1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What’s the main question being asked or hypothesis being tested in this study?
Speech and neurocognitive function are important domains of investigation in head and neck cancer (HNC) patients. Research has also indicated that speech and neurocognitive problems may already be present in HNC patients at baseline, prior to treatment (Bond et al., 2012; Borggreven et al., 2007; Cnossen et al., 2012). A previous study reporting on cognitive impairment pre-treatment did not control for effects of substance abuse (Bond et al., 2012). Moreover, self-reported neurocognitive function had a substantial floor effect for this measure, preventing any conclusions about self-perceived neurocognitive function in this group (Bond et al., 2012).

Neurocognitive function and speech are intertwined in speaking and both seem to show pre-treatment deficits in HNC patients. To date, no study has reported on baseline in speech and (self-reported and objective) neurocognitive function in the same group of patients.

The proposed study has the following aims:

(1) Document neurocognitive function, as measured objectively and through patients' self-report, pre-treatment in HNC patients using data from the NET-QUBIC Kubus project.

(2) Identify demographic (gender, age, education, literacy), behavioural (alcohol and drug abuse), and disease-related (tumour location and stage) features associated with low neurocognitive function in HNC patients. We hypothesise that substance abuse will be a strong predictor of cognitive function.

(3) Characterise the relationship between neurocognitive function and self-perceived speech function in HNC patients pre-treatment.

3) Describe the key dependent variable(s) specifying how they will be measured.
All variables provided by the NET-QUBIC Kubus database:
- Demographic: Gender, age, education, literacy
- Behavioural: Alcohol and drugs study-specific questionnaires
- Disease: Tumour location and stage

- Speech: Speech Handicap Index (Rinkel et al., 2008)

- Objective neuropsychological tests: Hopkins Verbal Learning Test (immediate and delayed recall); Trail making A and B; Controlled Oral Word Association Test
- Self-reported cognition: Cognitive Failures Questionnaire (Broadbent et al., 1982)

4) How many and which conditions will participants be assigned to?
Not applicable

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
Aim (1): normative data will be used to convert the patients' scores of the neuropsychological tests into t-scores, adjusted for age and education. A global deficit score will be calculated as the average of the t-scores converted to deficit scores (cf. Bode et al., 2012). The relationship between self-perceived and objectively measured cognitive function will be assessed with Spearman’s correlation analysis for each variable separately, multiple comparisons corrected via the Holm-Bonferroni method, and for the global deficit score.

Aim (2): LASSO regression will be used to assess the relative contribution of demographic (Gender, age, education, literacy), behavioural (alcohol and drug abuse), and disease-related (tumour site and location) factors on the global deficit score. Of particular interest is the extent to which alcohol abuse can explain low cognitive functioning in the HNC patients.

Aim (3): Strength and direction of the relationship between neurocognitive and speech function will be assessed with non-parametric (Spearman's) correlation analysis for each neurocognitive variable separately, multiple comparisons corrected via the Holm-Bonferroni method. For this analysis, the neuropsychological tests of interest are Trailmaking B (a measure of executive function), Controlled Oral Word Association Test (a measure of verbal fluency, which depends both on linguistic, motor, and executive processes), and delayed recall (a measure of verbal long-term memory).

6) Any secondary analyses?

Verify authenticity: http://aspredicted.org/blind.php?x=wn7uk2
7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
Analyses will be run on all baseline data currently available at the net-qubic database. At the time of writing, the sample size varies between 206 and 218 participants, depending on the assessment.

8) Anything else you would like to pre-register? (e.g., data exclusions, variables collected for exploratory purposes, unusual analyses planned?)
This study is part of a very large database effort (https://researchers.kubusproject.nl/general-information). Data have been collected already, but no analyses have been conducted yet. The very first data release has been launched in June 2016 and proposals for accessing those data have been granted in the past couple of months, so no one has had enough time to analyse any data. Besides, the present project is just one within the whole effort and the variables assessed here have not being the topic of investigation by anyone else. Thus, although the data have already been collected, as of this date, the data have not been accessed by the researcher conducting the present study. All research plans reported here have been formulated without any knowledge of results originating from the data.
This document describes changes in the manuscript's analyses relative to the pre-registered analyses.

1) The original plan was to calculate a global deficit score and use it to correlate with patient-reported neurocognitive functioning, in addition to the individual neurocognitive tests. Upon reflection, this score seemed less meaningful than the individual tests themselves, so this analysis was not performed.

2) The original plan did not include the subscales of the Speech Handicap Index (SHI). However, given that a relationship was found between patient-reported outcomes (i.e., total SHI score and cognitive failures score), additional analyses were included for the SHI subscales, which reflect psychosocial and speech function separately. This analysis was intended to find additional support for the hypothesis that the correlation between the patient-reported outcomes reflects a more neurocognitive/psychosocial relationship, rather than a peripheral, motor-speech one.

3) The original plan was to perform LASSO regression using demographic (gender, age, education, literacy), behavioural (alcohol and drug abuse), and disease-related (tumour site and location) factors on the global deficit score. However, LASSO regression proved unsuitable for the present dataset due to sample size (i.e., still too small for this type of statistical analysis). In addition, drug abuse and literacy did not show enough variance to be included as predictors in the model. Given that we did not calculate a global deficit score, the regression analyses were conducted on the individual neuropsychological tests instead.

All additional analyses reported in the manuscript were added during the review process.