Author’s response to reviews

Title: Rationale and design of the CAROLINA® - cognition substudy: a randomised controlled trial on cognitive outcomes of linagliptin versus glimepiride in patients with type 2 diabetes mellitus

Authors:

Geert Jan Biessels (g.j.biessels@umcutrecht.nl)

Jolien Janssen (J.Janssen-9@umcutrecht.nl)

Esther van den Berg (E.vandenberg@erasmusmc.nl)

Bernard Zinman (zinman@lunenfeld.ca)

Mark Espeland (mespelan@wakehealth.edu)

Michaela Mattheus (michaela.mattheus@boehringer-ingelheim.com)

Odd Erik Johansen (odd-erik.johansen@boehringer-ingelheim.com)

Version: 1 Date: 14 Jul 2017

Author’s response to reviews:

Response to Reviewers’ Reports:

Naomi Chaytor (Reviewer 1): The authors have addressed all my concerns adequately (when possible, since the trial is underway).

Authors’ response:

Thank you for reviewing our paper.

Michael Donohue (Reviewer 2): This reviewer's primary interest was the primary analysis approach. The planned approach is to submit dichotomized residualized change on the MMSE, and a composite of trail making and verbal fluency tests, to a logistic regression. The MMSE and composite are first "residualized" to form the Regression Based Index (RBI) which is adjusted, via ANCOVA, for age, language, education, and baseline performance. Cognitive decline will be defined by an MMSE or composite RBI in the lowest 16th percentile (yielding an expect 20-22% of subjects in aggregate with cognitive decline). There are three potential concerns with this approach.

Authors’ response:
We thank the reviewer for the careful considerations on our analytical approach. The point that dichotomizing continuous outcome measures has implications for the analyses is well taken. We will respond to each of the three potential concerns of the reviewer below.

We should probably have emphasized more why we choose to dichotomize the cognitive test results for our primary outcome measure. In the intro we state: “Over the past years it has become clear, also from observational studies, that the average decline in cognition over time associated with diabetes is relatively slow, limiting the sensitivity of follow up studies to detect meaningful differences. Importantly however, among patients with T2DM there is heterogeneity in the rate of cognitive decline, where some have accelerated decline which in some cases progress to dementia.” This is the key reason why we choose to dichotomize; we intend to capture the subgroup that has the worst cognitive outcome, rather than averaging cognitive scores of people who do fairly well, with those with poor outcome. We have now further emphasized this point in the section “Primary outcome considerations p.15-16”, also addressing the potential impact that this approach has on the primary analysis.

The added paragraph reads:

Clearly, dichotomizing the cognitive test results for the primary outcome measure does have implications for the analyses. It is also different from the approach of previous studies in the field [10-13]. Of note, our rationale for the dichotomy is that it has become apparent that cognitive decline in older individuals with T2DM is clearly not a unitary construct [2]. On average – at the group level - cognition declines only very slowly over time [10-13]. Yet, there is a subset of individuals with accelerated decline [2]. While ideally this accelerated cognitive decline would be defined in terms of incident dementia or mild cognitive impairment, this was not deemed to be feasible in the present multinational, multicenter study, because of variability in diagnostic approaches. We therefore choose the pragmatic approach as described above, which is likely to capture the patients with the worst cognitive outcome, although not in terms of a fixed diagnostic construct. Dichotomizing the cognitive test results based on the RBI could result in an underestimation of the standard error of the primary estimate of group difference in rate of cognitive decline. It also comes at the expense of information loss and power. Yet, it was decided to sacrifice some statistical power in order to enable the possibility of having a more powerful statement at the end of the trial. Moreover, the actual change in cognitive performance at end of follow-up (i.e. change in performance from baseline as a continuous measure) is an additional predefined outcome measure to confirm the results of the primary analysis.

The first concern is that uncertainty in the first stage model is not reflected in the second stage model. This could result in an underestimation of the standard error of the primary estimate of group difference in rate of cognitive decline, and inflation of Type I error. The rationale for this approach appears to be that it "reduces the impact of learning effects." It is unclear what the impact of learning effects would be, and how this approach reduces them compared to, say, a direct MMRM of group difference controlling for the same covariates. Learning effects, if more common in one group, might be an important indication of treatment effect.

Authors’ response:
It is our experience that what can be termed as learning effects (i.e. performance on cognitive function tests increasing across repeated administrations) is real, as you suggest. It may be differentially expressed among individuals, often being more pronounced among those whose cognitive function remains high over time. We agree that it is unclear what the impact of learning effects would be, however, since the tests are conducted at baseline (Dec 2010 – Dec 2012), week 160 (Dec 2013 – Jan 2016) and at study end (estimated to occur Jan 2019-March 2019), i.e. for most patients with a time-window of around 3 years between the tests, we believe this may reduce the magnitude of this effect. Further we also have no a priori reason to expect that the size of these potential effects differ between the treatment groups. Nevertheless, in our two stage approach to analysis, we use covariate adjustment in the first stage to adjust for “potential confounders such as age, language, education, baseline performance, learning effects ...” The adjustment for learning effects is systematic between intervention groups and is based on including a marker of how many times the test had been repeated prior to the current assessment. For our primary outcome, this is used to order scores at each exam to determine the 16th percentile cutpoint. Unless there are marked differences in patterns of missing assessments between intervention groups, this should provide a relatively balanced approach to order test performance consistently between groups.

You are certainly correct that the two-stage approach we describe will affect the distribution of summary scores and very likely reduce their variability, as compared with raw scores.

Please also note that the actual change in cognitive performance at end of follow-up (i.e. a continuous outcome measure; change in performance from baseline) is also a predefined outcome measure. To confirm conclusions from our primary analysis model, the change from baseline in MMSE, A&E, TMT, and overall VFT z-scores are analyzed based on a MMRM with the fixed, categorical effects of treatment, week, treatment-by-week interaction, with the continuous covariates of baseline z-score and baseline-by-week interaction. Additionally sensitivity analyses will be performed using the same model however including baseline covariates for, among others, age, language and years of formal education.

The second concern is whether this method meets the standard of being robust to data Missing at Random (MAR). Direct likelihood methods, such as the secondary MMRM analysis is robust to the MAR assumption, but it is unclear that the proposed primary multi-step approach would also be. Multiple imputation might be a useful tool which could be adapted to address both of the above concerns -- incorporating uncertainty in the RBI step and ensuring robustness to MAR. Multiple imputation could be used to supplement the proposed plan for handling missing cognitive data, and impute missing data that remains after the planned review.

Authors’ response:

We agree that missing data is a potential concern, however, we believe that the tactical and operational measures in place will minimize this. In trials with cognitive outcomes, missing data are often informative (i.e. related to participants’ level of cognitive function). We also agree with the recommendation to use multiple imputation as a means to gauge the potential biases related to missing data. The approach we have outlined in our protocol for addressing missing
data (i.e. computing summary measures from partial data and using LOCF) is common in several settings.

In addition, we also perform predefined MMRM analyses on the changes in z-scores to account for MAR and confirm conclusions of our primary analysis model. While the protocol and analysis for our ongoing trial is set, we appreciate your suggestion and will include multiple imputation as you suggest to support analysis and study publications in case of a positive trend.

Third, dichotomization, while possibly facilitating clinical interpretation, might come at the expense of information loss and power (see, e.g., Cohen, J. (1983) The cost of dichotomization. Applied Psychological Measurement, 7, 249-253.) While this is a fairly large study which is probably adequately powered, why not use the most statistically efficient method? In this case, this might mean analyzing a single continuous outcome, a composite of all three: MMSE, trail making, and verbal fluency.

Authors’ response:

We certainly agree that choice of a dichotomous outcome reduces statistical power, and we weighed adopting the approach that you suggest of using a continuous measure. However, we were drawn to the importance of have a clear message that we felt would resonate strongly if we have a positive finding. It is, we feel, more powerful to be able to state that treatment reduced the incidence of what may be described as clinically significant cognitive deficits rather than stating the treatment led to a relative difference in mean composite cognitive function. The rationale for this is now emphasized more in the discussion (see first response above). We were willing to sacrifice some statistical power in order to enable the possibility of having a more powerful statement at the end of the trial. However, as outlined in the paper, a predefined analysis will also be addressing the individual test scores.

" Awkward: "study treatment stop (referred to as end of follow-up previously)" Why not just use "end of follow-up"?"

Authors’ response:

Thank you for suggesting an improvement in our language. We have rephrased this to: “Cognitive tests are conducted at baseline, after 160 weeks and at planned end of follow-up (or at permanent treatment-discontinuation).

Additional small edits:

When preparing the Trial Statistical Analysis Plan for the sister trial of CAROLINA, CARMELINA, we discovered some small items that needed correction:

- calculation of the z-scores of the fluency test is done for the verbal and category fluency separately, this was not stated in a clear enough fashion and has now been rephrased
we erroneously indicated that fasting plasma glucose was measures prior to cognitive testing, this should have been “self-monitoring of blood glucose” and is now corrected on page 14.

we have described in more detail how we deal with “floor scores” at baseline on the TMT and fluency tests (page 15)