HIGH SCHOOL EDITION
Journal of Student Research
Volume 12 Issue 3 (2023)

Data Analysis of Interpretation Errors Made When Reading CT Angiograms of the Head and Neck

Yana Shah

Mills E. Godwin High School

ABSTRACT

A medical error is an adverse outcome of medical treatment that occurs more often than it should. When reading CT angiograms (CTAs) of the head and neck, radiologists receive detailed coverage from the aortic arch to the frontal sinus, along with information about targeted blood vessels. However, these large scans can also result in missed detection of incidental findings and anomalies. This observational study aimed to determine the frequency of missed incidental findings during the examination of CTAs of the head and neck and to identify the main reasons radiologists fail to report them. Five volunteer radiologists were each randomly assigned 50 previously read cases and used the RADPEER system to score the original report. In addition, if the case received a RADPEER score other than 1, an error classification (EC) score was assigned. It was hypothesized that there are many errors made by radiologists when reading CTAs of the head and neck and that in cases with missed findings, the main reason for the error is that the discrepancy was overlooked. The results of the study showed that 29.6% of the cases had an error, and of those cases, 86.5% had an EC score of 1, supporting both research hypotheses. Two chi-square tests were also performed, both of which were statistically significant. The significant number of errors could be due to satisfaction of search and the limited availability of subspecialty radiologists available to read complex cases in private practices, such as the one studied.

Introduction

Computed tomographic angiography (CTA) is a medical test that allows for the diagnosis of abnormalities involving blood vessels throughout the body. Through great technological advancement, this procedure has replaced digital subtraction angiography (DSA) for diagnostics, as it is quicker and non-invasive with fewer potential complications. Aside from recognizing problems with targeted vessels, CTAs are known for their ability to detect incidental findings (IFs), which are abnormalities within the anatomic coverage that are not related to the reason for examination (Chen et al., 2020). Because radiologists do not expect these findings, they are often overlooked which could possibly affect a patient's course of treatment.

Some indications for CTAs of the head and neck include, but are not limited to, carotid stenosis, brain aneurysms, arteriovenous malformations, and vascular injuries. While the purpose of CTAs of the head and neck is to provide high-resolution imaging of the targeted blood vessels, the scan provides full coverage of the anatomy from the aortic arch to the frontal sinus ("CT Neck/Brain Angiography", n.d.). Furthermore, these scans capture a detailed view of soft tissues and osseous structures, allowing for the diagnosis of IFs ranging from little clinical significance, such as a small thyroid nodule, to notable clinical significance, like the early detection of a tumor. Clinically significant findings can lead to additional medical care including other diagnostic procedures and treatments which can be very important to the patient's health. Missing significant clinical findings can delay diagnosis and treatment, which could be very harmful to the patient's health. (Lumbreras, 2010).

According to a study by Johns Hopkins University, over 250,000 people in the United States alone die from medical mistakes every year (Sipherd, 2018). In order to reduce this number, the American College of Radiology

(ACR) developed the RADPEER program, which is a system used by radiologists around the country to peer review their colleagues' interpretations and determine if any findings were missed within the original report ("RADPEER", n.d.). When first created in 2002, the reports were scored on a 4-point scale, in which one or two meant that there were no errors or an understandable error, and three or four meant that the error in the report should have been recorded in the original interpretation. As this approach gained widespread acceptance across the country, many doctors began to express frustration with the scoring system, believing that it left too much room for opinion to influence the true score. By 2016, the system had undergone two changes before the ACR finalized the scale. Now, all radiologists use a one to three scales when peer reviewing, one meaning "concur with interpretation", two meaning "discrepancy in interpretation/not ordinarily expected to be made (understandable miss)", and three meaning "discrepancy in interpretation/should be made most of the time". Furthermore, if a report is given the number two or three, it is also given the letter 'a' or 'b', 'a' meaning the miss is unlikely to be clinically significant, whereas 'b' represented a miss that is likely to be clinically significant (Goldberg-Stein et al., 2017).

Along with identifying errors in previous reports, peer-reviewing radiologists are trained to associate the discrepancy with a reason as to why it occurred. Diagnostic radiological errors can happen for a number of reasons such as human error, technical issues, and system flaws (Murphy, 2021). There are two major classification systems used to determine why an error was made, the Renfrew classification and the Brook classification. The Renfrew classification was proposed by Renfrew in 1992, but the one used today was revised by Kim and Mansfield in 2014. There are 12 different categories from type 1 to type 12 in this system, each one proposing a unique reason as to why the radiologist may have missed the finding (Kim & Mansfield, 2014). The Brooks classification system, on the other hand, only has four categories: latent errors, active failures or human error, external causes, and customer causes. The main difference between the two classifications is that the Brooks classification takes human error into greater consideration. Many radiology practices modify one of these two classification methods, creating their own classification system to determine the reason for the error (Murphy, 2021).

The purpose of this study is to determine the frequency of missed incidental findings during the examination of CTAs of the head and neck and to identify the main reason that radiologists are failing to report them. This should aid radiologists in minimizing misinterpretation errors and physicians in making clinical decisions. To determine if there are mistakes within the original interpretation, the peer-reviewing radiologist will be using the RADPEER scoring system: 1, 2a, 2b, 3a, and 3b, and will be using a modified version of the Renfrew classification system (8 error classification categories) to determine the reason for the error. There is no control because this is an observational study. It is hypothesized that there are a great number of errors made by radiologists when reading CT angiograms of the head and neck, and of the cases with a missed finding, the main reason for the miss is because the discrepancy was overlooked. This is because many studies similar to this one, such as one by Chen et al. written in 2020, found that over 5% of the cases studied had some sort of discrepancy, which is higher than expected by the ACR Committee, and because in 2014 Kim and Mansfield predicted that over 50% of radiological errors are caused by a missed finding (Goldberg-Stein et al., 2017).

Procedure

Before the study was conducted, pre-approval was obtained from the administrative leadership team at the teleradiology practice the cases were selected from. Two-hundred fifty Computed Tomography Angiograms (CTA's) originally interpreted between the dates January 1, 2020, and January 1, 2022, were selected at random by the mentor of this experiment. Additionally, all scans were of patients over eighteen. After being selected the cases were randomly divided into five groups (fifty cases in each group), and each group was assigned to one of the five radiologists who volunteered to participate in this study. The radiologists who volunteered were all certified by the American Board of Radiology and had over 10 years of experience. Each of the radiologists was given demographic information including sex and age at the time of the scan along with access to the original report.

To begin, the interpreting radiologists reviewed their assigned CTA scans and the original report and used the RADPEER system to score the original report with a score of 1, 2a, 2b, 3a, or 3b. Furthermore, if the case was given a RADPEER score other than 1 (no discrepancy), then an error classification (EC) score was assigned. There were eight set categories (type 1 to type 8 - see appendix) created prior to the experiment based on the Renfrew classification system. The data was collected in a raw data table, and mathematical and statistical analysis was performed.

Results

For this observational study, data analysis of the RADPEER scores of the CT angiograms and the error classification (EC) scores of the cases with an error was performed, and the results are shown in table 1, table 2, table 3, graph 1, and graph 2. Two research hypotheses were created, the first stating that there are a great number of errors made by radiologists when reading CT angiograms of the head and neck, and the second stating that of the cases with a missed finding, the main reason for the miss is because the discrepancy was overlooked. The central tendency of mode was determined for the RADPEER score (1) and EC score (1). This implies that the majority of the cases did not have any errors and that of those with errors, the most common reason for an error was simply because a finding was missed. Further analysis of the RADPEER score data showed that 176 (70.4%) of the cases had a score of 1, 39 (15.6%) of the cases had a score of 2a, 21 (8.4%) of the cases had a score of 2b, 8 (3.2%) of the cases had a score of 3a, and 6 (2.4%) of the cases had a score of 3b. This means that 29.6% of the cases had some sort of error (scores 2a, 2b, 3a, or 3b), leading to the conclusion that the research hypothesis can be supported as there are a significant number of errors made when radiologists read CT angiograms. A deeper review of the EC scores of the cases with errors showed that 64 (86.5%) of them had a score of 1, 8 (10.8%) of them had a score of 2, 1 (1.4%) of them had a score of 4, 1 (1.4%) of them had a score of six, and 0 (0%) of them had a score of 3, 5, 7, or 8. This further supports the second research hypothesis stating that the majority of the errors were caused by a missed finding (type 1).

A chi-square test was conducted for the RADPEER score data at a level of significance of 0.001 and a degree of freedom of 4. The null hypothesis created for this experiment stated that there would be no difference between the observed and expected values for the RADPEER scores of the selected cases. The calculated chi-square value for the observed vs. expected (540.265) was significantly higher than the critical chi-square value of 14.860. This draws the conclusion that the null hypothesis can be rejected and there is a significant difference between the observed and expected values. The probability that the results of this procedure were due to chance is less than 0.001, implying that there is a reason why the radiologists are making these errors and it is not simply due to human error. This data set is shown to be statistically significant, and there are a significant number of errors made by radiologists when reading CT angiograms.

A chi-square test was also performed for the EC score data at a level of significance of 0.001 and a degree of freedom of 7. The null hypothesis created for this experiment stated that there would be no difference between the observed and expected values for the EC scores for the cases with a discrepancy. The calculated chi-square value for the observed vs. expected (22.534) was higher than the critical chi-square value of 20.278. This allows for the null hypothesis to be rejected and means that there is a significant difference between the observed and expected values. The probability that the results of this procedure were due to chance is less than 0.001, implying that there is the main reason for error for the discrepancies. Lastly, this data set is also shown to be statistically significant, meaning there is most likely a correlation between the presence of an error and the reason for the error.

Conclusion

The purpose of this study was to determine the frequency of missed incidental findings during the examination of CTAs of the head and neck and to identify the factors that can contribute to radiologists failing to report them. To determine this, 250 previously reviewed CT angiograms were randomly selected and distributed to five volunteering radiologists. The cases were then reviewed and a RADPEER score and EC score was determined. Two research hypotheses were created, the first stating that there are a great number of errors made by radiologists when reading CT angiograms of the head and neck, and the second stating that of the cases with a missed finding, the main reason for the miss is because the discrepancy was overlooked. After analyzing the data, it was determined that although most of the cases were found to contain no errors, there is still a significant amount of discrepancy and the leading cause for an error is due to a missed finding, meaning that both research hypotheses can be supported. To determine the significance of the data, two chi-square tests were performed, one for each dependent variable. Both the RADPEER score observed vs. expected and the EC score observed vs. expected were statistically significant. This suggests that radiologists are making more errors than expected by the American College of Radiology and that their error classification distribution is not as expected by Kim and Mansfield, implying that there is probably a relationship between the presence of an error and its cause (Goldberg-Stein et al., 2017; Kim & Mansfield, 2014).

Over the years, few studies have investigated the number of errors made by radiologists when reading CTAs of the head and neck specifically. One observational study conducted by K. Lian et al. in 2011 that evaluated the benefits of double reading CT angiograms of the head and neck recorded that 4% of cases read had clinically significant findings that were not initially reported. However, the physicians who double-read the cases were of different experience, which could have caused the 4% to be an underestimate of the true number of mistakes within the sample group (Lian et al., 2011). A more recent study by Chen et al. found that 20.91% of the cases he and his colleagues reviewed had one or more clinically significant IFs, which is much higher than the 10.8% found in this study. One difference between these two studies that may contribute to the different percentages could be the number of physicians who reviewed each case. In this study, the case was reevaluated by one randomly assigned physician, however, in the study by Chen et al., all the cases were evaluated by at least two physicians and sometimes even a third to ensure all errors were accounted for (Chen et al., 2020).

There are a number of reasons why radiologists make errors when reading scans, some of which are not directly their fault such as incomplete patient history. According to Kim and Mansfield, however, around 64% of mistakes are due to unreported findings (Kim & Mansfield, 2014). In this study, 86.5% of the cases with incidental findings (both clinically significant and insignificant) were caused by the same reason, which is significantly higher than the predicted value. One reason for this may be due to the idea of satisfaction of search, in which radiologists unintentionally start to overlook other findings in the study after diagnosing or identifying one abnormality. Another possible explanation may be because the institution under study is a private practice, so because there are fewer subspecialty radiologists available, general radiologists are also assigned to read scans like CTAs of the head and neck, which are better read by neuroradiologists because they are more trained to do so (Adamo et al., 2021). While the high percentage in this study is most likely due to the small sample size of this data set, it still supports the idea that the leading cause of mistakes in radiology is due to the incorrect reading of the scans given and should be improved by all radiology departments across the world.

This study did contain a few sources of error that could be improved before performing again. For one, the sample size was somewhat small for the type of data being collected. This could have affected the true error rate of the institution studied and therefore a bigger sample size (between 500 and 1500 cases) should be collected for future papers. The fact that both dependent variables are qualitative is seen to be another source of error. Although every category for both DVs had distinctive characteristics, differences in opinion could still have an impact on the data set. To minimize this, it is recommended that two physicians analyze each case and jointly decide on a RADPEER and EC score. If there is still disagreement, a third physician may offer their opinions to help reach a more conclusive decision. A retrospective double-blind study, in which the radiologists are not given access to the original report and the old and new reports are compared, could be carried out for further research. Additionally, general radiologists and subspecialty radiologists read the same cases in private clinics like this one, therefore the error rates of general radiologists vs. neuroradiologists while reading CTAs of the head and neck should also be evaluated.

Works Cited

Adamo, S. H., Gereke, B. J., Shomstein, S., & Schmidt, J. (2021). From “Satisfaction of search” to “subsequent search misses”: A review of multiple-target search errors across radiology and Cognitive Science. Cognitive Research: Principles and Implications, 6(1). https://doi.org/10.1186/s41235-021-00318-w

Chen, G., Xue, Y., Wei, J., & Duan, Q. (2020). The undiagnosed potential clinically significant incidental findings of neck CTA. Medicine, 99(43). https://doi.org/10.1097/md.0000000000022440

CT neck/brain angiography. Cedars. (n.d.). Retrieved November 27, 2022, from https://www.cedars-sinai.org/programs/imaging-center/exams/neuroradiology/ct-neck-brain-angiography.html

Goldberg-Stein, S., Frigini, L. A., Long, S., Metwalli, Z., Nguyen, X. V., Parker, M., & Abujudeh, H. (2017). ACR RADPEER Committee White Paper with 2016 updates: Revised scoring system, new classifications, self-review, and subspecialized reports. Journal of the American College of Radiology, 14(8), 1080–1086. https://doi.org/10.1016/j.jacr.2017.03.023

Kim, Y. W., & Mansfield, L. T. (2014). Fool me twice: Delayed diagnoses in radiology with emphasis on perpetuated errors. American Journal of Roentgenology, 202(3), 465–470. https://doi.org/10.2214/ajr.13.11493

Lian, K., Bharatha, A., Aviv, R. I., & Symons, S. P. (2011). Interpretation errors in CT angiography of the head and neck and the benefit of double reading. American Journal of Neuroradiology, 32(11), 2132–2135. https://doi.org/10.3174/ajnr.a2678

Lumbreras, B., Donat, L., & Hernández-Aguado, I. (2010). Incidental findings in imaging diagnostic tests: A systematic review. The British Journal of Radiology, 83(988), 276–289. https://doi.org/10.1259/bjr/98067945

Murphy, A. (2021, September 16). Errors in Diagnostic Radiology: Radiology Reference Article. Radiopaedia Blog RSS. Retrieved November 27, 2022, from https://radiopaedia.org/articles/errors-in-diagnostic-radiology?lang=us

RADPEER. American College of Radiology. (n.d.). Retrieved November 27, 2022, from https://www.acr.org/Clinical-Resources/RADPEER

Sipherd, R. (2018, February 28). The third-leading cause of death in us most doctors don't want you to know about. CNBC. Retrieved November 27, 2022, from https://www.cnbc.com/2018/02/22/medical-errors-third-leading-cause-of-death-in-america.html?scrlybrkr=f7f505cb#:~:text=According%20to%20a%20recent%20study,after%20heart%20disease%20and%20cancer.