PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Diagnosis of the cause of death without autopsy: can virtual autopsy with postmortem CT improve clinical diagnosis?
AUTHORS	Sonnemans, Lianne Kubat, Bela Prokop, Mathias Klein, Willemijn

VERSION 1 – REVIEW

REVIEWER	Kunihiro Inai Division of Molecular Pathology, Department of Pathological Sciences, Faculty of Medical Sciences, University of Fukui, Japan
REVIEW RETURNED	02-Aug-2017

GENERAL COMMENTS	Sonnemans et al. describe the manuscript regarding the diagnostic accuracy of postmortem CT focused on the immediate cause of death. The content of this paper is valuable for the well-understanding of postmortem CT, especially in the patients with hospital autopsies. However, this reviewer considers that several revises will be needed for accepting this article. Major comments 1. From previous manuscripts, the specificity of diagnostic accuracy of postmortem CT is depended on the background of the cadavers such as about 30% accuracy of non-traumatic out-of-hospital deaths, 70% of in-hospital deaths, and >80% of traumatic deaths. In this research, the study cohorts were involved in both in- and out-of-hospital deaths including trauma, the frequency of intrinsic and extrinsic deaths, and/or the incidence of pre-existing diseases (underlying causes of death) in the study population may affect the diagnostic accuracy. So, the authors should make a table of patient profiles including the total number of in- and out-of-hospital death and details of their pre-existing diseases. 2. Several studies use the different research protocol for diagnostic accuracy of postmortem CT. For example, Inai et al. permit to see antemortem medical records in both pathologists and radiologists during their diagnosis, whereas Roberts et al. completely hide the cadaver information for the diagnosis of postmortem imaging. Interestingly, radiologists correctly diagnosed two arrhythmias in this study. However, physiological events are usually impossible to recognize from the postmortem CT images without any supports of antemortem clinical records. Nevertheless, no information is given whether the doctors for postmortem analyses could access the

	medical records or not in this manuscript. Thus, the author should disclose the information in the Materials and Methods. 3. In table 2, the authors classified the diagnostic accuracy of the immediate cause of death before and after postmortem CT. In the table 2, the author describes the malignancies including breast cancer and esophageal cancer as the immediate cause of death. Cancer itself doesn't belong to the immediate cause of death but should be involved as the underlying cause of death (namely, pre-existing disease). If the authors persist in cancer as the immediate cause of death, they should describe it as cachexia due to breast cancer, because cachexia represents the condition of real cancer death. If the patient was dead by liver failure, and/or multi-organ failure due to metastases, the authors have to describe as liver failure due to liver metastases of breast cancer, or multi-organ failure due to multiple metastases of esophageal cancer. 4. Although myocardial infarction and pulmonary embolism are notoriously difficult to diagnose by postmortem CT, some of these patients were seemed to be diagnosed by postmortem imaging. These results are very amazing and impressive. Unfortunately, the figures were just given postmortem CT images only. Because all of the study cohorts in this study were performed both hospital autopsy and postmortem CT, macroscopic and microscopic views of autopsies should be added in the figures from the viewpoint of pathology-radiology-correlation. In addition, this reviewer would like the authors to make additional figure including the postmortem CT, macroscopic and microscopic photos of the case of myocardial infarction exactly diagnosed by postmortem CT. Consequently, the quality of this manuscript would become superior to the initial version. Minor revision 1. In the 2nd line of the Discussion, the author described as "using conventional autopsy as the reference standard." This sentence may be removed in the Materials and Methods.

REVIEWER	Dominic Wichmann University Medical Center Hamburg-Eppendorf, Germany
REVIEW RETURNED	22-Aug-2017

GENERAL COMMENTS	Dear Lianne, To summarize I could get 90% of the information from your paper at first sight. The rest might be clear to you and your colleagues but for me as an external it was not "self explaining" because I was not involved in the conception of the study. I hope I can help to improve the manuscript with my comments. The result section line 181 -190 is difficult to read and needs clarification. If I understand you correctly I would propose something like: Clinicians were not able to assign an immediate cause of death in 29 cases (34%), virtual autopsy was not able to do so in seventeen cases (20%) and in 10 cases (12%) no immediate cause of death was found by any modality. What is the definition for "immediate cause of death" and "correct

	cause of death". Without stating this it is hard to understand why correct cause of death is 53% when clinicians are not able to identify the immediate cause of death in 29 cases (34%) what would result in the 66%. This may cause confusion in countries with different practices for death certificates, e.g. coroner system in Anglo-American countries or the clinician system in Germany. We don't have this to definitions, but of cause I agree the a correct cause identified by an autopsy has a higher value than the "clinical cause". Line 211 a radiologist is not able to diagnose an arrhythmia, This is only possible by ECG. A cardiac fibrosis may let suspect to an arrhythmia but can't "prove" it. If I have read your manuscript correct you concentrate only on causes of death. In this context your statement in line 284 may be correct, but we have also investigated the value of PMCT +/- angiography in a prospective study. The improvement by adding PMCT is one of the major results (Wichmann et al 2014 Ann Int Med) Table 2 Your grouping by type of pathology is good. The table itself I find a bit confusing, especially if all correct and incorrect diagnoses are given. May be you could modify it Table 3 To be honest I have difficulties to understand the calculation for sensitivity and specificity. What is the numerator / denominator? How are the calculations made.

REVIEWER	Tomohiro Shinozaki The University of Tokyo, Japan
REVIEW RETURNED	25-Sep-2017

GENERAL COMMENTS	As a statistical reviewer, I assessed statistical aspects of the manuscript "Diagnosis of the cause of death without autopsy: can virtual autopsy with postmortem CT improve clinical diagnosis? " by Sonnemans and others. Overall, I found the manuscript appropriately analyzed the data and presented their results; I have a few additional requests that I believe will improve the readability of the paper. 1) Results (p. 9, l. 181): They presented the frequencies of failure of COD assignment (n = 29 before PMCT, 17 after that, and 10 after autopsy). Whether COD is assigned or not by each procedure seems to have a critical implication in this study, I recommend using a figure to illustrate the "flow" of COD assignment according to the study flow. Please refer the attached file as an example. 2) Results (p. 9, l. 191-) and table 1: As authors correctly chose the McNemer's test, the data has matched-paired nature when comparing the performance of before vs. after PMCT assessments. I prefer to present 2-by-2 tables for each outcome: that is, each cadaver is classified into 4 categories by correct/incorrect diagnosis before/after PMCT. These tables enable readers to immediately understand the actual numbers of agreement/disagreement of diagnoses, which also make the results transparent. I also prefer to quantify the improvement by PMCT; the

	authors can present not only p-values (by McNemer's test) but also Mantel-Haenszel odds ratio (also calculated by the ratio of "discordant" pairs) or Mantel-Haenszel risk difference, along with 95% confidence intervals. 3) Results (p. 10, l. 199-): Throughout the results, they presented only percentages of characteristics without numerator/denominator, which may make it difficult for readers to understand what each proportion means. The following percentages are better to be accompanied by these numbers: - "73%" in line 203 is "11/15 = 73%" from table 2? - Instead of "In the group of perfusion disorders, all pulmonary embolisms were..." (line 210), "In the group of perfusion disorders, all pulmonary embolisms (n = 3) were..." may be easy to read. - "44%" in line 216 is "7/16 = 44%"? - What is the cases in "71% of these cases" in the same line? - "85%" in line 222 is 11/13? 4) Results (p.11, l. 234-) and table 3: (a) The above comment 3 is applied to this table: actual numbers may help readers understand the meaning and precision of percentages. (b) Most readers (including I) may not be able to replicate the calculation of sens/spec percentage based on table 2. Please clarify the classification of types of pathology (A) and anatomical systems (B). (c) p-values should be precisely presented even if they are over 5% level (eg, p = 0.56); see, the ASA statement on p-values (http://amstat.tandfonline.com/doi/abs/10.1080/00031305.2016.1154108). Also, please specify the statistical test used (again, McNemer's test may be an appropriate choice). (d) I doubt the value of reporting specificity in this context with multiple (4-fold) choice of diagnosis. Please clarify the implication of high specificity. (e) How much correct specification of types of pathology and anatomical systems, rather than COD itself directly evaluated thus far, is useful in real settings? 5) I cannot understand the aim and specified models for logistic regression (l. 240-). The detailed description of the analysis is needed, and please show whole results including (adjusted?) odds ratios (95% CI) in addition to p-values.

REVIEWER	Angie Wade University College London, UK
REVIEW RETURNED	26-Sep-2017

GENERAL COMMENTS	This study compares virtual autopsy with clinical diagnosis. There are several points to address before proper interpretation can be made of this data. In particular, the results are too dependent on statistical significance of quite small samples. The discussion must be rewritten taking into account the imprecision of results (for example, stating that PMCT only moderate added value for infections compared to usefulness for cardiac causes – this is not proven – see point 5 below). If statements are to be made regarding the relative usage in different pathologies and anatomical systems then these comparisons must be made formally via suitable significance tests, with differences presented with confidence intervals). Other points are given below: 1. The percentage of eligible deaths available (2155) who were included in the analyses was relatively low (77/2155). A further 9

	deaths were added who were not clinically examined at the hospital or by the hospital emergency service were also added. For this latter group, what is the denominator of all possible such cases that could theoretically have been included if they had undertaken the necessary investigations? 2. Presumably the relatively low inclusion rate was due to non-consent of autopsy and this is unavoidable. However, the authors should clarify the consent process. Were all eligible approached for consent (and if not, how were they selected)? Of those approached for consent, how many refused and are any reasons recorded for refusal? Is there any information available at all about the potential biases of those included in the analyses? For instance, in their clinical histories or their available scans? It is important to ascertain the extent to which the results can be generalised and this is dependent on the representativeness (or not) of the sample analysed. 3. What was the rationale for having those with minimal previous experience analysing the PMCT images? Presumably this infers a lower bound to the sensitivity that may be attained and the extent to which PMCT improves diagnosis? 4. The interval between death and PMCT should be reported more fully ie. not just the mean. I would imagine the distribution of times to death would be skew and so the median would be a more valid summary measures. Some indication of the range of times should also be given, - maybe the minimum and maximum times, plus the standard deviation or 25th/75th centiles dependent on whether mean or median is best summary measure. 5. Summary percentages are given with confidence intervals and this is helpful. However confidence intervals should also be given for the differences in percentages (tables 1 and 3 and associated text). For table 3 in particular, some of the percentages are based on very small samples and so the confidence intervals will be wide. It is not surprising that these are non-significant given the sample sizes, but the confidence limits will show whether the data remain compatible with quite large and clinically important differences in either sensitivity or specificity. For example, there is a difference of 23% in sensitivity for haemorrhage (69% to 92%) which is dismissed as NS, whereas the smaller difference of 20% for perfusion disorder is highlighted as significant and important. 6. CA needs defining (table 1). 7. Table 3 and associated text: p-values must be given (use of NS is unacceptable). 8. Logistic regression: How was radiologist experience measured? The text at the bottom of page 11 relating to these regressions seems to suggest that it is time rather than individual radiologist experience that is assessed for association with correct diagnosis. The Odds Ratios should be given with confidence intervals and the limits of these used in interpretation

VERSION 1 – AUTHOR RESPONSE

Response to reviewers

Reviewer 1

Major comments

1. We included a table of patient profiles (Table 1).
2. Radiologists and pathologists had full access to medical records. This is additional information is added to the Materials and methods section.
3. Correct point. We have adjusted these underlying causes of death into the immediate causes of death. (Table 4)
4. Myocardial infarction was correctly diagnosed after PMCT in seven cases. In five of those seven cases, myocardial infarction was not visible on PMCT and diagnosis was based on clinical findings in the absence of (other) significant findings on PMCT. This is now more clearly mentioned in the results section. The two other cases where imaging was suspect for myocardial infarction are described in the main text as no macroscopic photos were made during autopsy. A microscopic picture of a coronary artery thrombus is, in our opinion, not very contributory.

Minor revision

1. Revised.

Reviewer 2

• Revised results section line 181-190. Figure 1 is added for clarification.

• Immediate cause of death is now explained in the introduction section. With correct cause of death we meant a correctly diagnosed cause of death, or in other words, in accordance with autopsy (reference standard). Without PMCT, clinicians assigned a cause of death in 57/86=66%. However this was not always the correct diagnosis, as given by autopsy. Only in 40 of the 76 cases in which autopsy was able to assign a cause of death, the cause of death was correctly diagnosed by clinicians before PMCT (40/76=53% sensitivity). In 57-40=17 cases, the cause of death as diagnosed by clinicians was in disagreement with the cause of death at autopsy. I think the addition of Figure 1 and Table 3A will also help here for a better understanding.

• Correct. I have expanded the explanation describing that cardiac arrhythmia was the most likely based on the secondary symptoms observed in combination with the absence of other significant findings on imaging.

• This study investigated the diagnostic performance of postmortem CT angiography compared to traditional autopsy rather than investigating the improvement of the clinical performance by use of PMCT with traditional autopsy as reference standard. I revised the sentence in '.. of unenhanced PMCT compared to ..' rather than only '.. of PMCT compared to ..'.

• I deleted the incorrectly diagnosed causes of death to in order to improve the readability of Table 4.

• Sensitivity for infectious type of pathology = (number of cases where the type of pathology involved was correctly assigned as infectious) / (total number of cases with an infectious cause of death at autopsy). The denominator is given in the first column.

Specificity for infectious type of pathology = (number of cases where the type of pathology was correctly assigned as non-infectious) / (total number of cases without an infectious cause of death at autopsy). The denominator is 78 minus the number given in the first column, so for infections this is 78-26=52.

Reviewer 3

1. We included a flowchart. See Figure 1.
2. 2-by-2 tables were included. See Table 3A-C.
3. Numerators and denominators were included in the main text.

4. The denominators for sensitivity are given in the first column.
   This classifications of subgroups represent the most common types of pathology and most affected anatomical systems.
   Implication of high specificity -> there are less false-positives so PMCT can be used to identify pathologies rather than excluding them, as the sensitivity is lower than the specificity.
   I don't understand comment (e).
5. We revised the logistic regression analysis in a more detailed description including odds ratios with 95% confidence intervals.

Reviewer 4

1. It is impossible to determine the denominator of all possible cases, because these 9 cases deceased at home and were brought to the hospitals mortuary for PMCT and autopsy examination. All the deceased in the local area can be brought to our hospital for PMCT and autopsy examination when this is desired by the relatives. Furthermore, there are no strict borders delineating the area covered by our hospital. I have revised the text to make this more clear.
2. Clinicians had to ask consent for PMCT and autopsy in all cases of death. We added this information in the methods section. In the beginning of the discussion we explained that the reason for the low consent was not investigated as motives for performing or not performing a PMCT-scan were not extensively documented. Nonetheless, some clinicians mentioned that they only requested for PMCT in case of refusal of conventional autopsy. As mentioned in the 'strengths and limitations section', the low consent rate for autopsy and PMCT probably resulted in a selection-bias towards patients with diagnostic difficulties or unresolved issues, resulting in an underestimation of the diagnostic performance compared to more general causes of death.
3. This is a misunderstanding. We did not prefer radiologists with minimal previous experience in reading PMCT images, but this was unavoidable as postmortem imaging is a relatively new field of expertise. We revised the text in the 'materials and methods' section.
4. Revised into median + interquartile range.
5. Confidence intervals are added in Table 5.
6. CA = conventional autopsy. Added in the caption of the Table.
7. Revised.
8. Indeed, time related; the number of years of experience in reading postmortem CT images. Odds ratios and confidence intervals are given.

VERSION 2 – REVIEW

REVIEWER	Tomohiro Shinozaki The University of Tokyo
REVIEW RETURNED	02-Nov-2017

GENERAL COMMENTS	The paper was improved a lot but a few points remained to be corrected. - Comparison measure and CI should be included in addition to p-values in Table 2. For example, Mantel-Haenszel probability ratio is easily calculated by Table 3. Below I show the calculation by taking example of COD: (1) a point estimate = 49/40 = 1.225 (this is exactly the same as 64%/53%); (2) log of point estimate = ln(1.225) = 0.203; (3) Standard error of log of point estimate = sqrt([4 + 13] / [49 * 40])

	= 0.0931; (4) 95% confidence limits in log-ratio scale = 0.203 -/+ 1.96*0.0931 = 0.021, 0.385; (5) convert into ratio scale: exp(0.021) = 1.02, exp(0.385) = 1.47. In the main text around line 201, this result may be mentioned as "The number of correctly identified causes of death after performing a PMCT scan (64% [95% CI 53%-75%]) increased from before the scan (53% [95% CI: 41%-64%]) by 1.23 times (95% CI 1.02-1.47; p=0.049 by McNemar's test)." You can find the formula in intermediate/advanced epidemiology textbooks (eg. Modern Epidemiology, third ed). - The order of Tables 2 and 3 should be reversed. - I cannot still understand what logistic regression modeled. What are the "radiologists' improvement in reporting PMCT-scans over the four years of initial experience" and "each year of experience". Please clearly specify the outcome variable and explanatory variable(s).

REVIEWER	Angie Wade University College London, UK
REVIEW RETURNED	12-Nov-2017

GENERAL COMMENTS	Although the authors have addressed many of my concerns, they have still not given confidence intervals for the differences between pre and post CT and between different systems. These are crucial to any discussion of differences observed and their potential importance. In particular, the headline increase in sensitivity for overall diagnosis based on p=0.049. The confidence interval for the difference will show the data to be compatible with very small differences as well as more meaningful values. The footnote related to the definition of NS should be removed from table 5.

REVIEWER	Dominic Wichmann University Medical Centre Hamburg-Eppendorf Department of Intensive Care Medicine Martinistr. 52 20246 Hamburg Germany
REVIEW RETURNED	13-Nov-2017

GENERAL COMMENTS	To the editorial office of BMJ Open, Thank you for letting me participate in the review process of the manuscript "Can virtual autopsy with postmortem CT improve clinical diagnosis of cause of death? A retrospective observational cohort study in a Dutch tertiary referral centre." (BMJ Open 2017-018834). The manuscript of Sonnemans et al. has substantially improved after revision. All my comments have been addressed in a satisfying way. Please correct a little typo in line 430: „ATherefore…" Sincerely Yours Dominic Wichmann

REVIEWER	Kunihiro Inai Division of Molecular Pathology, Faculty of Medical Sciences, University of Fukui
REVIEW RETURNED	13-Nov-2017

GENERAL COMMENTS	All my previous queries and comments has been properly resolved. Thus, I consider that the new version of the manuscript from Sonnemans et al. is now suitable for publication. No additional changes are needed.

VERSION 2 – AUTHOR RESPONSE

Response to reviewer comments:

Reviewer 2

- Corrected.

Reviewer 3

- I have included the differences in sensitivity and specificity before and after PMCT was performed, including 95% confidence intervals as recommended also by reviewer 4.
- I placed Table 3 in front of the tables which present the sensitivities/specificities.
- 'Radiologists' improvement in reporting PMCT-scans over the four years of initial experience' is the improvement in sensitivity for immediate cause of death/ type of pathology/anatomical system by radiologists reporting the PMCT-images. As all radiologists had minimal previous experience in interpreting PMCT-images, we wondered whether the number of correctly made diagnoses at PMCT increased with (and could be explained by) the experience in reporting PMCT-images. The experience was expressed as the time interval from the moment of start of the study (July 2012).

Reviewer 4

- I have replaced the McNemar p-values by 95% confidence intervals of the observed differences as requested. Indeed, I think these values are more meaningful for interpretation of the results.
- Footnote table 5 is corrected.

VERSION 3 – REVIEW

REVIEWER	Tomohiro Shinozaki The University of Tokyo, Japan
REVIEW RETURNED	30-Dec-2017

GENERAL COMMENTS	Including 95% confidence intervals instead of reporting just p-values is good, but their calculation is simply incorrect because the authors naively assumed independent binomial sampling for sensitivities before/after PMCT: Confidence intervals for risk difference estimates should incorporate within-patient correlation. I suggested to use Mantel-Haenszel risk ratio in previous comment, though its risk difference is probably a better alternative. Why did the authors use

	this simple stratification technique? Applying Mantel-Haenszel difference, for example, difference in sensitivities of immediate cause of death before (53%, 95% CI 41% to 64%; this CI is correct) and after PMCT (64%, 95% CI 53% to 75%, this is also correct) is 11.8% with 95% CI of 1.6% to 22.1%, which excludes null value and is consistent with formerly provided appropriate McNemar's test.

VERSION 3 – AUTHOR RESPONSE

Dear editor,

It is a great honour that our manuscript has been selected for publication. The comment of reviewer 3 was very useful as we used independent binomial sampling instead of dependent binomial sampling. Once again, thank you for time. I am looking forward to your response.

Sincerely,
Lianne Sonnemans

VERSION 4 – REVIEW

REVIEWER	Tomohiro Shinozaki The University of Tokyo, Japan
REVIEW RETURNED	22-Jan-2018

GENERAL COMMENTS	Thank you for addressing my comment. I confirmed that the provided values for confidence intervals in main results were correctly calculated by Mantel-Haenszel difference. Please specify the method anywhere, eg, "Mantel-Haenszel difference with Sato's variance estimator (Biometrics 1989:45;1323-4)" so that readers can replicate the results (though I do not know the actual procedure the authors employed in this revision).