PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Neuroinflammation in cognitive decline post-cardiac surgery (the FOCUS study): an observational study protocol
AUTHORS	Peters van Ton, Annemieke; Duindam, Harmke; van Tuijl, Julia; Li, Wilson; Dieker, Hendrik-Jan; Riksen, Niels; Meijer, F.J.; Kessels, Roy; Kohn, Nils; van der Hoeven, J.G; Pickkers, Peter; Rijpkema, Mark; Abdo, Wilson

VERSION 1 – REVIEW

REVIEWER	Miles Berger Duke University Medical Center, Durham NC (usa)
REVIEW RETURNED	16-Nov-2020

GENERAL COMMENTS	This is a timely and relevant paper to the field of periop neurocognitive disorders describing the study protocol for the FOCUS study, designed to study the relationship between neuroinflammation (as assessed by PET TSPO binding) and postoperative cognitive dysfunction after cardiac surgery. The following comments and questions are intended to improve the manuscript by clarifying relevant details and by putting this work in fuller context of other related studies in the field. “This study is the first to examine in vivo neuroinflammation both prior to and after cardiac surgery”- this is an overstatement and should be revised. Other studies have measured CSF inflammatory cytokine levels in cardiac surgery patients before and after surgery, so this is not really the first study to examine in vivo neuroinflammation before and after cardiac surgery. Pls revise. Is there a control group of non-surgical patients, preferably patients with similar CAD who are medically managed? If not, could the authors please discuss why not? This paper should reference other similar studies addressing this same general question (role of neuroinflammation in POCD; i.e.PMIDs 32417770, 30674067, 31930549) and should discuss how the findings of this study are expected to complement those of these other studies. P 4 the word “enrollment” should be spelled with two l’s In the section on exclusion criteria, will patients on anti inflammatory or immunomodulatory drugs be excluded from the trial? If not, the authors should discuss how the results in such patients will be accounted for, due to the potential confounding effect of these drugs. Will patients on CNS-active drugs besides neuroleptics be excluded, such as anti anxiety or antidepressant drugs? If not the authors should discuss how these potential confounders will be addressed in the analysis.

	P 6 in the section on resting state fmRI, the authors should cite and discuss the prior rs-fmri studies in cardiothoracic surgery patients (PMIDs 29164774, 27858963, 27858963, 29034164) P 7 can the authors provide additional detail on how the flow cytometry will be performed, or reference a prior paper using this method? P 7 bottom- “Perioperative care will be executed conform the standard of care protocols for CABG, minimizing confounders due to medical management.” . . . awkward wording, please revise sentence. P 9- “A cognitive domain will be classified as impaired when the average rating of tests is >1” …meaning unclear, please rephrase. Do the authors mean a domain is impaired when the Z score drop is >1? P 9- how many cognitive domains will there be in the proposed analysis? Please list the domains. Re the statement that POCD will be defined in part by “the average test rating has declined in more than one domain compared to baseline”… how much of a drop is required to constitute a “decline” ? Any drop? If so, then will any patient who drops by a z score of 0.01 in 2 domains count as POCD? This seems a bit lenient… POCD has also been viewed as a continuous phenotype, i.e as part of a normal distribution. As such, do the authors plan on any continuous correlation analyses between changes in TSPO binding and overall cognitive changes across tests, or within domains, as continuous variables? Such analyses may provide significantly more statistical power than the dichotomous analyses proposed. Sample size- 30 patients is effectively a pilot study, since this same size will be too small to perform a proper multiple variable analysis of predictors of POCD, given that only 15 patients are expected to develop POCD. This sample size will be inadequate to adjust for confounders as a result. This should be discussed as a limitation of this study, and this should be explicitly described as a pilot study. The authors say that the study will be stopped if increased PET tracer uptake is not observed after CABG surgery compared to the preoperative baseline tracer uptake (bottom of p 10). How does this account for the possibility that increased postop uptake may be seen only in the patients with POCD, and decreased uptake may be seen in patients without POCD, such that on average there is no postop change? P 11- pls fix spelling of word “enrollment”

REVIEWER	Julie Lasselin Stockholm University, Sweden
REVIEW RETURNED	03-Dec-2020

GENERAL COMMENTS	This is an interesting study planned by van Ton et al that aims at assessing a potential role of neuroinflammation in the development of cognitive disturbances after cardiac surgery in middle-aged and old patients. I have only minor comments: - In the abstract, the authors states that ”serial extensive neurocognitive assessments will be performed”, while the

	neurocognitive evaluation is conducted at only two time points outside of the hospitalization period. - In the introduction, the authors state that ”Severe systemic inflammation can result in increased permeability of the blood-brain barrier (BBB), enabling systemic inflammatory cytokines to enter the brain.[10, 11] As a result, systemic inflammation may induce activation of the innate immune cells of the brain, the microglia and astrocytes, leading to a neuroinflammatory response.[12]”. This suggests that activation of microglia and astrocytes can only result from a severe systemic inflammation that causes permeability of the BBB, while neuroinflammation can also occurs when systemic inflammation is lower and without obvious changes in the permeability of the BBB (see for instance, D’Mello and Swain, Curr Top Behav Neurosci, 2017, doi: 10.1007/7854_2016_37). - In the introduction, the authors could more clearly define what they mean by “substantial systemic inflammation” that is observed after cardiac surgery. - In the exclusion criteria, an active infection could also be ruled out by measuring levels of CRP (e.g. very high levels of CRP such as >20-50 mg/L). - The authors could provide more indication about the validation of the second generation TSPO ligand to measure neuroinflammation. - The authors indicate that various systemic inflammatory markers will be measured, as well as danger-associated molecular patterns, but do not indicate by which mean they will be measured.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Dr. Miles Berger, Duke University Hospital

Comments to the Author:

This is a timely and relevant paper to the field of periop neurocognitive disorders describing the study protocol for the FOCUS study, designed to study the relationship between neuroinflammation (as assessed by PET TSPO binding) and postoperative cognitive dysfunction after cardiac surgery.

The following comments and questions are intended to improve the manuscript by clarifying relevant details and by putting this work in fuller context of other related studies in the field.

1. “This study is the first to examine in vivo neuroinflammation both prior to and after cardiac surgery”- this is an overstatement and should be revised. Other studies have measured CSF inflammatory cytokine levels in cardiac surgery patients before and after surgery, so this is not really the first study to examine in vivo neuroinflammation before and after cardiac surgery. Pls revise.

Authors’ comments: In with this comment we additionally specified this in the manuscript.

Authors’ actions: Page 2:

“This study is the first to examine in vivo neuroinflammation using TSPO PET neuroimaging both prior to and after cardiac surgery”

2. Is there a control group of non-surgical patients, preferably patients with similar CAD who are medically managed? If not, could the authors please discuss why not?

Authors' comments: This is an interesting suggestion for a different study. Our aim here is to assess whether a systemic inflammatory state induces neuroinflammation and whether this is associated with neurocognitive outcomes. Our study setup could serve as a proof-of-principle study for disease states accompanied with severe systemic inflammation (sepsis, trauma, COVID, etc), while including medically treated patients would serve as a proof-of-principle setup for disease states with low systemic inflammation (diabetes, obesity, low physical activity, etc). A medically treated group is therefore different and would also need a large number of patients because they don't suffer from a severe inflammatory state and are likely to experience less neuroinflammation. This comes with high costs as MRI and TSPO PET imaging are very expensive and it would also mean that the study participants are not really well comparable to the surgery group we use. Those that undergo cardiac surgery develop a severe systemic inflammatory response while those that are treated medically do not. This control group is therefore not needed to investigate our study aims and would expose more participants to unnecessary radioactivity.

Authors' actions: None to the manuscript.

3. This paper should reference other similar studies addressing this same general question (role of neuroinflammation in POCD; i.e.PMIDs 32417770, 30674067, 31930549) and should discuss how the findings of this study are expected to complement those of these other studies.

Authors' comments: To remain within the word limits we shortly discussed these references on page 12.

Authors' actions: on page 12:
"This adds to previous and ongoing observational work perioperatively combining blood and cerebrospinal fluid parameters with MR neuroimaging, cognition and electroencephalogram recordings."

4. P 4 the word "enrollment" should be spelled with two l's

Authors' comments: For the British Medical Journal (BMJ) Open, we used British English, in which enrolment is the correct spelling according to the Cambridge Dictionary. If the editor prefers otherwise, please let us know, so we can adapt the spelling in our manuscript.

Authors' actions: None

5. In the section on exclusion criteria, will patients on anti-inflammatory or immunomodulatory drugs be excluded from the trial? If not, the authors should discuss how the results in such patients will be accounted for, due to the potential confounding effect of these drugs.

Authors' comments: Yes, immunocompromised patients (due to immunomodulatory drugs or underlying conditions) will be excluded from the trial.

Authors' actions: We clarified this more in Table 1.

6. Will patients on CNS-active drugs besides neuroleptics be excluded, such as anti-anxiety or antidepressant drugs? If not the authors should discuss how these potential confounders will be addressed in the analysis.

Authors' comments: The use of anti-anxiety or antidepressant drugs is not an exclusion criterium. For anxiolytics, all patients will receive benzodiazepines and propofol, as these are part of standard of

anaesthetic care. We expect minimal confounding effects of antidepressants since we compare the individuals' delta between pre- and postoperative TSPO binding.

Authors' actions: none

7. P 6 in the section on resting state fmRI, the authors should cite and discuss the prior rs-fmri studies in cardiothoracic surgery patients (PMIDs 29164774, 27858963, 27858963, 29034164)

Authors' comments: Due to manuscript length restrictions and the fact that functional MRI is only a secondary outcome, we decided to discuss only the study from Browndyke et al., since this work is most related to our protocol.

Authors' actions: We added the following on page 6:

"A previous study found positive associations between alterations in resting-state functional connectivity in the brain's default mode network and global cognitive change after cardiac surgery[34], and the current study enables us extend these findings by studying the relationship with (neuro)inflammation."

8. P 7 can the authors provide additional detail on how the flow cytometry will be performed, or reference a prior paper using this method?

Authors' comments: We added the suggested references on page 7.

Authors' actions: We added on page 7:

"Flow cytometry analysis of whole blood will be performed as described previously [Leijte et al., Critical Care 2020, Noz et al., J Am Heart Assoc 2019], to study the inflammatory phenotype of the cells (including expression of human leukocyte antigen-DR (HLA-DR), C-C chemokine receptor type 2 (CCR2), CD11b, CD14, and CD16).

9. P 7 bottom- "Perioperative care will be executed conform the standard of care protocols for CABG, minimizing confounders due to medical management." . . . awkward wording, please revise sentence.

Authors' comments: We revised this sentence

Authors' actions: "Perioperative care will be delivered according to the regular clinical protocol for CABG. This minimizes the risk of potential confounders due to variability in medical management."

10. P 9- "A cognitive domain will be classified as impaired when the average rating of tests is >1" …meaning unclear, please rephrase. Do the authors mean a domain is impaired when the Z score drop is >1?

Authors' comments:
We rephrased our explanation and introduced the term 'weighted test score' to clarify that the overall cognitive domain score (impaired or not impaired) is not based on average z-scores, but on clinically more relevant weighted scores.

The z-scores for each test will be calculated using standardized normative data. Furthermore, we classify these individual z-scores into three – clinically relevant - categories, with a corresponding weighted score (0, 1 or 2) per test. Each cognitive domain is comprised of one or more neuropsychological tests. The overall domain classification (impaired or not impaired) is based on the *average* of weighted test scores within a domain. Every neuropsychological assessment will result in five domain scores, with both an averaged z-score (continuous variable) and a dichotomized classification per domain (impaired or not impaired).

We consider a domain impaired when the *weighted* domain score is >1. To clarify this, we show an example of calculating the test results of one domain in the table below. The individual test scores within a domain create an overall domain score (both as a standardized z-score and a classified score).

Example of analysis of a domain score (in this case: **Executive Function**)				
Timepoint→	Baseline		Hospital discharge	
Tests↓	z-score	Classification (average score)	z-score	Classification (average score)
Trail making test B	-1.70	2 (Impaired)	-1.66	2 (Impaired)
Stroop colour-word test	-0.5	0 (Normal)	-0.7	0 (Normal)
WAIS – digit span	-1.30	1 (Below average)	-1.70	2 (Impaired)
Overall domain score	**-1.17**	**1 (not impaired)**	**-1.35**	**1.33 (impaired)**
Classification scores per test: 0 =normal (if SD >-1), 1 =below average (if SD -1 to -1.65), 2 =impaired (if SD < -1.65)				
Classification scores per domain: average score of all tests. >1 = impaired, ≤1 = not impaired				

Authors' actions: We revised the following paragraph on page 9, introducing the term "weighted score":

*"*Overall domain z-scores will be calculated by averaging the z-scores of the individual tests within a specific domain. Calculation of test performance will be supervised by an experienced clinical neuropsychologist. Additionally, z-scores of each individual test will be clinically classified as either being within the normal range (0), below average (1), or impaired (2), compared to the aforementioned normative data. "Normal performance" is defined as performance above -1 SD from the normative mean. "Below average" as between -1 SD and -1.65 SD from the normative mean (the lowest 16% of the normal population), and "impaired" as below -1.65 SD from the normative mean (the lowest 5% of the normal population).[41, 42] Thus, an overall weighted performance score will be calculated for all five cognitive domains. A cognitive domain as a whole will be clinically classified as *impaired* when the average weighted score of all tests within the domain is >1."

11. P 9- how many cognitive domains will there be in the proposed analysis? Please list the domains.

Authors' comments: In our previous version of the protocol, table 3 explains the different cognitive subdomains assessed by the individual tests. Based on the feedback from the reviewer, we added a new column showing the five coinciding (main) domains: executive function, information processing speed, episodic memory, visuoconstructive ability and language.

Authors' actions:
Revision of table 3:
* Changed header of column "*Assessment of*" in "*Subdomain*"
* Added extra column "*Domain*", disclosing the domain each tests belongs to.

12. Re the statement that POCD will be defined in part by "the average test rating has declined in more than one domain compared to baseline"… how much of a drop is required to constitute a "decline" ? Any drop? If so, then will any patient who drops by a z score of 0.01 in 2 domains count as POCD? This seems a bit lenient…

Authors' comments:
When creating our definition of POCD, we attempted to emphasize *clinically relevant* outcome measurements. We fully agree with the statement of the reviewer that a 0.01 SD change of cognitive function cannot be considered clinically relevant. Therefore, the definition of POCD is based on a calculation using the *weighted* scores,. This ensures that z-scores will have to change meaningfully in order to change the categorized weighted score per test. This way, when an overall weighted domain

score changes, this necessarily reflects a significantly clinical change in cognitive function within a domain.

Authors' actions:
Revision of our definition of POCD on page 10:
"Our definition of POCD is: 1) a patient is *newly impaired* in one or more cognitive domains at hospital discharge, compared to baseline, OR 2) the overall weighted performance score deteriorated in more than one domain at hospital discharge, compared to baseline."

13. POCD has also been viewed as a continuous phenotype, i.e as part of a normal distribution. As such, do the authors plan on any continuous correlation analyses between changes in TSPO binding and overall cognitive changes across tests, or within domains, as continuous variables? Such analyses may provide significantly more statistical power than the dichotomous analyses proposed.

Authors' comments: POCD is a clinical syndrome. Therefore, we designed this weighted classification model to define its presence or absence. However, we do collect continuous cognitive outcome measurements (z-scores per test and domain), although they have not been described as our primary outcome. Based on the suggestions of the reviewer, continuous cognitive outcome variables will be used in additional secondary analyses, correlated with tracer uptake. Page 9 already describes that standardized z-scores are collected, which will be used as a continuous variable for cognitive function. Due to the maximum word count, we have not described these additional secondary analyses in further detail.

Authors' actions: None

14. Sample size- 30 patients is effectively a pilot study, since this same size will be too small to perform a proper multiple variable analysis of predictors of POCD, given that only 15 patients are expected to develop POCD. This sample size will be inadequate to adjust for confounders as a result. This should be discussed as a limitation of this study, and this should be explicitly described as a pilot study.

Authors' comments: We agree with the reviewer.

Authors' actions: We adapted the manuscript accordingly on page 13 and after the abstract:

"Finally, the power of this pilot study is inadequate to allow adjustment for confounding factors."

15. The authors say that the study will be stopped if increased PET tracer uptake is not observed after CABG surgery compared to the preoperative baseline tracer uptake (bottom of p 10). How does this account for the possibility that increased postop uptake may be seen only in the patients with POCD, and decreased uptake may be seen in patients without POCD, such that on average there is no postop change?

Authors' comments: The independent DSMB will be unblinded for the stratification of patients in both groups (with and without POCD), to perform this interim analysis.

Authors' actions: None

16. P 11- pls fix spelling of word "enrollment"

Authors' comments: For the British Medical Journal (BMJ) Open, we used British English, in which enrolment is the correct spelling according to the Cambridge Dictionary. If the editor prefers otherwise, please let us know, so we can adapt the spelling in our manuscript.

Authors' actions: None

***

Reviewer: 2
Dr. Julie Lasselin, Stockholm University, Karolinska Institute

Comments to the Author:
This is an interesting study planned by van Ton et al that aims at assessing a potential role of neuroinflammation in the development of cognitive disturbances after cardiac surgery in middle-aged and old patients.

I have only minor comments:

1. In the abstract, the authors states that "serial extensive neurocognitive assessments will be performed", while the neurocognitive evaluation is conducted at only two time points outside of the hospitalization period.

Authors' comments: "Serial" refers to the total of four consecutive neuropsychological examinations performed during the study period in each patient.

Authors' actions: None

2. In the introduction, the authors state that "Severe systemic inflammation can result in increased permeability of the blood-brain barrier (BBB), enabling systemic inflammatory cytokines to enter the brain.[10, 11] As a result, systemic inflammation may induce activation of the innate immune cells of the brain, the microglia and astrocytes, leading to a neuroinflammatory response.[12]". This suggests that activation of microglia and astrocytes can only result from a severe systemic inflammation that causes permeability of the BBB, while neuroinflammation can also occurs when systemic inflammation is lower and without obvious changes in the permeability of the BBB (see for instance, D'Mello and Swain, Curr Top Behav Neurosci, 2017, doi: 10.1007/7854_2016_37).

Authors' comments: We agree with the reviewer that less severe systemic inflammation can also induce neuroinflammation.

Authors' actions: We revised the following sentence on page 3, and included the sugggested reference:

"Systemic inflammation can result in increased communication and signalling from the periphery to the brain [10-11]. As a result, systemic inflammation may induce activation of the innate immune cells of the brain, the microglia and astrocytes, leading to a neuroinflammatory response.

3. In the introduction, the authors could more clearly define what they mean by "substantial systemic inflammation" that is observed after cardiac surgery.

Authors' comments: none

Authors' actions: We revised the sentence on page 3:

"Cardiac surgery produces substantial systemic inflammation (reflected by leucocytosis and significant dysregulation of cytokines and other inflammatory mediators, affecting various physical processes) due to multiple stimuli such as sternotomy, extracorporeal circulation (ECC), associated transient endotoxemia, and aortic cross-clamping. "

4. In the exclusion criteria, an active infection could also be ruled out by measuring levels of CRP (e.g. very high levels of CRP such as >20-50 mg/L).

Authors' comments: The exclusion criteria are assessed by medical doctors who will combine all clinical data of patients including laboratory values such as CRP to judge whether an active infection is present and whether or not a patient should therefore be excluded from study participation. The patients that enter this study are patients from the outpatient clinic that are on the waiting list for an elective CABG. CRP is not used as a single exclusion criterium. When an infection is suspected full infection screening including CRP is performed, which is part of the routine clinical work-up by the surgeons because an active infection is considered an contra-indication to undergo the elective CABG.

Authors' actions: None.

5. The authors could provide more indication about the validation of the second generation TSPO ligand to measure neuroinflammation.

Authors' comments: We agree and adjusted the manuscript accordingly.

Authors' actions: As suggested we added the following sentence on page 6:

"This generation of TSPO ligands outperforms the first generation TSPO tracer, isoquinoline carboxamide (^{11}C-PK11195) on TSPO binding affinity and PET imaging properties, but requires polymorphism genotyping and stratifying according to binding affinity status[32]."

6. The authors indicate that various systemic inflammatory markers will be measured, as well as danger-associated molecular patterns, but do not indicate by which mean they will be measured.

Authors' comments: none

Authors' actions: As suggested this information is added on page 6&7:

"…circulating pro- and anti-inflammatory cytokines (including tumour necrosis factor (TNF-α), interleukin-6 (IL-6), interleukin-1β (IL-1β), interleukin-10 (IL-10), and interleukin-1 receptor antagonist (IL-1RA)) by using simultaneous Luminex assays."

"In addition, danger-associated molecular patterns (DAMPs, including high mobility group box 1 (HMGB1), heat shock protein-70 (HSP70), calgranulin-C (S100A12), calprotectin (S100A8/9), nuclear DNA, and mitochondrial DNA) will be measured as reported previously [38], in plasma centrifuged for a second time at high speed (16000 g)."

VERSION 2 – REVIEW

REVIEWER	Berger, Miles Duke University Hospital, Anesthesiology
REVIEW RETURNED	13-Feb-2021

GENERAL COMMENTS	Interesting study and well written paper, which should be accepted. The following comments are additional items they may wish to address: COuld the authors add a reference for the DOS delirium assessment, and describe how they will measure Inter-rater reliability for the delirium assessments performed by their team? POCD criteria #2 "when the overall weighted performance score deteriorated in more than one domain at hospital discharge compared to baseline. Accordingly, all patients will be dichotomized into two groups: with or without POCD at hospital discharge". …..if there are 5 domains, a lot of patients will randomly have decline in 2+ domains post surgery. How much of a decline in these domains will be required as a threshold to meet this criterion #2 for POCD? How will this address both sensitivity and specificity? Merely finding patients that decline in 2 domains (and improve in 3 others) following surgery is likely to pick up a lot of false positives, i.e. patients who do not really have POCD, and just have a random slight decline in 2 domains, if no thresholds are used here for the amount of decline. Also the authors will likely have more statistical power to detect relationships between biomarkers (ie imaging or blood based) and cognition if cognition is analyzed as a continuous outcome (mean of the domain scores) rather than as a dichotomous trait. This is relevant since cognitive function is a continuously distributed variable among humans- the population distribution of cognitive function is not a bimodal distribution consistent with two discrete different groups (i.e. impaired and non impaired). Multiple linear regression can be applied to correct for possible confounding factors- for which confounders? How specific is this PET tracer? Does it bind to activated peripheral blood monocytes? If so, then couldn't the signal seen in the brain reflect either infiltrating activated peripheral blood monocytes, or resident microglia or astrocytes? HOw will the data be interpreted if this ligand thus lacks cellular specificity?

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Dr. Miles Berger, Duke University Hospital

Comments to the Author:

Interesting study and well written paper, which should be accepted. The following comments are additional items they may wish to address:

1. Could the authors add a reference for the DOS delirium assessment, and describe how they will measure Inter-rater reliability for the delirium assessments performed by their team?

Authors' comments: We added references for the DOS and CAM-ICU delirium assessments. Screening of delirium is standard of care during hospitalisation and performed three times daily. All nurses are trained and experienced in these screening assessment methods. Given the previously reported inter-rater reliabilities with high overall percent agreement [Mueller et al. J Nurs Measurement 2017], we will not measure this specifically in this cohort, since this is not our primary study outcome.

Authors' actions: We added references for the DOS and CAM-ICU delirium assessments on page 7.

2. POCD criteria #2 "when the overall weighted performance score deteriorated in more than one domain at hospital discharge compared to baseline. Accordingly, all patients will be dichotomized into two groups: with or without POCD at hospital discharge". …..if there are 5 domains, a lot of patients will randomly have decline in 2+ domains post surgery. How much of a decline in these domains will be required as a threshold to meet this criterion #2 for POCD? How will this address both sensitivity and specificity? Merely finding patients that decline in 2 domains (and improve in 3 others) following surgery is likely to pick up a lot of false positives, i.e. patients who do not really have POCD, and just have a random slight decline in 2 domains, if no thresholds are used here for the amount of decline.

   Authors' comments: To prevent diagnosing false positives as POCD, we use the clinical classification based on normative data. A decline in a cognitive domain is defined by a deterioration in clinical classification from baseline. So patients who scored "normal performance" (above -1 SD from the normative mean) at baseline and dropped to a "below average" (between -1 SD and -1.65 SD from the normative mean = lowest 16% of the population) or to "impaired" (below -1.65 SD from the normative mean = the lowest 5%) postoperatively are considered to have deteriorated on that specific cognitive domain. Similarly, if a patient already scored "below average" at baseline and deteriorated postoperatively to "impaired", this is considered a postoperative decline on this domain. If such deterioration is present in more than one domain this fulfils the POCD criteria. This classification method has been used previously in other neuropsychological studies, and is well applicable for POCD definition. These previous studies are referred in the manuscript [van den Berg et al., JNNP 2005; Reukers et al., BMC Inf Dis 2020]

   Authors' actions: To clarify this better we added the following on page 9:

   "We will define the presence of POCD when 1) patients are impaired in one or more domains at hospital discharge, compared to baseline, or 2) when the overall weighted performance score deteriorated (from classification score 0 or 1 to 1 or 2, respectively) in more than one domain at hospital discharge compared to baseline."

3. Also the authors will likely have more statistical power to detect relationships between biomarkers (ie imaging or blood based) and cognition if cognition is analyzed as a continuous outcome (mean of the domain scores) rather than as a dichotomous trait. This is relevant since cognitive function is a continuously distributed variable among humans- the population distribution of cognitive function is not a bimodal distribution consistent with two discrete different groups (i.e. impaired and non-impaired).

   Authors' comments: Thanks for this suggestion! It was already in our manuscript, but could be easily overseen and was not so specific, so we emphasized this better now in the revised manuscript.

   Authors' actions: On page 9 we changes the following:

   "~~*Multiple linear regression models will be used to study the relationship between tracer uptake and neuropsychological outcomes."*~~ INTO:

"*Additionally, the (change in) mean of the cognitive domain scores is analysed as continuous dependent outcome, with (change in) TSPO expression and (change in) systemic inflammatory markers as predictors in multiple linear regression models.*"

4. Multiple linear regression can be applied to correct for possible confounding factors- for which confounders?

   Authors' comments: We removed this sentence, since possible confounding factors are already mentioned two sentences earlier. Age, sex, Clinical Frailty Scale, Charlson Comorbidity Index, Hospital Anxiety and Depression Scale (HADS), the RAND-36 item health survey at inclusion, and newly developed structural brain lesions can be included as covariates. Pre-existent frailty, comorbidity, anxiety, depressive symptoms, or quality of life could affect cognitive performance in this study and are therefore potential confounders. Likewise, patients who develop perioperative microbleeds, lacunar or territorial infarcts, could show more postoperative neuroinflammation or worse cognitive performance, and this should be taken into account in our analyses. The study power does not allow inclusion of all these potential confounders in our regression models. Therefore, we will test whether differences exist between the two distinct groups (POCD yes/no) or whether correlations exist between cognitive performance or TSPO expression and these markers to identify relevant confounders.

   Authors' actions: We removed this sentence on page 10 to prevent repetition of potential confounders mentioned at page 9:

   ~~"*Multiple linear regression can be applied to correct for possible confounding factors.*"~~

5. How specific is this PET tracer? Does it bind to activated peripheral blood monocytes? If so, then couldn't the signal seen in the brain reflect either infiltrating activated peripheral blood monocytes, or resident microglia or astrocytes? How will the data be interpreted if this ligand thus lacks cellular specificity?

   Authors' comments: Indeed, there is a lot of debate on the cellular and functional interpretation of TSPO PET neuroimaging. TSPO expression is not specific to microglia and astrocytes, and the measured PET signal can be driven by other factors such as recruitment of peripheral monocytes to the brain tissue, adherence of circulating leukocytes to the vascular epithelium, and expression of TSPO in neurons or vascular endothelial cells. This has been added to the limitation section of the manuscript. Although this technique has its limitations, there is currently no better alternative available to image neuroinflammation in humans non-invasively, and it has shown clear differences between healthy subjects and diseased controls [Bradburn et al. Ageing Res Rev 2019; Kreisl et al. Lancet Neurology 2020; Zurcher et al; Neuroimage Clin 2015; Lavisse et al. Parkinsonism Relat Disorder 2020] and showed associations with disease activity [Kreisl et al; Brain 2013], indicating that a potentially important pathophysiological hallmark is being assessed.

   Authors' actions: We added the following as limitation to our manuscript:

   "Unfortunately, TSPO expression is not specific to microglia and astrocytes, and the measured PET signal can be affected by recruitment of peripheral monocytes to the brain, adherence of circulating leukocytes to the vascular epithelium, or TSPO expression in neurons or vascular endothelial cells."