Research Article

Interobserver Variability in Interpretation of VQ SPECT and its Impact on Patient Management

Boktor RR[1,2,3]*, Poon A[1,4], Berlangieri SU[1], Tauro A[1], Lee ST[1,2,3,4], Gong S[1,5], Thomas SS[1] and Scott AM[1,2,3,4]

[1]Department of Molecular Imaging and Therapy, Austin Health, Melbourne, VIC Australia
[2]Olivia Newton-John Cancer Research Institute, Melbourne, VIC, Australia
[3]School of Cancer Medicine, La Trobe University, Melbourne, VIC, Australia
[4]Faculty of Medicine, University of Melbourne, Melbourne, VIC, Australia
[5]School of Engineering and Mathematical Sciences, La Trobe University, Melbourne, VIC, Australia

*Corresponding author: Boktor RR, Department of Molecular Imaging and Therapy, Austin Health, 145-161 Studley Rd, Heidelberg, VIC 3084, Australia

Abstract

Background: Pulmonary Embolism (PE) is a major cause of morbidity, mortality and hospitalization. Ventilation Perfusion lung scan (VQ) is a powerful tool in diagnosing PE. It has been noted that there are some variations between highly experienced physicians in interpreting VQ SPECT due to lack of widely accepted reporting guidelines.

Aim of the Study: Is to measure the interobserver variability in interpreting VQ scans, and then re-measure it again after applying standardized guidelines.

Methods: Two cohorts of patients were included in this study the first included 347 patients and the second 290. Interobserver variability between 4 experienced physicians was measured on the first cohort and re-measured on the second cohort after applying 10 points agreed standardized guidelines.

Results: Showed substantial increase in the percentage of agreement between all the physicians after applying the agreed 10 points standardized diagnostic criteria. This was apparent in all the categories with the highest agreement achieved when comparing 2 physicians. Kappa value increased from 0.346 to 0.4665 between the 4 Physicians, from low 0.3 to high 0.4 range between 3 Physicians and from as low as 0.2762 to the maximum of 0.5516 between 2 physicians. Unclassified number decreased between the 2 cohorts from 16.5% to 8% and subsequently decreasing false positive cases from 7.5% to 1.7%.

Conclusion: Adherence to reporting guidelines increases the interobserver agreement in interpreting VQ SPECT leading to better patient outcomes and increased referrer confidence in reporting VQ SPECT.

Keywords: VQ SPECT; Pulmonary embolism; Interobserver variability

Introduction

Pulmonary embolism remains a diagnostic challenge and both missed diagnosis and over diagnosis have undesirable clinical consequences. Untreated PE is reported to have a mortality rate of up to 30% [1] while anticoagulant therapy exposes patients to a significant risk of bleeding [2], hence the need for accurate and precise diagnosis.

The diagnosis of pulmonary embolism can be made by imaging with either VQ or Computed Tomography of the Pulmonary Arteries (CTPA). Lung scintigraphy has been used for more than 50 years for the diagnosis of pulmonary embolism. It is a safe study with no absolute contraindication, but the planar images have some limitations which can impact on the sensitivity of the diagnosis. However, the equipment, imaging techniques and protocols, radiotracers, viewing platforms, and interpretation have significantly evolved over the years. More recently the routine use of SPECT VQ scintigraphy has improved the diagnostic performance of the study [3–6] and reduced the percentage of non diagnostic studies [7–10].

Reporting planar VQ scan initially used The Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED) criteria, which remained the standard interpreting guidelines for long time [11]. It is a probabilistic reporting criterion which does not encompass a binary reporting system of presence or absence of PE in the study. The introduction of the SPECT technique in VQ which is also used in many other nuclear medicine procedures has coincide with a change in reporting algorithm from a probabilistic to a binary method. In 2019, the European Association of Nuclear Medicine (EANM) published guidelines for VQ scintigraphy strongly recommending the use of SPECT and advocating the use of a binary reporting method [12].

Austin Health is one of the largest tertiary hospitals in Australia and performs more than 800 VQ SPECT scans per year. The diagnostic pathways at Austin Health first risk stratify patients for possible PE using the Wells score [13,14]. High probability patients are triaged to imaging. VQ scan is the first imaging choice in patients with allergy to iodinated contrast, females of reproductive age (<50 years old), and patients with impaired renal function (eGFR<45) (Figure 1). VQ is also the first imaging modality of choice in pregnant patients and in cases where CTPA is equivocal or technically inadequate. All other patients with suspected PE undergo CTPA. This triage protocol was based on clinical consensus among senior clinicians and taking into account radiation exposure and risks of CTPA.

Despite early reports of lower interobserver variation when

Figure 1: Austin Health diagnostic strategies in suspected pulmonary embolism.
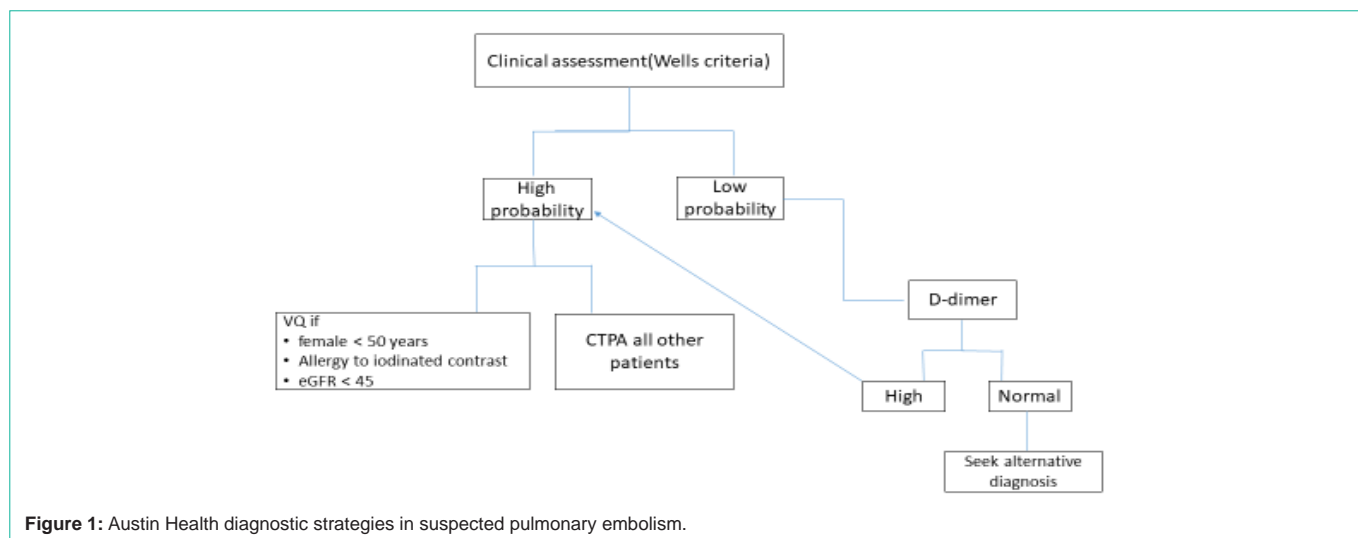
interpreting VQ SPECT on small cohorts of patients [15,16], with its routine use, we have noted some variations between highly experienced nuclear medicine physicians in interpreting VQ SPECT scans which has been identified when reviewing follow up study.

Since VQ SPECT scans are a powerful tool and a corner stone in guiding treatment in patients with PE, interobserver variability should be kept to its lowest possible.

The aims of the study were to: 1) evaluate inter-observer variability in reporting VQ SPECT studies, 2) determine if use of defined reporting criteria impacted on inter-observer variability in reporting VQ SPECT studies, and 3) evaluate if the use of defined reporting criteria reduced the number of equivocal reports of VQ SPECT studies.

Patients and Methods

This is a single institution retrospective study. Ethics approval was obtained from the Austin Health Human Research Ethics Committee (HREC/57162/Austin-2019).The imaging data base of the Department of Molecular Imaging and Therapy at Austin Health was reviewed. All patients who had VQ SPECT from 15th July 2019 to 30th October 2019 at our Institution were included. A total of 347 patients were included in this first cohort. All the scans were blindly reviewed by 4 experienced nuclear medicine physicians each with at least 10 years' experience. Each physician was provided with all the clinical information that was available at the time of the original report. The data were extracted from the request form, the routine history taking sheet performed at the time of scan, technologist worksheet, and any clinical notes written in the original VQ SPECT report in addition to useful clinical notes from the hospital electronic medical record. All the patients' medical information past the time of the scan was not available to the reading physicians and all the physicians were blinded to the original VQ SPECT report. Results were categorized as positive, negative, or equivocal. There were no agreed pre-determined set criteria in our institution for definition of positive study.

All the scans were anonymized and viewed as a 3D display using MEDVIEW DELTA manager V12 (Medimage Inc., MI,

USA). Iterative Reconstruction Non Attenuation Corrected (IRNC) ventilation and perfusion images and low dose CT which was available for many of the studies were used.

A standardized department VQ SPECT protocol was performed in all patients. Ventilation SPECT images (60 frames, each 22 seconds)) were performed firstly using average dose of 40 MBqTc-99m Technegas followed by perfusion SPECT images (60 frames, each 10 seconds) after intravenous injection of 100-200 MBq Tc-99m macro aggregated albumin on GE Discovery 670 gamma Camera (GE Healthcare, Buc Cedex, France). Low dose CT (120 Kev, average mAs 70-100) was obtained in most of the cases at the discretion of the attending nuclear medicine physician who reported the initial scan except for pregnant women and unequivocal positive or negative scans. For pregnant patients, the dose is reduced by half for both ventilation and perfusion agents.

The results of this first cohort review were statistically analyzed and interobserver variability calculated. Standardized criteria for interpreting VQ SPECT as positive or negative were then established and agreed by all the nuclear medicine physicians in the department. The data base of the Department of Molecular Imaging and Therapy was again reviewed and all patients who had VQ SPECT from 15th November 2019 to 30th March 2020 were included. A total of 290 patients were included in the 2nd cohort. These studies were interpreted independently by all 4 physicians using the agreed standardized diagnostic criteria. Once again all scans were anonymized and viewed using the same imaging settings as the first patient cohort.

The agreed standardized criteria we established for interpretation of VQ SPECT and diagnosing a positive study in the second cohort consisted of 10 points guidelines, aligned with the EANM, 2019 VQ SPECT guidelines.

1.   PE is considered if at least 2 subsegmental or one segmental mismatched perfusion defect (EANM , 2019 VQ SPECT guidelines) [12].

2.   Defects should conform to the pulmonary vascular anatomy, and are peripheral, wedge shaped and pleural based.

3. Defect is seen in more than one slice and more than one plane.

4. Defects that correspond to underlying lung vessel, opacity, pulmonary hilum, mass or structural changes are not PE.

5. Any single sub-segmental mismatched defect is not considered PE.

6. Known artefacts such as rind artefact on ventilation, fissure artefact, stripe sign, reverse mismatch, and triple match should be always kept in mind during interpretation and are not considered PE.

7. Use 3D display platform to assess the defects in the standard 3 orthogonal planes: coronal, sagittal and transaxial views. Defects should have similar appearance (triangular, pleural based and peripheral) on the 3 views to be called positive.

8. Adherence to binary reporting (positive or negative) unless severe underlying lung disease renders images non diagnostic.

9. Consider holistic view rather than absolute findings, meaning taking into consideration the clinical presentation, risk factors and laboratory findings. For example, patients with non elevated D-dimer are very unlikely to have pulmonary embolism.

10. Consider low dose CT in all the cases, unless there is contraindication (pregnancy), unequivocal positive study or completely negative study.

The impact of the VQ SPECT scan result on patient management was also measured in this study. A true positive study was defined as one with either one or combination of: A) 3 reporting physicians agreement, B) patient was treated with anticoagulation with resolution or persistent mismatched defects on follow up VQ scan, or C)positive CTPA. A true negative was defined as one with either one or combination of: A) 3 reporting physicians agreement, B) patients received no anticoagulation with no recurrent presentation with similar symptoms for 3 months, or C) negative CTPA. Patients who could not be categorized as positive or negative based on these criteria were defined as unclassified. These results were compared to the original VQ SPECT report.

Statistics

The percentage agreement among two, three, and four physicians was calculated for assessment criterion. The interobserver variability corrected for chance between any two physicians was evaluated with the weighted Cohen's kappa coefficient at the 95% confidence interval. The interobserver variability among three or four physicians was evaluated with the weighted Fleiss' kappa coefficient. The strength of agreement for the kappa statistics was categorized using the scale initially proposed by Landis and Koch [17]. A kappa value less than 0.01 would be considered to be no agreement, 0.01 - 0.20 poor, 0.21 - 0.40 fair agreement, 0.41 - 0.60 moderate agreement, 0.61 - 0.80 substantial agreement, and 0.81 - 1.00 good agreement.

All statistical analyses were performed with SPSS statistical software version 26 for Windows.

Results

Patient Population

The first cohort included 347 patients of whom 221 had VQ SPECT CT and 126 had only VQ SPECT.

62 patients were follow up studies and had a previous VQ scan, but were treated as a stand-alone scan without reference to previous studies for the purpose of this analysis (Table 1).

The number of positive and negative cases reported by the 4 physicians are comparable between 3 physicians and different for the fourth. Physician C showed a high number of positive and equivocal cases (27.4% and 13.5 % respectively), almost double the numbers of the other 3 Physicians and consequently a lower number of negative cases (Table 2).

The overall interobserver agreement among the 4 physicians was 58.79 % with overall Fleiss' Kappa value of 0.346 indicating fair (less than moderate) strength of agreement with p value <0.05.

Agreement has also performed between 3 and 2 readers. There was overall moderate agreement (75.5%) with overall Fleiss' Kappa value of 0.4231 among physicians A,B, and D which was superior to 4 physicians together. Agreement between physician C and the other 3 physicians was 61.38%, again indicating fair (less than moderate) strength of agreement category. A better agreement was seen between 2 physicians reaching a maximum of 83.86% between physician A,B and D interchangeably with Cohen's kappa value of 0.457 lying in the moderate category and a minimum 67.15 % and Cohen's Kappa value of 0.313 lying in the fair category between physician C and the other physicians (Supplementary 1).

Table 1: Characteristics of patients in the first cohort.

No. of Patients	Total n= 347	
Age range	(17-95 year)	
Sex		
Male	111	
Female	236	
Indication for VQ	N	%
Follow up for previous PE	62	18
Pregnant with either one or combination of (tachycardia, SOB or chest pain)	23	6.5
Female <50 with at least one symptom (tachycardia, shortness of breath, chest pain) and at least one risk factor (contraceptive pills, long flight, post operative, …)	139	40
Poor renal function or allergy to contrast agents in high probability patients.	40	11.5
Equivocal or non diagnostic CTPA in high probability patients.	21	6
Symptomatic patients with no risk factors but elevated D-dimer.	45	13
Pulmonary hypertension to exclude chronic PE	17	5

Table 2: Number of positive, negative and equivocal cases by each physician in the first cohort of patients.

Reviewer	Total cases	Number(%) of cases		
		"positive"	"negative"	"equivocal"
Physician A	347	37(10.6%)	283(81.5%)	27(7.7%)
Physician B	347	50(14.4%)	282(81.2%)	15(4.3%)
Physician C	347	95(27.4%)	205(59.08%)	47(13.5%)
Physician D	347	49(14.2%)	292(84.1%)	6(1.7%)

Table 3: Characteristics of patients in the second cohort.

No. of Patients	(total n= 290)	
Age range	(17-94 year)	
Sex		
Male	84	
Female	206	
Indication for VQ	N	%
Follow up for previous PE	56	19.5
Pregnant with either one or combination of (tachycardia, SOB or chest pain)	28	9.5
Female <50 with at least one symptom (tachycardia, shortness of breath, chest pain) and at least one risk factor (contraceptive pills, long flight, post operative, …)	103	35.5
Poor renal function or allergy to contrast agents in high probability patients.	35	12
Equivocal or non diagnostic CTPA in high probability patients.	11	4
Symptomatic patients with no risk factors but elevated D-dimer.	45	15.5
Pulmonary hypertension to exclude chronic PE	12	4

The above figures show that physician C contributed to a larger percentage of the variability between the physicians. Physicians A,B and D had a 75.5 % agreement with Fleiss' Kappa value 0.534 within the moderate agreement category and even greater agreement between each two of them reaching up to 83.86 % with Cohen's Kappa value of 0.457.

36 patients had CTPA within 48 hours of the VQ. In those cases the initial investigation report (either VQ or CTPA) was suggestive but not confirmatory and hence the other imaging modality was performed. A total of 22 cases (61%) had concordant results; 6 cases were positive on CTPA and original VQ report and 16 cases were negative. A total of 14 cases (39%) were disconcordant with different combinations of negative, positive and equivocal on the two studies.

The second cohort included 290 patients, 107 had VQ SPECT and 183 had VQ SPECT CT. A total of 56 patients were follow up studies and had a previous VQ scan, but were again treated as a standalone scan for the purpose of this analysis (Table 3).

In the second cohort, 2 physicians reported a lower number of equivocal cases to around 1% adhering to the binary method of interpretation. Physician C who had the highest positive number of cases in the first cohort, showed a reduced percentage of positive cases in the second cohort (16.9 % as compared to 27.4%) indicating increased specificity by adhering to the reporting guidelines. Physician B also showed an increased number of negative cases (90 % as compared to 81 %) (Table 4 & Figure 2).

The 4 physicians agreement in the second cohort increased from 58.79 % to 74.14 % with an increase in the Fleiss' kappa value from 0.3465 (fair) to 0.4665 (moderate agreement) with p value < 0.05 which indicates improved agreement among the 4 physicians (Figure 3).

The agreement among 3 physicians also increased up to 84.48 % from 75.5 % between the most agreeable physicians (A,B and D) with Fleiss' Kappa value increasing from 0.4231 to 0.5203. Also, when physician C was involved, agreement increased from around 61 % to

Table 4: Number of positive, negative and equivocal cases by each physician in the second cohort of Patients.

Reviewer	Total cases	Number(%) of cases		
		"positive"	"negative"	"equivocal"
Physician A	290	28 (9.66%)	244(84.14%)	18 (6.2%)
Physician B	290	26(8.97%)	261(90.0%)	3(1.03%)
Physician C	290	49(16.9 %)	210(72.14 %)	31(10.69%)
Physician D	290	39(13.4%)	250(86.12%)	1(0.34%)

around 76 % with Fleiss' Kappa value increasing from 0.3 to 0.4 range and from the fair to moderate category of agreement.

The agreement between 2 physicians has increased approaching 88- 89 % between physician A,B, and D interchangeably with Cohen's kappa value increasing up to 0.551 . Similarly where physician C has been involved agreement increased from around 66-68% to 80 % with Cohen's Kappa value increasing up to 0.45 from the lowest value of 0.27 with agreement improving from fair to moderate category (Supplementary 1).

23 patients had CTPA within 48 hours of the VQ. In those cases the initial investigation report (either VQ or CTPA) was suggestive but not confirmatory and hence the other imaging modality was performed.
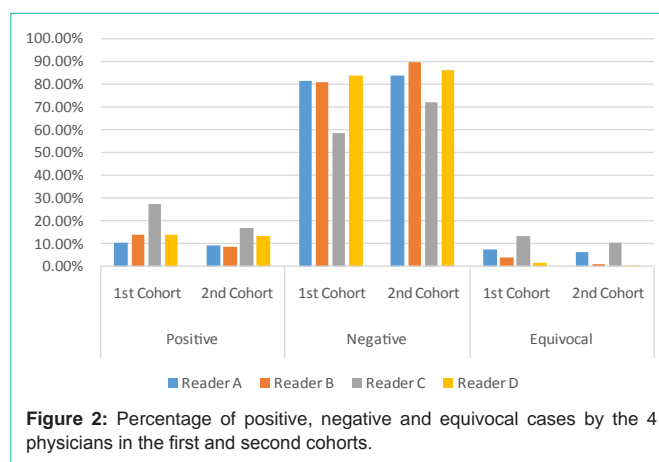

Figure 2: Percentage of positive, negative and equivocal cases by the 4 physicians in the first and second cohorts.
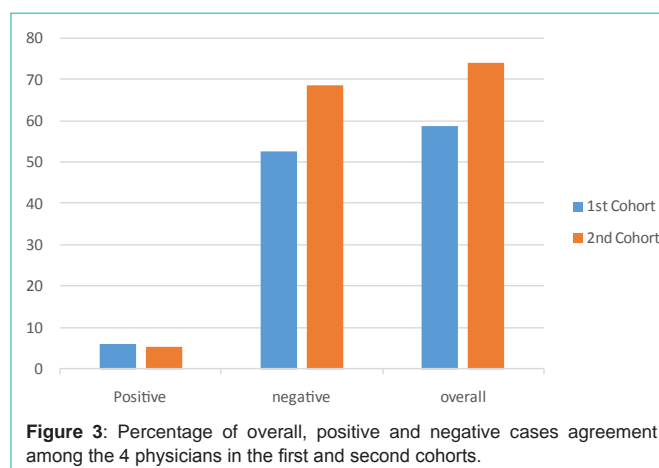

Figure 3: Percentage of overall, positive and negative cases agreement among the 4 physicians in the first and second cohorts.

A total of 8 cases were positive on CTPA and all concordant with the original VQ report. A total of 15 cases were negative and again all concordant with the original VQ report.

Impact on Patient Management

At Austin Health, treatment decision of pulmonary embolism is based on VQ SPECT or CTPA result. If the scan is positive, patient is likely to be anti coagulated.

In the first cohort, according to the pre-set definition of true positive and true negative, 253 patients out of the 347 were negative of whom 29 patients were originally reported as positive for small sub segmental defects on the original report. 37 were positive; all were reported as positive on the original report and have been anticoagulated. 57 patients (16.5%) were unclassified as positive or negative as they were not meeting the pre-set definition. 26 (7.5%) of those 57 were reported as positive and treated accordingly.

In the second cohort, the number of unclassified patients has decreased, 23 patients (8%) out of 290 could not be classified as positive or negative. Only 5 of them (1.7%) were reported as positive on the original report and treated accordingly. 26 patients were positive and all reported as positive on the original report and anticoagulated. 241 were negative of whom 25 patients were reported as positive for small subsegmental defects on the original report.

These figures show decrease in the number of the unclassified cases between the 2 cohorts (reduced from 16.5% to 8%) indicating improved agreement which subsequently decreased the number of false positive cases (decreased from 7.5% to 1.7%).

Discussion

At our Institution VQ scan is the imaging modality of first choice for young females in child bearing period, pregnant women, patients with impaired renal function and patients with contrast allergy with suspected pulmonary embolism. This has also been recommended and indicated by the European societies in institutions where a nuclear medicine facility is available and there is expertise in its interpretation [18-19].

In this study, we measured the interobserver variability between 4 experienced nuclear medicine physicians in 2 separate cohorts of patients. The first cohort without applying agreed standardized guidelines for interpreting SPECT VQ and the second cohort after applying the guidelines. Our results showed substantial increase in the percentage of agreement between all the nuclear medicine physicians after applying standardized diagnostic criteria adopted from the EANM guidelines. This was apparent in all the categories with the highest agreement achieved when comparing 2 physicians. Kappa value increased from 0.346 to 0.4665 between the 4 physicians, from low 0.3 to high 0.4 range between 3 physicians and from as low as 0.2762 to the maximum of 0.5516 between 2 physicians. Consequently, this has impact on patients' management as our results showed decrease in the number of false positive cases from 7.5% to 1.7%. This indicates the importance of applying standardized guidelines to maximize interobserver agreement in interpreting VQ studies.

There have been few prior studies that have looked at the interobserver variability in reporting VQ studies. In a study performed in 2003, by Hagen et al., on 328 patients comparing the interobserver variability in interpreting planar VQ scan between 2 physicians using 3 sets of criteria; the PIOPED, Hull, and the Gestalt. The variability assessed by Kappa score was 0.7, 0.79 and 0.65 respectively. The differences in Kappa values between the Hull and PIOPED criteria and between the Hull criteria and Gestalt interpretation were statistically significant (P < 0.05 and P <0.001, respectively) [20].

Liu and Larcos performed a study on 165 patients with 2 specialist reviewing VQ SPECT CT and perfusion SPECT CT. Intraobserver agreement with VQ SPECT was perfect (K (cohen's Kappa)=0.91 for physician 1 and K=0.95 for physician 2, P < 0.001), but not with perfusion only SPECT/CT (K= 0.4 for R1 and k = 0.62 for R2; P < 0.001). Inter-observer agreement was moderate for VQ SPECT (k = 0.65) and VQ SPECT/CT (k = 0.63)[21]. In our study, the highest interobserver agreement between 2 Physicians was also moderate with the highest K = 0.55 in the second cohort which has increased from K=0.45 in the first cohort and moved from fair to moderate agreement.

Apart from the EANM guidelines for interpreting VQ SPECT, published in 2009 and updated in 2019, the literature lacks a robust method and practice guidelines to apply in interpreting SPECT VQ. A survey conducted in 3 countries; Canada, France and Australia in 2015 involving 331 centers in the 3 countries showed 60 % of the centers who perform SPECT VQ use EANM guidelines (binary interpretation with a diagnostic threshold of 1 segmental or 2 subsegmental mismatched defects). 20% did not use standardized criteria, 11% used a binary reporting interpretation with a diagnostic cut off of 1 subsegmental mismatched defect, and 8% used the revised PIOPED criteria. The proportion of sites using the various VQ SPECT interpretation criteria was broadly consistent among the 3 countries [22].

Another study by Le Roux et al in 2013 assessed the performance of SPECT VQ using various interpreting criteria in comparison to a validated independent diagnostic strategy on 249 patients. They found of all the tested criteria, the best performance was achieved using a diagnostic cut off of at least 1 segmental or 2subsegmental mismatches, with sensitivity and specificity of 0.92 and 0.91 respectively. With a negative SPECT VQ result, the post test probability of PE was 0.010, 0.037, and 0.119 for a low, intermediate, and high clinical probability respectively. With a positive result, the post test probability of PE was 0.531, 0.814, and 0.939 for a low, intermediate, and high probability respectively [23].

In our study, the main reason for interobsever variability was found to be using personal expertise in interpreting the scans rather than agreed standardized criteria. Despite every physician having a long history of experience in interpreting VQ SPECT, the personal criteria of positive and negative studies are variable as is the case in many other nuclear medicine centres. Further analysis of the reasons for personal criteria variability revealed that some physicians used a single mismatched subsegmental defect as a criterion for PE. Physicians may depend on the absolute scan findings rather than having a holistic approach taking into consideration the symptoms, D-dimer, risk factors and pre-test probability. Interpretation of VQ SPECT studies requires careful evaluation of technical issues such as impact of anatomy on interpretation of scans, such as fissures or

stripe artifacts, and careful comparison of VQ SPECT to low dose CT (when performed) may also assist in this diagnosis.

Non adherence to a binary way of reporting and using probabilistic way increases the uncertainty and equivocal results. The use of single plan e.g transaxial plan instead of 3D display with the three orthogonal views may also result in equivocal studies.

In preparing our guidelines we have adopted the most widely known, agreed and used EANM guidelines for VQ SPECT for diagnosis of PE. PE is diagnosed based on 1 segmental or 2 subsegmental mismatch. Matched and reverse mismatch defects are not PE. Equivocal cases are those with widespread matched perfusion/ventilation defects such as in the setting of sever COPD. We then added our approach to help increase the specificity and accuracy of VQ SPECT interpretation, resulting in a 10 points standardized guidelines.

We found adding low dose CT to VQ SPECT is one of the most important criteria to increase the accuracy and confidence in diagnosing PE. This is attributed to the fact of anatomical correlation of perfusion defects increasing the confidence of whether it is corresponding to a true parenchymal lung segment or other structure. This is particularly applicable in cases of sever parenchymal lung disease like chronic obstructive airway disease, interstitial lung disease and severe inflammation/infection. In a study performed by Gutte, et al., on 81 patients, the specificity of scintigraphy increased from 88% to 100 % using VQ SPECT CT as compared to VQ SPECT. Inconclusive rate of VQ SPECT CT was zero compared to 5 % for VQ SPECT. Sensitivity was identical in both [24].

Many studies have assessed the accuracy of VQ SPECT in PE diagnosis [25-27] and have concluded that VQ SPECT is superior [27] or equal to CTPA [25]. However, in some clinical communities the accuracy is still questionable as in some of the published systematic reviews and meta-analyses. There is wide heterogeneity within the studies in terms of the reference standards and criteria for interpretation [28]. Hence the need for widely acceptable interpretation criteria to follow. Currently there is an ongoing systematic review on the performance of VQ SPECT for diagnosis of PE using objective and widely acceptable tools trying to avoid the heterogeneity and bias in some of the previously published studies [28]. In our study, the results of CTPA and VQ SPECT were concordant in 61% of the patients who had both in the first cohort, and increased to 100% concordance in the second cohort after applying the 10 points standardized guidelines.

Interobserver variability in interpreting VQ SPECT has implication on patients' management. In the first cohort, the unclassified patients were 57 (16.5%) of whom 26 (7.5%) have been treated, which could potentially be negative as the agreement was 50% or less. In the second cohort, the number of unclassified patients has decreased to 23 patients (8%) with only 5 patients (1.7%) were reported as positive and treated. Following standardized reporting guidelines decreased the interobserver variability and therefore decreased exposing questionable negative patients to the risk of anticoagulation therapy. All the positive cases were reported as positive on the original report with subsequent adequate treatment. The variability was in the negative and equivocal cases which results in increasing the percentage of negative cases having unnecessary anticoagulation therapy.

This study has some limitations, firstly the number of cases in the second cohort was less than in the first cohort (290 as compared to 347), however this difference should not contribute significantly to the statistical analysis. Secondly, some follow up studies were included, but they were treated as a single time point study with no reference to the baseline study. Thirdly, two different cohort of patient were compared, but this was designed to avoid any possible bias from the physicians reviewing cases they had already seen before.

Conclusion

Adherence to reporting guidelines increases the interobserver agreement in interpreting VQ SPECT leading to better patient outcomes and increased referrer confidence in reporting VQ SPECT.

Declarations

All authors read and approved the final version of the manuscript.

Availability of Data and Material

All the data used in this research is available.

Ethics Approval

Ethics approval has been obtained.

References

1. DW Barritt, SC Jordan. Anticoagulant drugs in the treatment of pulmonary embolism. A controlled trial. Lancet. 1960; 1: 1309–1312.

2. M Carrier, G Le Gal, PS Wells, MA Rodger. Systematic review: case-fatality rates of recurrent venous thromboembolism and major bleeding events among patients treated for venous thromboembolism Ann Intern Med. 2010; 152: 578–589.

3. Roach PJ, Schembri GP, Bailey DL. VQ scanning using SPECT and SPECT/CT. J Nucl Med. 2013; 54: 1588–1596.

4. Mortensen J, Gutte H. SPECT/CT and pulmonary embolism. Eur J Nucl MedMol Imaging. 2014; 41: S81–S90.

5. Le Duc-Pennec A, Le Roux PY, Cornily JC, Jaffrelot M, Delluc A, et al. Diagnostic accuracy of singlephotonemission tomography ventilation/perfusion lung scan in the diagnosis of pulmonary embolism. Chest. 2012; 141: 381–387.

6. Bajc M, Olsson CG, Olsson B, Palmer J, Jonson B. Diagnostic evaluation ofplanar and tomographic ventilation/perfusion lung images in patients with suspected pulmonary emboli. Clin Physiol Funct Imaging. 2004; 24: 249–256.

7. Miles S, Rogers KM, Thomas P, Soans B, Attia J, et al. A comparison of single-photon emissionCT lung scintigraphy and CT pulmonary angiography for the diagnosis of pulmonary embolism . Chest. 2009; 136: 1546–1553.

8. Bajc M, Olsson B, Palmer J, Jonson B. Ventilation/perfusion SPECT for diagnostics of pulmonary embolism in clinical practice. J Intern Med. 2008; 264: 379–387.

9. Leblanc M, Paul N. VQ SPECT and computed tomographic pulmonary angiography. Semin Nucl Med. 2010; 40: 426–441.

10. Pierre-Yves Le Roux, Philippe Robin, Pierre-Yves Salaun. New developments and future challenges of nuclear medicine and molecular imaging for pulmonary embolism. Thrombosis Research. 2018; 163: 236–241.

11. PIOPED Investigators. Value of the ventilation/perfusion scan in acute pulmonary embolism: results of the prospective investigation of pulmonary embolism diagnosis (PIOPED). JAMA. 1990; 263: 2753–2759.

12. Bajc M, Schümichen C, Grüning T, Lindqvist A, Le Roux PY, et al. EANM

guideline for ventilation/perfusion single-photon emission computed tomography (SPECT) for diagnosis of pulmonary embolism and beyond. Eur J Nucl Med Mol Imaging. 2019; 46: 2429–2451.

13. Marc Righini , Grégoire Le Gal , Drahomir Aujesky, Pierre-Marie Roy, Olivier Sanchez, et al. Diagnosis of pulmonary embolism by multidetector CT alone or combined with venous ultrasonography of the leg: a randomised non-inferiority trial. Lancet. 2008; 371: 1343–52.

14. Henrike J Schouten, G J Geersing, H L Koek, Nicolaas P A Zuithoff, Kristel J M Janssen, et al. Diagnostic accuracy of conventional or age adjusted D-dimer cut-off values in older patients with suspected venous thromboembolism: systematic review and meta-analysis. BMJ. 2013; 346: 2492-2505.

15. Bajc M, Olsson CG, Olsson B, Palmer J, Jonson B. Diagnostic evaluation ofplanar and tomographic ventilation/perfusion lung images in patients with suspected pulmonary emboli. Clin Physiol Funct Imaging. 2004; 24: 249–256.

16. Collart JP, Roelants V, Vanpee D, Lacrosse M, Trigaux JP, et al. Is a lung perfusion scan obtainedby using single photon emission computed tomography able to improve theradionuclide diagnosis of pulmonary embolism? Nucl Med Commun. 2002; 23: 1107–1113.

17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33: 159–74.

18. Hansen SL, de Nijs R, Mortensen J, Bonilla JJ, Amador NM, et al. Ventilation–perfusion SPECT versus CTPA in young adult females with suspected pulmonary embolism. Eur Respir J. 2020; 55: 2000448.

19. Konstantinides SV. The optimal imaging test for diagnosis of acute pulmonary embolism: a second chance for lung scintigraphy?. Eur Respir J. 2020; 55: 2001426.

20. Petronella J Hagen, Ieneke JC Hartmann, Otto S Hoekstra, Marcel PM Stokkel, Pieter E Postmus, et al. Comparison of Observer Variability andAccuracy of Different Criteria for Lung Scan Interpretation. Nucl Med. 2003; 44:739–744.

21. Jui Liu, George Larcos. Radionuclide lung scans for suspected acute pulmonary embolism: Single photon emission computed tomography (SPECT) or hybrid SPECT/CT?. J Med Imaging Radiat Oncol. 2019; 63: 731-736.

22. Pierre-Yves Le Roux, Matthieu Pelletier-Galarneau, Romain De Laroche, Michael S Hofman, Lionel S Zuckier, et al. Pulmonary Scintigraphy for the Diagnosis of Acute PulmonaryEmbolism: A Survey of Current Practices in Australia, Canada, and France. J Nucl Med. 2015; 56: 1212–1217.

23. Pierre-Yves Le Roux, Philippe Robin, Aurélien Delluc, Ronan Abgral, Alexandra Le Duc-Pennec, et al. VQ SPECT Interpretation for Pulmonary Embolism Diagnosis: Which Criteria to Use?. J Nucl Med. 2013; 54: 1077–1081.

24. Gutte H, Mortensen J, Jensen CV, Johnbeck CB, von der Recke P, et al. Detection of pulmonary embolism with combined ventilation-perfusion SPECT and low-dose CT: head-to-head comparison with multidetector CT angiography. J Nucl Med. 2009; 50: 1987–1992.

25. Phillips JJ, Straiton J, Staff RT. Planar and SPECT ventilation/perfusion imaging and computed tomography for the diagnosis of pulmonary embolism: A systematic review and meta-analysis of the literature, and cost and dose comparison. Eur J Radiol. 2015; 84: 1392–400.

26. Kan Y, Yuan L, Meeks JK, Li C, Liu W, et al. The accuracy of VQ SPECT in the diagnosis of pulmonary embolism: a meta-analysis. Acta Radiol. 2015; 56: 565–72.

27. Hess S, Frary EC, Gerke O, Madsen PH. State-of-the-Art Imaging in Pulmonary Embolism: Ventilation/Perfusion Single-Photon Emission Computed Tomography versus Computed Tomography Angiography-Controversies, Results, and Recommendations from a Systematic. Review. Semin Thromb Hemost. 2016; 42: 833–45.

28. Pierre-Yves Le Roux, Philippe Robin, Cécile Tromeur, Alexandra Davis,Helia Robert-Ebadi, et al. SPECT VQ for the diagnosis of pulmonary embolism: protocol for a systematic review and meta-analysis of diagnostic accuracy and clinical outcome. BMJ Open. 2018; 8: e022024.