. The issue of statistical versus statistical significance seems very relevant in the current study. With such a large sample, the actual group differences appear quite small and, yet, are statistically significant indicating a possible problem with overpowering the study. It may be helpful to include either effect size measures to provide greater interpretive information about the actual magnitude of the effects found and/or to discuss clinical meaningfulness in terms of the actual means for each group. For example, the baseline differences between groups (cut for 80th percentile of GI complaints) on BMI yields an extremely small effect size despite the significant p value. Taking effect size into consideration in selecting variables for entry into the model may help to streamline the analyses, as well as aid in more accurate interpretation of results. Alternately, the authors may wish to split the group randomly, using one half of the participants to develop the model and the remaining half to confirm the model to decrease overfitting to the data and to better improve generalizability.

14. Given the large amount of participants and data, the authors also may wish to consider using a more sophisticated statistical approach, such as structural equation modeling, that best capitalizes on the study’s power while preventing overfitting of the data. This would require the authors to make their hypotheses about relationships among variables more clear, but would allow for adding or pruning paths in an empiric fashion, as well. This approach also would allow for more information about possible mediation or moderation associated with specific variables (e.g., genders impact on the relationship between GI complaints and LTSA).
Discussion
15. The point made about there being a general tendency of symptom reporting driving the association between GI complaints and extended sick leave is an interesting one that may be testable using a structural equation model or hierarchical multiple regression and looking at several variables simultaneously for contribution to LTSA as an outcome. As alluded to above, this manuscript may be better served by a focus on larger scale model-building and testing, rather than on examining individual relationships in a somewhat separate manner.

16. The Discussion should include some interpretation of study results from the point of view of social and economic cost, given the initial rationale provided about why the study questions are important. In addition, it would be helpful to include some discussion of potential clinical targets for intervention to ameliorate the identified problem, if available, or future directions to clarify what might be appropriate targets for future intervention.

Minor Essential Revisions
General
17. The authors should verify that AMA style is used throughout the manuscript, including for inline citations, etc.

Abstract
18. Greater specificity of language regarding study aims and hypotheses would be helpful in the Objective.

19. Greater clarity about what relationships were predictive (e.g., with LTSA) and what relationships were concurrent (e.g., with anxiety and depression) both in the Abstract and results/tables would be helpful.

20. In the Conclusions, the authors state that “associations are partially explained by symptoms of anxiety and depression, and to a greater extent the presence of other somatic symptoms;” however, this language appears inaccurate, as the statistical tests conducted did not test this type of meditational model.

Background
21. Greater specificity is warranted in language related to organicity. Recent literature increasingly has identified biological (or organic) contributors to the types of GI complaints asked about in the current study (historically considered “not to have a medical explanation”, using the words of the current authors). While an “organic disease” or “a clear organic cause” may not be found to fully explain the GI complaints endorsed, the authors should be cautious in their language to reflect current thinking in this area.

Method
22. The participation rate given of 63% is higher than the actual percentage included in the final analysis. The actual final participation rate should be given after accounting for all of those who did not participate, including those who did not complete the questionnaires or clinical exams, deletions due to preceding
disability pensions, lack of paid employment, and casewise deletion due to missing variables. Rationale for casewise deletion should be presented here, rather than in the discussion. In addition, these decisions (and the relatively large proportion of the population who were not included for some reason), along with the selection of a Norwegian population and the limited age range involved, should be discussed in terms of generalizability to other groups.

23. Was this a secondary analysis of existing data or an actual prospective longitudinal study? Clarification on this point would be helpful, as well as citation to the original HUSK study manuscript which would provide further details on participant selection and data collection methodology, if available.

24. The concept of absences less than 14 days in length providing a “wash out” period, or how the “wash out” period would be helpful, was not clear to this reader.

25. Organization of the Method could be enhanced by labeling variables as predictor variables and target or outcome variables and reordering them to appear together as a block (e.g., GI complaints, somatic symptoms, physical conditions, psychological conditions, and demographic variables before presenting extended sick leave, the target).

Results

26. Table 1 should be broken down into two tables to better present baseline differences related to LTSA. As currently organized, this half of the table is difficult to follow and does not provide mean/SD information for those not experiencing an LTSA during the study period.

27. Please explain more fully the sentence in the second to last paragraph of the results which states “Those with more GI complaints more often had an ICPC-diagnosis from chapter “D-Digestive” warranting the sickness absence”.

28. All tables should be referenced in appropriate areas of the text and all statistical information mentioned, even in the Discussion, should be present in a table or within the text itself (e.g., information related to “joint pain and stiffness”).

29. Table 4 is difficult to read, as it is the reverse of an earlier table, with presentation of the models across the top, rather than the side. Some degree of consistency in presentation would be helpful to the reader.

Discussion

30. The authors state that there “was no attrition during follow up”, yet this is a bit misleading in that numerous potential participants were dropped from analyses due to incomplete data, or other issues, limiting the pool of participants on the front end of the study. This should be clarified in the strengths and limitations more clearly. In addition, it is not clear whether any participants could have moved out of the healthcare system being studied and, thus, could have looked like no LTSA when, in fact, they were no longer in the pool.

31. The Discussion may benefit from reorganization, with the initial focus on addressing the study’s aims and hypotheses (clarified as discussed in the Background), moving on to interpretation and clinical implications, followed by strengths and weaknesses/future directions. As it currently reads, it is difficult to
follow the authors' logical flow and identify the main findings and implications of this study. Further specifics cannot be given at this time, as the content of the Discussion may change considerably if the Background becomes more focused and the statistic plan is altered.

Discretionary Revisions

General
32. The authors may wish to consider changing the title to better reflect main study findings.

Abstract
33. Please consider eliminating the use of acronyms in the Abstract, unless they are defined on first appearance in the Abstract.

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

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.

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