Author's response to reviews

Title: Single blind randomized Phase III trial to investigate the benefit of a focal lesion ablative microboost in prostate cancer (FLAME-trial)

Authors:

Irene M. Lips (i.m.lips@umcutrecht.nl)
Uulke A. van der Heide (u.vd.heide@nki.nl)
Karin Haustermans (karin.haustermans@uzleuven.be)
Emiel van Lin (e.vanLin@rther.umcn.nl)
Floris Pos (f.pos@nki.nl)
Stefan P.G. Franken (s.franken-3@umcutrecht.nl)
Alexis N.T.J. Kotte (akotte@umcutrecht.nl)
Carla van Gils (c.vangils@umcutrecht.nl)
Marco van Vulpen (m.vanvulpen@umcutrecht.nl)

Version: 2 Date: 3 November 2011

Author's response to reviews: see over
Dear Editor-in-Chief,

Please find attached our revised manuscript entitled: “Single blind randomized Phase III trial to investigate the benefit of a focal lesion ablative microboost in prostate cancer (FLAME-trial)”.

Thank you for the review procedure and the helpful comments on our manuscript.

We are pleased to be able to respond to the reviewer’s comments. We have corrected the manuscript following the reviewer’s comments. All changes in the text are indicated below.

Yours sincerely,

Irene Lips, M.D., Ph.D.
Corresponding author
We thank the reviewer for all valuable suggestions for improvement of the manuscript.

Response to the comments from the reviewer:

- **Eligibility criteria. What is a WHO score >2? WHO score of what?**
  The WHO (World Health Organization) performance scoring system defines >2 as ‘symptomatic, >50% during the day in bed, but not bedbound (score 3)’ or ‘bedbound’ (score 4). To make this clear, we added the reference Oken et al, 1982 and ‘performance’ and ‘(= symptomatic, >50% during the day in bed, but not bedbound (score 3) or bedbound (score 4)).’ to the text.

- **Justify the use of multiple overlapping questionnaires. It is very peculiar to have patients fill in so many questions, especially when most have nothing to do with prostate cancer or treatment side-effects. Questionnaires every 6 months for 10 years? This is a massive amount of data. It is probably 100 questions per time point, multiplied by 600 patients, times 20 questions, that is over 1 million data points. This huge burden for patients and research staff needs to be justified.**
  The disease- and treatment-related side effects are only a component of health-related quality of life (Litwin et al, 1999). To address the components of overall well-being, a general instrument is used in addition to the disease-specific questionnaires, which is frequently described in literature (Inoue et al, BJU Int, 2009; Nout et al, JCO, 2011; Kornblith et al, JCO, 2011).
  To be able to determine the point in time at which the QoL changes, for example due to side-effects after treatment, it is important to measure QoL every 6 months after treatment. It takes only 20 minutes for a patient to fill in the questionnaires.

- **The sample size is peculiar for two reasons. First, the use of a one sided p value is very unusual, indeed, I cannot remember seeing it before in a Phase III trial of a new cancer intervention.**
  The increase in radiation dose in the experimental arm will result in more loss of tumor cells in the experimental arm, if anything. Therefore an increase in biochemical failure in the experimental arm is improbable and thus a one sided p-value was chosen.

  Second, the analysis is based on time to event data (Cox modeling) but the sample size is based on binary comparison of proportions. A more standard sample size approach would be to take the time to event nature of the data into account, calculating the number of events expected within a given length of follow-up.
  Indeed for the analysis of our binary endpoint, defined as the five-year freedom from biochemical failure rate, a Chi-square test will be used. Therefore, we added to the data analysis section: ‘To test the difference between the two treatment arms in the proportions of patients who are still free from biochemical failure after five years the Chi square test will be used.’
  The time to event analysis for biochemical failure will be performed as a secondary analysis. The trial is well powered for such an analysis, because a one-sided log rank survival power analysis shows that with an overall sample size of 566 subjects (283 patients per treatment arm) a power of 86% at a 5% significance level is achieved to detect a difference of 10% (64% and 74% free from biochemical failure after 5 years, in the control arm and experimental arm, respectively). This is under the assumption that the patients entered the study during an accrual period of 5 years, 50% of the enrollment was complete when 70% of the accrual time had past and that 20% of the patients in both arms were lost to follow-up.
during the follow-up period of 5 years (Lips et al, BJU int. 2009).

Is it really the case that all patients will be followed for five years? State the length of follow-up after accrual of the last patient.
Yes, all patients will be followed for at least five years.

- The QoL analysis is hugely underpowered. You are talking all that data from patients and then just saying "better or worse?". It might be that all patients get worse, so no difference between groups, but patients in one treatment group don't get as bad as the other. Please use a standard approach such as ANCOVA, with baseline score as a covariate (see Vickers and Altman, BMJ)
We chose to power our trial for the primary endpoint and not to reveal statistically significant differences in the secondary endpoints such as QoL. We agree that a more extensive analysis for the QoL would be a better approach so therefore we changed this part into: “To analyze differences in QoL between the two treatment groups over the time points, a general linear model repeated-measures analysis of covariance will be performed (Vickers et al, 2001).”