Evaluating discrepancy rates of radiology resident provisional reports for cross-sectional body imaging studies at a tertiary hospital

Jonathan Kia-Sheng Phua1 and Lionel Tim-Ee Cheng2

Abstract

Introduction: Urgent radiological studies obtained during on-call hours are often preliminarily read by on-call residents before consultant radiologists finalise the reports at a later time. Such provisional radiology reports provide important information to guide initial patient management. This study aims to determine discrepancy rates between provisional reports and final interpretations, and to assess the clinical significance of such discrepancies.

Methods: This retrospective quality assurance project reviewed a total of 1218 cross-sectional imaging studies of the body (thorax, abdomen and pelvis) done between July 2015 and May 2016 during on-call hours. The studies included 1201 Computed tomography (CT) scans and 17 Magnetic Resonance Imaging (MRI) scans. Studies with incomplete or unavailable reports were excluded. Conclusions of both the provisional and final reports of each study were reviewed for concordance, with reference to the full report if needed. Discrepancies were graded according to the ACR 2016 RADPEER scoring system.

Results: There were 1210 studies with complete reports. Discrepant reports were noted in 183 (15.1%) studies. Of these, 89 (7.3%) were assessed to be clinically significant and the majority of these (55) were due to interpretations which should be made most of the time. CT of the abdomen and pelvis were the most prone to discrepant reports, accounting for 148 cases (80.9%).

Conclusion: The majority of preliminary reports for on-call body scans were concordant with final interpretations. The discrepancy rates for provisional body scan reports provided by residents while on call were comparable to those previously reported in literature.

Keywords
Diagnostic radiology, body imaging, cross-sectional imaging, discrepancy rates, residency training

Introduction

Most major hospitals and trauma centres provide round-the-clock services to ensure uninterrupted delivery of patient care. As not all hospitals have the resources to provide round-the-clock subspeciality expertise, many hospitals have on-call radiology residents providing preliminary reads of urgent radiological studies done after usual working hours.1 By providing such coverage, residents learn independent responsibility in scan interpretation which is integral to their training.2 Consultant radiologists then finalise the reports after variable time intervals, depending on institution practice. Medical teams use provisional radiology reports to guide timе-sensitive interventions. Inadequate radiological reports can lead to unnecessary investigations, delays in diagnosis and possible harm to the patient.3 Hence, such preliminary reports need to be accurate.

1Medical Officer, Department of Diagnostic Radiology, Singapore General Hospital, Singapore, Singapore
2Senior Consultant, Department of Diagnostic Radiology, Singapore General Hospital, Singapore, Singapore

Corresponding Author:
Jonathan Kia-Sheng Phua, Medical Officer, Department of Diagnostic Radiology, Singapore General Hospital, Outram Road, Singapore 169608, Singapore.
Email: phuakiasheng@gmail.com
This study aims to determine the rate of discrepancy between provisional reports and final interpretations, and to assess the clinical significance of such discrepancies.

Methods

A total of 1218 cross-sectional imaging studies of the body (thorax, abdomen and pelvis) done during on-call hours from 1 July 2015 to 31 May 2016 were reviewed as part of a department quality assurance project which qualified for IRB exemption. Conclusions of provisional and final reports were reviewed, and if discrepancies were found, the full provisional report was reviewed. After assessing the full provisional report, any confirmed discrepancy was then graded based on the American College of Radiology (ACR) 2016 RADPEER Scoring System which is summarised in Table 1. A clinically significant discrepancy was defined as one which results in a change in diagnosis or treatment. A non-clinically significant discrepancy would not change the main diagnosis or treatment plan (e.g. a missed old rib fracture in a CT pulmonary angiogram to exclude pulmonary embolism). An overview of how each preliminary report was assessed is illustrated in Figure 1.

Results

Of the 1218 scans done during on-call hours from 1 July 2015 to 31 May 2016, 8 had incomplete records. A total of 1210 scan reports (1195 CT scans and 15 MRI scans) were then reviewed. 653 (54.0%) scans were done after office hours on weekdays from 6 p.m. to 8.30 a.m. while 557 (46.0%) were done on weekends and public holidays.

A breakdown of all the types of scans done during on-call hours is illustrated in Table 2. CT scans covering the abdomen and pelvis were the most common studies done during on-call hours, mostly for clinical purposes.

Table 1. The ACR 2016 RADPEER Discrepancy Scoring System.

| Score | Meaning | Subclassification |
|-------|---------|------------------|
| 1     | Concur with interpretation | — |
| 2     | Discrepancy in interpretation not ordinarily expected to be made (understandable miss) | a. Unlikely to be clinically significant b. Likely to be clinically significant |
| 3     | Discrepancy in interpretation should be made most of the time | a. Unlikely to be clinically significant b. Likely to be clinically significant |

Table 2. Breakdown of types of scans done during on-call hours.

| Type of scans | Number of scans |
|---------------|-----------------|
| CT Abdomen    | 8               |
| CT Abdomen and Pelvis | 758             |
| CT Chest      | 77              |
| CT Chest (Pulmonary Angiography) | 118             |
| CT Chest and Abdomen | 7               |
| CT Chest, Abdomen and Pelvis | 152             |
| CT Chest, High Resolution | 4               |
| CT Kidneys    | 2               |
| CT KUB (Non Contrast) | 32              |
| CT Liver      | 13              |
| CT Pelvis     | 4               |
| CT Urography  | 20              |
| MRI Abdomen   | 3               |
| MRI Cholangio-Pancreaticography (MRCP) | 7               |
| MRI Pelvis    | 4               |
| MRI Prostate  | 1               |

*CT: Computed Tomography.
time-sensitive clinical questions such as the cause of an acute abdomen (Figure 2).

There were 1027 (84.9%) provisional scan reports without discrepancies which were graded as RADPEER 1. Provisional reports for 183 (15.1%) scans had discrepancies and were graded according to the ACR 2016 RADPEER classification (Table 3). Among these, 110 reports had discrepancies which were understandable misses, and these were subclassified into 2a (76 reports) and 2b (34 reports).

There were 73 preliminary scan reports with discrepancies in interpretation which should be made most of the time. Of these, 55 were assessed to be likely clinically significant discrepancies (category 3b).

Just over half of the discrepant reports (97 reports or 53.0%) were done between 6 p.m. and 8.30 a.m. on weekdays, while the remainder (86 reports or 47.0%) were done on weekends and public holidays. The vast majority of discrepant preliminary scan reports (148 reports or 80.9%) involved the abdomen and pelvis.

A total of 7.3% of scans had clinically significant discrepancies in preliminary reports (category 2b + 3b), while non-clinically significant discrepancies stood at 7.8% (category 2a + 3a). Examples of discrepancies under the various categories are summarised in Tables 4–7. Figures 3–5 illustrate examples of discrepancies graded as RADPEER 2b while Figures 6–8 illustrate those graded as RADPEER 3b.

| Table 3. Breakdown of discrepancy rates based on ACR 2016 RADPEER Classification System. |
|-----------------------------------------------|
| RADPEER grade | Number of scans (percentage of total) | Cumulative number of scans (percentage of total) |
|----------------|-------------------------------------|-----------------------------------------------|
| 1              | 1027 (84.9%)                        | 1027 (84.9%)                                   |
| 2a             | 76 (6.3%)                           | 183 (15.1%)                                    |
| 2b             | 34 (2.8%)                           |                                               |
| 3a             | 18 (1.5%)                           |                                               |
| 3b             | 55 (4.5%)                           |                                               |

| Table 4. Discrepancies graded under RADPEER 2a (understandable miss, unlikely to be clinically significant). |
|---------------------------------------------------------------|
| Category of discrepancies | Undercalled discrepancies | Overcalled discrepancies |
|----------------------------|---------------------------|--------------------------|
| Inflammation/Infection   | Gall bladder: Subtle gas from known fistula with adjacent colon, mild mural thickening | Omental infarction       |
|                           | Lung: Mild focal centrilobular nodules, small ground glass opacity, mild bronchial wall thickening |                           |
|                           | Others: Renal scarring, mild oesophageal mural thickening, subtle loculated fluid |                           |
| Neoplasia                 | Small lesions: Cystic pancreatic lesions, thyroid nodules, gall bladder polyps |                           |
|                           | GU: Uterine fibroids, prostatomegaly |                           |
| Vascular                  | Mild vascular mural thickening, progression of known thrombus, portosystemic shunt, possible venous stenosis (subclavian) |                           |
| Others                    | Incidental tiny stones (gall bladder, renal) |                           |
|                           | Others: Colonic diverticulosis, small hydrocele, mild hepatic steatosis, pelvic floor laxity, subcutaneous oedema |                           |

| Table 5. Discrepancies graded under RADPEER 2b (understandable miss, likely to be clinically significant). |
|---------------------------------------------------------------|
| Category of discrepancies | Undercalled discrepancies | Overcalled discrepancies |
|----------------------------|---------------------------|--------------------------|
| Inflammation/Infection   | Genitourinary: Pyelonephritis, early renal transplant infection, early cystitis, pelvic inflammatory disease, subtle colovaginal fistula | Compartimentalised gall bladder (called abscess) |
|                           | Gastrointestinal: Perforated appendicitis called as ileocolic intussusception, D2 diverticulitis, proctocolitis, Hepato-Pancreato-Biliary: Cholecystitis, mild peripancreatic fat stranding | Peristalsis (called colitis), Splenic infarcts (called abscess) |
|                           | Others: Post-operative collection |                           |
| Trauma                    | Post-operative: Possible urinary leak, possible bile leak |
| Neoplasia                 | Gastrointestinal: Subtle right hepatic lobe mass, appendiceal malignancy (called abscess), gastric fundus nodule (1.5 cm) |                           |
|                           | Hepato-Pancreato-Biliary: Gall bladder neoplasm, subtle pancreatic head lesion |                           |
|                           | Lung: Hypodense nodule in atelectatic lung, small nodules |                           |
|                           | Others: Subtle peritoneal nodularity |                           |
| Vascular                  | Subtle pulmonary emboli, narrowing of portal vein branches, pulmonary trunk dilatation |                           |
| Others                    | Cirrhotic, small bowel dilatation | Ovary with prominent follicles called hydroaslapinxx |

*More common discrepancies are highlighted in bold font.*
Discussion

On-call residents play an important role in the provision of uninterrupted patient care for public sector hospitals in Singapore. As part of the healthcare team on duty, radiology residents are responsible for providing timely and accurate provisional scan reports to guide the clinical teams.

Table 6. Discrepancies graded under RADPEER 3a (correct interpretation should be made most of the time, but discrepancy is unlikely to be clinically significant).

| Category of discrepancies | Undercalled discrepancies | Overcalled discrepancies |
|---------------------------|---------------------------|--------------------------|
| Inflammation/Infection    | Subcutaneous inflammation, mild ureteric wall enhancement, retroperitoneal fluid, pleural effusions | Pulmonary centrilobular nodularity |
| Trauma                    |                           |                          |
| Neoplasia                 | Known breast nodule, adrenal myelolipoma, borderline lymphadenopathy |                          |
| Vascular                  | Myocardial thinning and calcification from prior infarct, contrast reflux into hepatic veins, splenic infarct |                          |
| Others                    | Lung: Left lower lobectomy |                          |
|                           | Hepato-Pancreato-Biliary: Prominent intrahepatic biliary tree, known cirrhosis |                          |
|                           | Gastrointestinal: Mild intussusception, diverticulosis, inguinal hernias |                          |
|                           | Genitourinary: Small renal hypodensity |                          |

Table 7. Discrepancies graded under RADPEER 3b (correct interpretation should be made most of the time, and discrepancy is likely to be clinically significant).

| Category of discrepancies | Undercalled discrepancies | Overcalled discrepancies |
|---------------------------|---------------------------|--------------------------|
| Inflammation/Infection    | Gastrointestinal: appendicitis, bowel perforation with abscess, colitis | Colitis, appendicitis, pyelonephritis |
|                           | Genitourinary: Tubo-ovarian abscess |                          |
|                           | Others: Incision site infection |                          |
| Trauma                    | Hepatic laceration/contusion |                          |
| Neoplasia                 | Hepato-Pancreato-Biliary: Pancreatic head mass, LUQ mass interpreted as spleen | Liver metastases |
|                           | Gastrointestinal: Colonic stricture |                          |
|                           | Genitourinary: Ovarian tumour (called fibroid) |                          |
|                           | Bone: Metastasis (pubic rami, clavicles, vertebrae) |                          |
|                           | Lung: Pleural nodularity |                          |
| Vascular                  | Venous thrombosis (femoral vein, portal vein), splenic infarct, arterial occlusion, pulmonary embolism, active retroperitoneal bleed, missed new endoleak |                          |
| Others                    | Airway narrowing, pericoeliac lymphadenopathy, paravertebral nodule, large bowel obstruction |                          |

*More common discrepancies are highlighted in bold font.*

Figure 3. (a and b) Coronal CT images. Graded as RADPEER 2b - Gall bladder mass with enlarged necrotic lymph nodes, suspicious for gall bladder carcinoma which was provisionally reported as ‘perforated acute cholecystitis’.
It is therefore important for radiology educators to understand common ‘misses’ that residents may make while on call. Armed with information on common discrepancies in preliminary scan reports, educators can enhance training programmes to improve reporting accuracy.

Kim and Mansfield suggested 12 types of errors in diagnostic radiology, ranging from issues related to the person doing the interpretation (e.g. lack of knowledge, faulty reasoning and complacency) to factors involving the scan (e.g. limitations of scan or technique). In our series, we noticed the following broad categories of ‘misses’.

Firstly, the lesion was simply not detected or misinterpreted as not significant. Residents may have missed the lesion entirely, or picked it up but dismissed it as a normal finding or an artefact.

Secondly, the anatomical region in question may be tricky to evaluate. Some areas are notoriously difficult to interpret, such as the pancreas, bowel and post-transplant organs.

Thirdly, findings were missed in ‘blind spots’ that are commonly overlooked, such as bones, blood vessels and lung bases in studies focusing on the abdomen and pelvis.

Fourthly, residents may encounter difficult studies due to patient’s condition or scan protocol. Examples include post-operative cases with substantially altered anatomy, acutely unwell individuals with multiple pathologies or a scan with multiple phases acquired. Such cases may be more common in tertiary or quaternary hospitals where complicated subspeciality surgical services are available.

Lastly, residents may find themselves in a contextual conundrum during interpretation. Subtle findings such as focal fat stranding, mild mural thickening and gas pockets may or may not be relevant depending on the clinical context. One example is pneumoperitoneum which can either be a surgical
emergency or an unremarkable expected finding (in the context of recent peritoneal drainage, biopsy or laparotomy). Subtle but relevant findings in a given clinical presentation may be dismissed leading to inaccurate interpretation.

There is paucity of literature on discrepancies in interpretation of body imaging studies. Most prior studies used the terms ‘major’ and ‘minor’ discrepancies, which is analogous to our classification on whether a discrepancy is clinically significant or not. Our clinically significant (major) discrepancy rate (DR) was 7.3%, while that of non-clinically significant (minor) was 7.8%. In a study done by Howlett et al. on the accuracy of interpretation of emergency abdominal CT in patients who presented with non-traumatic abdominal pain in the United Kingdom, the major DR and minor DR was 4.6% and 8.4%, respectively, for both surgical and non-surgical scans provisionally reported by registrars.

A similar study by Tieng et al. on the interpretation of Emergency Department body CT scans by radiology residents has a major DR of 10% and minor DR of 20%. Another study in literature published in 1996 by Wechsler et al. on the effects of training and experience in interpretation of emergency body CT scans had a major DR of 7.8% and a minor of 5.9% among senior residents.

In a more recent study by Wu et al. on the discrepancy rates in acute abdominal CT, the major DR by registrar was 6.86% while the minor was 1.82%. The lower DRs in this study may be due to the fact that this study only involved patients who presented with non-traumatic abdominal pain who subsequently underwent emergency laparotomy. As these cases had emergency laparotomies performed, there could have been more overt radiological signs present on CT which were identified and hence accounting for the lower DRs.

Our study has several limitations. Firstly, only the conclusions of the provisional and final reports were reviewed at the first instance. Full provisional reports were not reviewed unless there was a discrepancy between the conclusions of the provisional and final reports. We assessed that most clinical decisions would be made based on report conclusions, and thus felt that this initial assessment of the conclusions only was reasonable.

Secondly, actual images from the studies were not re-read and interpreted for all cases. Hence, the “ground truth” was the final report and not a separate independent review of all the images. It is therefore plausible that some discrepancies may have been missed by our review if they were not evident in either the provisional or final report.

Thirdly, the scope of our study was limited to scans from the body subspeciality. To fully assess the performance of radiology residents on call, other studies beyond the thorax, abdomen and pelvis should also be reviewed. We chose to focus on the body scans as these were anecdotally identified by residents as being one of the more challenging scans encountered while on call. The overall DR is likely to be different, and possibly lower if studies from all subspecialities were assessed together.

Fourthly, assessment of clinical significance was based on a change of diagnosis (if any) in the final report and the likely clinical management. The full electronic medical records for these patients were not reviewed. Future studies may consider reviews of the electronic medical records for a more robust assessment of the clinical impact of the discrepancies.

Finally, the discrepancies found were not systematically analysed for the root causes. For example, the respective reporting residents were not identified and approached to evaluate whether the discrepancy was due to an error in detection or interpretation. Further classification of discrepancies may help highlight useful lessons for educating future batches of radiology residents.

Notwithstanding the abovementioned limitations, we feel that our study shows that the overall discrepancy rates, and in particular, clinically significant discrepancies for provisional body imaging radiology reports by our residents is low and comparable to that of other previously published studies. This shows that the skills acquired by our residents during the training programmes allows them to perform at a similar level while on call compared to their counterparts from other countries. Further studies of such discrepancies and the trends over time may also provide useful information for refinements to existing training programmes.

Conclusion

The majority of preliminary reports for on-call body CT scans were concordant with final interpretations. Our study revealed an overall DR of 15.1% and a clinically significant discrepancy rate of 7.3%, which is comparable to those previously reported in literature. Further studies can be done to assess DRs in different subspecialties across radiology, to identify commonly missed regions or conditions and further tailor training programmes for future batches of radiology residents.

Acknowledgements

We would like to thank the Department of Diagnostic Radiology for the assistance rendered in this research.

Author Contributions

Lionel Tim-Ee Cheng was involved in conceptualisation, protocol development and data analysis. Jonathan Kia-Sheng Phua was involved in data collection, data analysis and drafting of the manuscript.
All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

Ethical approval
This was a department quality assurance project which qualified for IRB exemption.

Informed Consent
This was a retrospective department quality assurance project with no direct or indirect patient involvement.

Availability of data and materials
The datasets generated and/or analysed during the current study are available from the Department of Diagnostic Radiology.

ORCID iD
Jonathan Kia-Sheng Phua  🏷️ https://orcid.org/0000-0003-2324-9952

References
1. Tieng N, Grinberg D and Li SF. Discrepancies in interpretation of ED body Computed Tomographic scans by Radiology residents. *Am J Emerg Med* 2007; 25(1): 45–48.
2. Carney E, Kempf J, DeCarvalho V, et al. Preliminary interpretations of after-hours CT and sonography by radiology residents versus final interpretations by body imaging radiologists at a level 1 trauma center. *Am J Roentgenology* 2003; 181(2): 367–373.
3. Wu Y, Das B, Shah V, et al. An audit of local discrepancy rates in acute abdominal CT: Does subspecialist reporting reduce discrepancy rates? *Clin Radiol* 2020; 75(11): 879.e7–879.e11.
4. Goldberg-Stein S, Frigini LA, Long S, et al. ACR RADPEER committee white paper with 2016 updates: revised scoring system, new classifications, self-review, and subspecialized reports. *J Am Coll Radiol* 2017; 14(8): 1080–1086.
5. Kim YW and Mansfield LT. Fool me twice: delayed diagnoses in radiology with emphasis on perpetuated errors. *Am J Roentgenology* 2014; 202(3): 465–470.
6. Howlett DC, Drinkwater K, Frost C, et al. The accuracy of interpretation of emergency abdominal CT in adult patients who present with non-traumatic abdominal pain: results of a UK national audit. *Clin Radiol* 2017; 72(1): 41–51.
7. Wechsler RJ, Spettell CM, Kurtz AB, et al. Effects of training and experience in interpretation of Emergency Body CT Scans. *Radiology* 1996; 199(3): 717–720.