PhD evaluation — Alkayyali

SDC

4 February 2015
Version 2

Compiled Wednesday 4th February, 2015, 10:17
from: `/home/bendix/sdc/eval/PhD/alkayyali/eval1.tex`

Bendix Carstensen Steno Diabetes Center, Gentofte, Denmark
& Department of Biostatistics, University of Copenhagen
bxc@steno.dk
http://BendixCarstensen.com

Contents

1 Evaluation of Alkayyali's papers

The evaluation of the thesis is based on 4 papers and appears in section 1.1; a more detailed background for the evaluation can be found in section 1.2. An erroneous reference to the title of the second paper has been corrected in section 1.1 relative to the conclusion given to University of Lund.

Comments concerning the revisions I received on 2nd February is added as section 2, followed by a final conclusion in section 2.1.

1.1 Statement concerning PhD-thesis

The focus of the 4 papers presented (one published, 3 manuscripts) is assessment of the effect of genetic variants on mortality and complications in diabetes patients, and has therefore required accessing genotypical as well as phenotypical information from several studies. The genotypical information acquisition seems well described (however I have very limited knowledge in this field), whereas the phenotypical (clinical) information is less consistently reported between studies.

In general the epidemiological analyses appear weak and not well explained in the papers, in particular the second paper ("*KCNJ11*/E23K polymorphism increases ...") has flaws in the explanation of methods as well as inconsistent event counts (12% extra deaths appear).

Moreover the fourth paper has severe methodological omissions that possibly invalidates the conclusions even though this question could have been resolved quite easily (with unknown result, though). Also it seems odd to make the last two papers (on the DOLCHE study) separate, the famine analysis could easily have been included in the first paper — it is basically a single extra variable (although pointing at different loci).

In conclusion I find it difficult to recommend to the University of Lund to pass a PhD with two major flawed manuscripts out of 4.

1.2 Specific comments

There are 4 papers, where I have the following comments, on which I have built my statement above:

Common variant in the *HMGA2* gene ...
: This is an analysis of genetic risk factors for nephropathy. The studies entering the analyses are not uniformly described which makes it difficult to evaluate the internal validity of the study. Although there seem to be follow-up data available, an over-simplified logistic regression of the end status of participants' nephropathy status have been used. The reason for this is not entirely clear.

KCNJ11/E23K polymorphism increases ...
: This is an analysis of (in particular) CVD mortality in 3+1+6 cohort studies focusing on genes identified through analyses of the first 3 studies. The studies are not described homogeneously, and it is not indicated when follow-up started for any of them. Despite this it is in the methods section stated that data were treated as left truncated and right censored. This seems to indicate that time since baseline is not the underlying timescale (in which case there would be no left truncation), but it is not stated which is the underlying timescale used.

In table 1 is given 276+200=476 CVD deaths in the two first studies, which is also found in table 1 under analysis with RS5219 as covariate, but for the other three SNPs there are suddenly *more* than 476 CVD deaths.

The total number of events (deaths) is not mentioned in the results section at all, only the number of persons in the studies. The total number of deaths cannot be checked because they are not given in table 1.

In conclusion, it is unclear what survival analysis were made, unclear how many CVD deaths there were available, and it appears the the author is not aware the the number of deaths is the determinant of the precision of the results.

This paper does not meet scientific standards for reporting survival studies.

Shared genetic susceptibility for micro- and macro-vascular ...
This is a prevalence study of complications in 2203 Ukrainian DM patients. It is well reported and transparently analyzed.

It is a bit odd that prevalence is termed "risk" in the statistics section. Of slightly more concern is that in table 1 the DN+ / DN− classification only seem to have 1888 patients and not 2203 as the others, that is 315 patients less, quite a number since there are only 347 with DN+.

It is also not argued why analysis of macro-vascular complications are adjusted for age at DM, while analyses of micro-vascular complications are adjusted for duration of DM. And in particular why none of them are adjusted for age at investigation.

The limitations of a cross-sectional study where some exposures may contribute to mortality and thus removal of patients from the study population is not mentioned.

Famine exposure and genetic susceptibility to ...
This is essentially a replica of the previous study, but including famine exposure, defined as birth prior to 1950. Since data are collected in 2011–?? — a very short period, the famine exposure is almost the same as age over 62 (approximately). Incidentally the information about the *timing* of the data collection is absent in this paper even though large parts of the material and methods sections are identical to those in the previous.

The potential massive confounding by age which is so closely associated to the famine measure is not discussed at all in the paper. It is stated that clinical characteristics did not differ after adjustment for age, yet none of the main analyses are adjusted for age; only for age at DM onset and duration of DM. Thus it must be strongly suspected that results relating to famine are mere age-effects. Which is a pity since it would have been easy to adjust for age at examination (interview).

2 Comments on the revised version and notes

- Common variant in the *HMGA2* gene ...
  To my comment on the rather restricted analysis by logistic regression instead of survival analysis is said:

> Comments: the study on patients with type 2 diabetes has cross-sectional design given that the time of the event (occurrence of nephropathy) was not available.

However, in the statistical analysis section in the paper it is stated that:

> Diabetes duration was calculated from the age at onset of diabetes until development of nephropathy for cases and from age at onset of diabetes until last visit for controls.

Both cannot possibly be correct. The latter (from the published paper) allows for a survival analysis.

- *KCNJ11*/E23K polymorphism increases ...
  This paper has now been revised and gives a sound account of the analyses and a fairly clear overview of the studies included.

  The percentages of all deaths and CVD deaths has accidentally been interchanged (p. 13, numbers 7.5% and 16.6%).

- Shared genetic susceptibility for micro- and macro-vascular ...
  In the original manuscript it read:

  > Logistic regression was used to estimate genetic effects size expressed as odds ratio (OR) of the cardiometabolic loci adjusted for sex and age-at-onset for macrovascular complications, and sex and diabetes duration for microvascular complications.

  which I interpreted as:

  - macrovascular outcome: adjustment for sex and age-at-onset (implying "of DM")
  - microvascular outcome: adjustment for sex and DM duration (at investigation)

  Hence my comment on the disparity of the two different ways of adjusting.

  This fairly minor point has not been addressed, but the section shown above has been replaced by:

  > Logistic regression was used to estimate genetic effects size expressed as odds ratio (OR) of the cardiometabolic loci adjusted for sex and age-at-onset of macrovascular complications, and age, sex and diabetes duration for microvascular complications.

  Thus it is now stated that the risk of CVD (really prevalence odds of CVD) is adjusted for the age at CVD onset, also in the note to table 2.

  This is however not possible, since age at CVD is not defined for persons without CVD — the variable "age at onset of CVD" is of course only defined for CVD cases, and therefore cannot be used in the logistic regression analysis. So some other analysis must have been made.

- Famine exposure and genetic susceptibility to ...
  This paper now also contains the same error as the previous — claimed adjustment for a variable only defined for cases.

  The very close correlation of the famine variable and the age (at investigation) variable has not been addressed at all. The analyses in table 2 and the supplementary table are models with *interactions* by famine, and the relevant sensitivity analysis would be to replace the famine variable with a dichotomized age-variable say over/under 62 at investigation. It would be most illustrative to see histograms of age at examination for persons born before and after 1950 (famine yes/no).

  Since famine exposure plays such a vital role it is a bit odd that the definition of it is in the results section, it should be in the methods section.

  If analyses by age instead of famine show the same (as is my guess), then the conclusion in the paper would have to be that there is not information in the data to separate famine effects from age effects — be that interaction effects or main effects.

2.1 Conclusion

Having read the replies to my original comments I have seen an explanation which is 100% inconsistent with what is written in the published paper; one paper nicely corrected; and two papers where an explanation of analyses are given that cannot possibly be correct, one of which still contains a potential major flaw that has not been addressed.

On this background I still cannot recommend to the University of Lund to pass a PhD based on this work.