The TARGIT-A trial: understanding non-inferiority and survival analysis

Joanne S Haviland MSc[1], Judith M Bliss MSc[2], Soeren M Bentzen PhD, DMSc[3], Jack Cuzick PhD[4]

[1] Faculty of Health Sciences, University of Southampton, Southampton, UK

[2] ICR-CTSU, Division of Clinical Studies, The Institute of Cancer Research, London, UK

[3] Division of Biostatistics & Medical Informatics, Department of Epidemiology & Public Health, University of Maryland School of Medicine, Baltimore, US

[4] Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London WC2A 3PX, UK

Corresponding author:

Joanne S Haviland

Faculty of Health Sciences, University of Southampton, Highfield, Southampton, SO17 1BJ, UK

Tel: +44 (0) 23 8059 7860

Fax: +44 (0) 23 8059 7967

Email: j.s.haviland@soton.ac.uk

Conflict of interest: none

In Vaidya's response to Hepel and Wazer[1], Figure 1 illustrates how the upper confidence limit for the difference between treatment outcomes allows assessment of non-inferiority in a trial[2]. This follows FDA guidelines[3]. It is imperative however that the underlying statistical analysis method and associated confidence limit are appropriate to the event of interest and the completeness of data. It is disappointing, therefore, that the TARGIT trialists persist in using a non-inferiority test statistic based on binomial proportions. Binomial analysis simply divides the number of recurrences by the total number of patients; thus subjects with one month or 5 years' follow-up contribute the same to the denominator. Moreover, subjects with very short follow-up are counted as *not* having had a LR. This is flawed as shown by a simple example. Assume two groups of 30 patients with long follow-up and 10 and 20 failures respectively: a statistically significant (P=0.01) difference of 33% in failure rates is observed. Adding 200 cases with very short follow-up to each group (contributing no additional failures), the difference in failure rates is now only 4% and no longer significant. In TARGIT-A fewer than 700 patients have at least 5 years' follow-up or an observed LR – i.e. have the full information required for a binomial analysis. In contrast, survival analysis methods use all available data, account for varying follow-up and timing of events, and incorporate censoring (a subject without an event at the time of last contact has a risk of failing in the future). The TARGIT trialists argue that their assessment is better because it uses all recorded events, in contrast to the "single snapshot point estimates of 5-year recurrence rates" obtained from Kaplan-Meier (K-M) analysis – revealing a fundamental misunderstanding of survival analysis.

Appropriate assessment of non-inferiority in the TARGIT-A trial would employ survival analysis to estimate the absolute difference in 5-year recurrence rates (protocol-specified primary endpoint), with a confidence interval (CI). The upper confidence limit would indicate whether or not the pre-specified threshold for non-inferiority had been crossed. Survival analysis provides a hazard ratio (HR) calculated

using all reported events, which can be applied to any time-specific rate to obtain an estimate of the difference in event rates between treatment groups (with CI)[4]. As the relevant figures were not presented in the TARGIT-A Lancet 2014 paper[5], it was previously necessary to estimate them indirectly from the information provided to establish an estimate which accounted for the variability in both treatment groups and not just TARGIT[6,7]. The TARGIT trialists can and should provide a proper analysis of LR rates at 5 years (with CI), to enable an unequivocal assessment of non-inferiority.

Vaidya's citation of Cuzick[8] reflects a fundamental misunderstanding of 1-sided versus 2-sided CIs as calculation of the upper limit of a 2-sided 90%CI provides the same limit of a 1-sided 95%CI. Vaidya confirms that the significance level set for the primary outcome changed from 5% in the protocol to 1% for the final analyses; therefore should assessment of non-inferiority not be based on the higher 1-sided 99%CI?

Another major misunderstanding is to state that predefined strata are not subgroups. P-values are designed for a single predefined hypothesis, and should not be applied to separate subgroups/strata without a Bonferroni correction. The TARGIT protocol clearly states that the main analysis will include all randomised patients and the focus on the pre-pathology subgroup was clearly post-hoc after seeing the results. The dangers of restricting results to subgroups are well-known[9].

References:

1. Hepel J, Wazer DE. A flawed study should not define a new standard of care. Int J Radiat Oncol Biol Phys 2015; 91(2):255-7.

2. Vaidya JS, Bulsara M, Wenz F, et al. Pride, prejudice, or science: Attitudes towards the results of the TARGIT-A trial of targeted intraoperative radiotherapy for breast cancer. Int J Radiat Oncol Biol Phys 2015; DOI: 10.1016/j.ijrobp.2015.03.022.

3. Food and Drug Administration. Guidance for Industry: Non-inferiority clinical trials. 2010. http://www.fda.gov/downloads/Drugs/Guidances/UCM202140.pdf (accessed May 6, 2015).

4. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. BMJ 1999; 319: 1492–95.

5. Vaidya JS, Wenz F, Bulsara M, et al, on behalf of the TARGIT trialists' group. Risk-adapted targeted intraoperative radiotherapy versus whole-breast radiotherapy for breast cancer: 5-year results for local control and overall survival from the TARGIT-A randomised trial. Lancet 2014; 383: 603–13

6. Parmer MKB, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat Med 2008; 17: 2815–34.

7. Haviland JS, A'Hern RA, Bentzen SM, Whelan T, Bliss JM. Radiotherapy for breast cancer, the TARGIT-A trial. Lancet 2014; 383 (9930): 1716-7.

8. Cuzick J. Radiotherapy for breast cancer, the TARGIT-A trial. Lancet 2014; 383 (9930): 1716.

9. Cuzick J. Forest plots and the interpretation of subgroups. Lancet 2005; 365 (9467): 1308.