Emerg Radiol
DOI 10.1007/s10140-010-0899-2

ORIGINAL ARTICLE

Systematic survey of discrepancy rates in an international teleradiology service

Anjali Agrawal · Anurag Agrawal · Meenakshi Pandit · Arjun Kalyanpur

Received: 18 June 2010 / Accepted: 9 August 2010

Abstract International teleradiology services (ITS) to the United States are based on the principle of deploying American board-certified radiologists across global time zones to optimally distribute the workload. While errors may be reduced by circumventing the traditional night call, there is limited evidence on the actual error rates of teleradiology groups. We have a comprehensive quality assurance (QA) process in our practice, which includes a review of discrepancies between preliminary reports and the final reports by the on-site radiologists. We analyzed the discrepancy QA data to determine the error rates. Archived QA data for 126,449 cases over a period of 1 year (2008) were analyzed for the discrepancy rate, nature of errors, and possible contributory factors. The scores ranged from 0 (no error) to 5 (clinically significant in the acute setting) based on the level of clinical significance. A novel modified Lorenz plot was used to estimate the degree of underreporting and to estimate the true error rate. An internal review of 200 cases was performed to validate the findings. Of the total, there was a total of 227 confirmed errors (0.18%, 95% CI, 0.16 to 0.20). Of these, the majority were levels 2 and 3 (minor error and error of long-term significance but not in the acute setting). Even after correction for underreporting, error rates were less than 1% for clinically significant errors. ITS is associated with very low rates of clinically significant errors. Due to limited feedback, particularly for minor errors, an internal review is important.

A. Agrawal (✉) · M. Pandit · A. Kalyanpur
Teleradiology Solutions,
New Haven, CT, USA
e-mail: anjali.agrawal@telradsol.com
e-mail: draagrawal@gmail.com

A. Agrawal
Institute of Genomics and Integrative Biology,
Delhi, India

Keywords Teleradiology · QA · Quality assurance · Underreporting bias · Radiological errors

Introduction

International teleradiology services (ITS) to the United States are based on the principle of deploying certified radiologists and ancillary staff across global time zones to optimally distribute diagnostic radiology workload [1]. The need for round-the-clock radiology services exists due to high after-hours emergent imaging load [2]. This can be met either by radiologists in similar time zones working at odd hours, as in US-based teleradiology operations or by offshore-certified radiologists working during waking hours as in ITS [3]. Anecdotal and hard evidence suggest that fatigue is an important cause of errors in diagnostic radiology, implying that errors may be reduced by radiologists reading during normal working/waking hours [4]. This also improves the quality of life for radiologists by almost eliminating the need to stay awake during odd hours [5], yet this creates a different set of challenges related to coordination between multiple sites, logistics, and sometimes local infrastructural issues [3]. While studies on discordant rates of teleradiology companies working from within the United States have been published [6], there is no published evidence of error rates of international teleradiology companies.

The authors are associated with a teleradiology group that provides radiology-related services to over 63 hospitals in the United States across 16 different states. It is Joint Commission-accredited and in full compliance with the Health Insurance Portability and Accountability Act. The group currently comprises of 13 American Board of Radiology (ABR)-certified radiologists based in India, China, Europe, Israel, and the United States, of which eight were reading for the group in 2008, from India, China, and the United States. There is an established quality assurance (QA) process wherein feedback is actively solicited from hospital-based radiology groups regarding discrepancies between preliminary reads submitted by the group and final reports by their on-site radiologists. All reported discrepancies are noted, reviewed, graded, and archived by a QA team composed of senior radiologists of the group who had more than 3 years of emergency room (ER) teleradiology experience and an interest in QA processes. Since these are based on feedback from client radiology groups, underestimation is a major concern; however, we do not know of any previous publications that attempt to address this important problem. We analyzed the entire data for the year 2008 to gain more insight into the discrepancy rates of ITS and the factors determining them. We also attempted to gain insight into the extent of underreporting of errors by client radiology groups in a teleradiology practice using a novel analytic method.

Materials and methods

Radiology reads protocol Due to technological advancements, methods of image transmission and communication are in constant evolution. A general scheme of the workflow applicable to 2008 is summarized. Radiology technologists of client hospitals submitted a requisition for radiology read to the group's operations center in Bangalore, India using either fax or electronic radiology information system (RIS) entry. The images were transferred over VPN to a web-based PACS (eRAD, Greenville, South Carolina) and downloaded from there to the individual radiologist's workstation using lossless JPEG transfer. A firewall and site to site virtual private networks were used to secure the data. Some studies were read directly from the client hospital's radiology viewing software. Preliminary reports were faxed back to clients using a prototype RIS. Critical or unexpected findings requiring immediate intervention or unexpected findings warranting follow-up were called in to the physician or nurse-in-charge. Turnaround times for the reports were maintained. Performance of eight radiologists was analyzed; the radiologists were ABR-certified, state licensed, and credentialed at the institutions where the images originated, in keeping with the ACR Standard for International Teleradiology [7].

Quality assurance protocol To facilitate a smooth and effective quality assurance process, a standardized format for recording discrepancy between final and preliminary reports was provided to all clients. All reported discrepancies were analyzed by a QA team to judge whether an error had occurred, which was further graded for level of clinical significance with regard to patient care. This QA report was communicated to the primary radiologist and hospital-based radiologists along with a recommended plan to minimize future errors. All high-level discrepancies were reviewed by a QA team of three senior radiologists.

Data collection and analysis The total numbers of cases, number of discrepancies, and clinical significance of discrepancy based on review (levels 0 to 5; Table 1) were obtained for the year. Notably, discrepancies are differences between final reports by client radiologists and initial reports by the group while errors are discrepancies that arose due to erroneous initial reports. As shown in Table 1, a level 0 discrepancy is not an error. Details pertaining to the radiologist providing the interpretation and study type were also collated. A random sample of 200 cases was collected for internal grading and validation. Errors above level 3 were considered to be of high clinical impact and equal or below level 3 to be of low clinical impact in the emergent setting. The error rate as a percentage of cases was calculated for the group, study type, clinical impact, and individual radiologist. Ninety-five percent confidence intervals of the estimate were reported. Differences in error rates between radiologists and case types were tested for statistical significance using a two-tailed Chi-square analysis of proportions ($p<0.05$).

Estimation of error rate from the discrepancy data As mentioned previously, error for the purpose of our study means a discrepancy, where the preliminary report contained an error, i.e., any non-zero discrepancy level in our QA system (Table 1). Error rate was initially estimated as the rate of discrepancies excluding level 0 and was separately calculated for all errors as well as those with

Table 1 Grading of discrepancies

Level 0	No error
Level 1	Typographical or format errors
Level 2	Minor discrepancy, not of clinical significance
Level 3	Discrepancy of clinical significance but not in the acute setting, for example, lesion requiring follow-up
Level 4	Acute clinically significant discrepancy requiring an immediate change in patient management
Level 5	Missed finding of major/critical/life-threatening significance

clinically significant errors in the acute setting. Since each and every discrepancy may not be reported by the hospital-based radiologists and level 0 discrepancies were excluded, our estimated error rates are likely lower than the true error rates. To better understand the true error rates and the degree of underreporting bias, it was assumed that error rates were independent of the client to whom radiological services were provided. In this scenario, minimum error rates are non-meaningful, represent significant underreporting, and result in distortion of central tendency estimates (mean or median). To overcome this limitation, cumulative errors were plotted as a function of total cases with data from client hospitals being added to the plot in ascending order of error rates, somewhat similar to a Lorenz plot [8]. An approximately constant error rate across client hospitals would lead to a linear or near linear plot; random variation would cause symmetrical concavity with flattening of the initial part and accentuation of the slope of the terminal part while underreporting by some (but not all) clients would cause marked concavity of the initial part of the curve (Fig. 1). It can be seen from Fig. 1 that error rates estimated from the slope of the upper bounds of the plot will be most representative of the true error rate when underreporting is present. The slope of a line fitted to the terminal 5% of the curve was used for this purpose. This was chosen by repeated simulations of scenarios with both random variation in error rates and underreporting, being less error-prone than taking the maximum error rate reported from any client (equivalent to taking the slope at the tip of the curve), which due to random variation may be an overestimate. Notably, the method presumes that over a large sample of clients, at least some will be reporting errors correctly. While this appears likely for clinically significant errors, it may not be true for trivial errors, which will therefore remain undercorrected and measurable only by a full internal review.

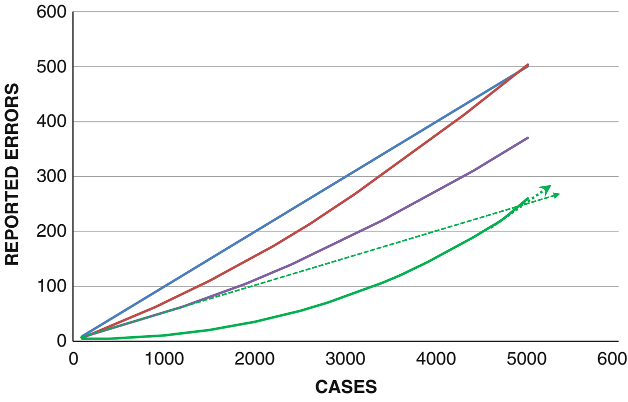

Fig. 1 A mathematical simulation of the effect of underreporting of errors on quality assurance data. A stable true error rate is assumed to exist and effects of random sampling variation around the true error rate and variable underreporting of errors is shown. The *dark blue line* shows an ideal case of no underreporting and no variation. The *red line* shows the effects of 20% random variation without any underreporting bias. The *purple line* shows the effect of variable underreporting (up to 50%), and the *green line* simulates the real-world with both underreporting (up to 100%) and random variation. It should be noted that for the real-world simulation, while the pooled error rate (slope of *dashed line* from origin) is lower than true error rate (slope of *dark blue line*), the slope of a line (*curved arrowhead*) fitted to the terminal region closely approximates the true error rate

Results

In 126,449 cases for which data was available, there was a total of 330 reported discrepancies (0.26%; 95% CI, 0.23 to 0.29) between preliminary report submitted by the primary radiologist and the final report, with 227 determined errors (0.18%, 95% CI, 0.16 to 0.20). Table 2 shows the distribution of these errors in terms of QA levels along with examples of common types of error. Of these, the majority were levels 2 and 3 (minor error and error of long-term significance but not in the acute setting).

Underreporting and corrections To determine the degree of underreporting of errors, we constructed modified Lorenz plots for all errors, all errors of any clinical significance (level 3 or more), and clinically significant errors in the acute setting (level 4 or more). Figure 2 shows that some clients reported no errors of any type while others reported more than the average. The estimate of the true error rate derived from the slope of the terminal portion of the curve was approximately 1.5% for all errors, 1% for errors of any clinical significance, and 0.2% for high clinical impact errors. These were much higher than the uncorrected rates (0.18%, 0.1%, 0.05%, respectively; $p < 0.001$). Underreporting appeared to be higher for errors of lower levels than high clinical impact errors as judged by the degree of concavity (Fig. 2b). This was also supported by the lower fold increase due to correction for high clinical impact errors compared to all errors (corrected/uncorrected; eightfold for all errors, tenfold for errors of some clinical significance, fourfold for high clinical impact errors).

Inter-observer variation To determine whether error rates were different between radiologists, we examined the type and rates of errors for each radiologist. The variation was not significant suggesting that the findings broadly apply to the entire group (Fig. 3).

Body part effect To determine whether error rates were different between anatomic body parts, we examined the proportion of errors for each body part vs. the proportion of studies (Fig. 4). CT head studies were the most common

Table 2 Illustration of the QA discrepancy grading system

Study	Level 1	Level 2	Level 3	Level 4	Level 5
227 errors (0.18%)	8	80	87	49	3
	Less than 0.01%	0.06%	0.07%	0.04%	Less than 0.01%
CT Head	Old right parietal lobe infarct rather than left parietal infarct	Maxillary sinus disease	Enlarged pituitary	Thin acute interhemispheric subdural hematoma	
		Chronic infarctions involving the inferior left cerebellar hemisphere and the left posterior frontal convexity	Ruptured dermoid cyst	3-mm right frontal epidural hematoma	
		Calcified granulomas in the brain without associated edema	Right frontal hypodensities, rule out demyelination or ischemia	Left frontal 4-mm depressed skull fracture	
		Small meningioma	Subtle right temporal bone fracture	3–5-mm acute subdural hematoma along the left cerebral convexity. No mass effect	
		Choroid fissure cyst	Small 3-mm aneurysm of the anterior communicating artery		
CT Abdomen/pelvis		Mild adrenal hyperplasia	Left ventricular pseudoaneurysm	Asymmetric cecal wall thickening near the terminal ileum region	Moderate-to-severe hemoperitoneum is noted with associated active hemorrhage
		Small loculated fluid collection in the ventral abdominal wall.	Partly included right femoral lesion with a sclerotic rim.	Larger right psoas muscle with decreased attenuation, which may represent non-acute hematoma	
		T7/T8 and –T8/T9 osteophytes and mild central spinal canal stenosis	Mild hepatosplenomegaly	Pyelonephritis	
		Duplicated left renal collecting system	Possible early cholecystitis	Small pneumoperitoneum	
		Pars interarticularis defects at L5 bilaterally	Ureteral calculus with mild	hydroureteronephrosis	
			Small clot in the left common iliac vein		
CT chest	Aneurysm of the descending thoracic aorta erroneously mentioned as infrarenal aorta	Pulmonary nodule with hypodense foci and calcification, which may represent a hamartoma	Chronic pulmonary thromboembolism	Subtle vertebral fracture	Pulmonary thromboembolism including central arteries
			Pancreatic neoplasm	Solitary subsegmental pulmonary embolus	
			8-mm pulmonary nodule		
			Mild cardiomegaly with pulmonary artery hypertension		
CT spine		Disc herniation mentioned at an incorrect level	Moderate canal stenosis due to degenerative changes	Non-displaced fracture C7 transverse process and C6/C7 facet joint	
			Mild non-acute appearing C4 compression	Mildly impacted fracture of the left clavicular head	
			Mild rotatory subluxation of C1 on C2 could be positional or trauma related		

Representative errors for each QA grade for common study types are shown and were associated with significantly lower error rates. CT chest and abdomen/pelvis studies were associated with the highest error rates. Low-frequency studies such as CT of extremities and maxillofacial region as well as ultrasound, MRI, nuclear medicine scans, and plain radiographs (categorized as "other CTs" and "other modalities") were not associated with increased overall error rates. While an increased proportion of high-grade errors can be seen for other modalities, this was not statistically significant when compared to the overall rate.

Discussion

In this study, we report for the first time the error rates in an international teleradiology practice. Distribution of workload across global time zones has already revolutionized fields like medical transcription and enabled 24-h support for critical applications. Our results show that it can be successfully applied to diagnostic radiology with low error rates. In our study of over a hundred thousand cases, there were only two potentially critical errors where patient care in the acute setting could have been affected to a large degree (level 5), and there were less than 0.2% errors overall. One of the two level 5 errors may have been avoided had adequate clinical information been available in the emergent setting. This was not a case of a missed finding, rather, it was a case of hemoperitoneum with clots, which was thought to represent a mass with ascites as the history of drop in hematocrit was not available. This emphasizes the fact that inadequate history can lead to significant interpretation errors. While we excluded level 0 discrepancies from this study, i.e., where there was no error after review by the QA team, inclusion of such discrepancies would have had only a minor effect (0.26%).

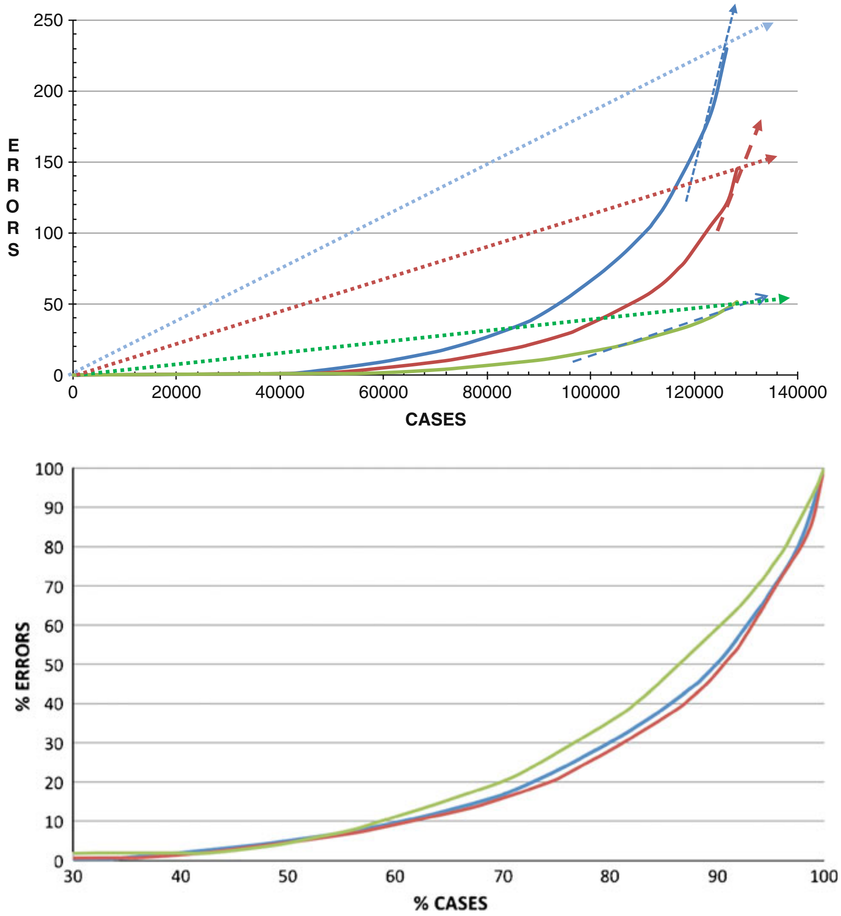

Fig. 2 Underreporting of errors appears to be highly prevalent in feedback-based quality assurance processes in teleradiology. A cumulative plot of the errors vs. cases was arranged such that errors are added in ascending order of reported error rates (Top). The slope of the *dotted line* starting from origin (0, 0) represents the uncorrected error rate while the *dashed line* parallel to the terminal portion of the curve represents the corrected error rate. The *blue line* depicts all errors, the *red line* depicts errors of any clinical significance (level 3 or more), and the *green line* depicts errors of acute clinical significance (level 4 or more, high-impact errors). It can be seen that there is marked concavity of the initial part of the plot suggesting that there is underreporting of errors (see Fig. 1). This is less for high-impact errors (*green line*) as seen in the bottom graph, after eliminating all the non-reporting hospitals (zero errors reported) and normalizing the data as percentage

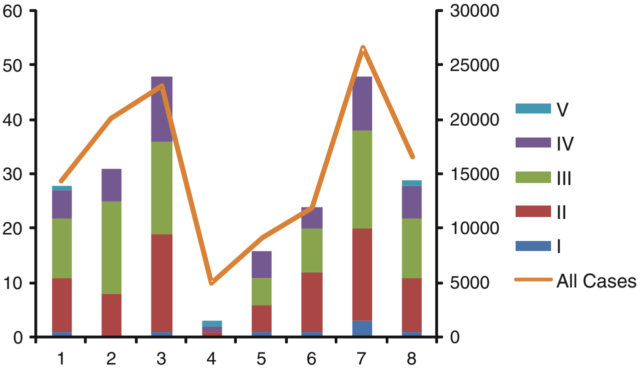

Fig. 3 Inter-radiologist variation. Distribution of errors is shown as stacked columns (left y-axis) for each radiologist (x-axis). The total number of cases read by each radiologist is shown by the *brown line* (right y-axis). Mostly, the number of errors is proportionate to the number of cases read and there is no statistically significant difference in error rates or type

The biggest concern in any estimation of error rates based on retrospective review is the degree of review and feedback. Our practice is unique in that our reports are preliminary and are reviewed by the on-site radiologists who then generate a final report the following day; however, while feedback is actively solicited, there is likely to be incomplete reporting, especially regarding minor errors. Our data supports that there is an underreporting of minor errors. First, level 1 errors were very infrequent and less than other levels, as shown in Table 2. This seems unlikely. Second, plots of the reported errors shows a pattern highly suggestive of underreporting (Fig. 2). More importantly, the pattern suggests a lower level of underreporting for clinically significant errors compared to clinically insignificant errors and allows corrected estimates to be made. We are not aware of any previous use of such analysis for this purpose. The underlying assumption for this method is that real error rates are mostly independent of client hospitals and that, at least, some hospitals are diligent enough to provide complete feedback regarding errors. In an emergency practice like ours with a stable radiologist base and case mix, the first assumption is likely to be valid over the large sample sizes reported in this study. An exception to this rule could be pediatric hospital clients since it is possible that general radiologists may have higher error rates while reading pediatric cases. Indeed, we noticed higher reported discrepancy rates for one of the two such hospitals but did not exclude them from the analysis since the magnitude of this effect is likely to be small compared to underreporting. This may result in a slightly upward bias to our corrected

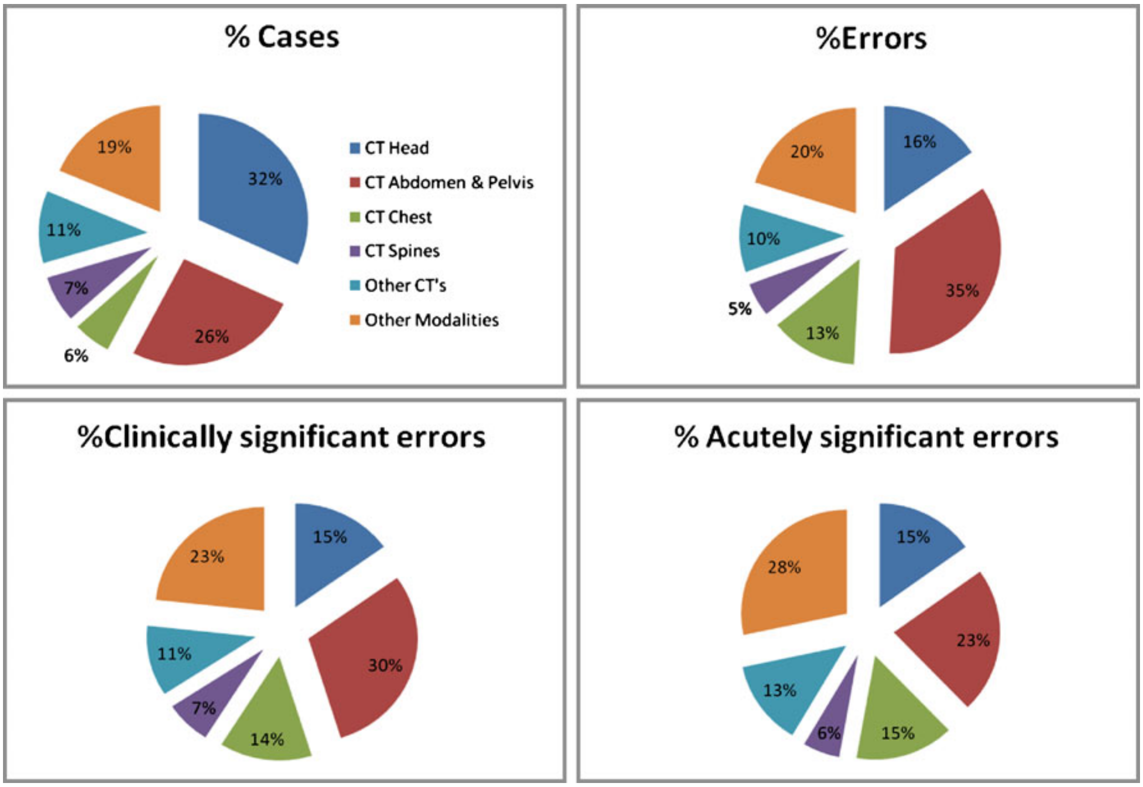

Fig. 4 Different study types are associated with different error rates. Distribution of errors and reads are shown as a fraction of the total for each major anatomic region for CTs and jointly for non-CT studies (US/NM/MRI/XR). While head CTs comprise the largest number of studies (32%), they account for only 16% of the errors. In contrast, CT chest and abdomen/pelvis studies account for 6% and 26% of the studies, respectively and are associated with 13% and 35% of the errors ($p<0.0001$)

estimates; however, the second assumption is likely to be true only for errors of high clinical significance and even the most diligent client hospital may provide less than complete feedback. This would result in undercorrection, more so for errors of low clinical significance. Within these limitations, the premises of the described method are sound and have been found to be reproducible by repeated mathematical simulations, some of which are shown in Fig. 1. To further confirm, we used corrected estimates of the error rates and compared them to the findings of an internal review where 200 cases were rigorously second-read by one of us, discrepancies noted and graded. The corrected error rates and the findings of the review were consistent with clinically significant errors (level 3 or more; 1% vs. 1.5%, respectively; p=NS) and were much higher than uncorrected rates (0.1%, $p<0.001$). Furthermore, no errors of acute clinical significance (level 4 or more) were found in the 200 randomly selected cases. This is consistent with the very low corrected error rate for such errors (0.2%, one error in 500 cases). The small sample size of the internal review limits further generalization; however, there were many more errors of levels 1 and 2 such as typographic mistakes and findings of no clinical significance such as simple cysts, small visceral hemangiomas and other “incidentalomas” (total error rate including errors of no clinical significance; internal review vs. corrected error rate, 10% vs. 1.5%; $p<0.01$). This indicates that clinically insignificant errors are common in preliminary reporting in the emergent setting and that clients do not provide feedback regarding such errors. This is not unexpected from busy clinical radiology practitioners working in the ER setting.

Interestingly in our study, the error rate was significantly different across anatomical regions. There was a higher error rate for CT chest or abdomen/pelvis compared to the pooled estimate while head CTs had the lowest error rate. This may relate to the lower number of images in head CTs compared to CT abdomen/pelvis or chest. There was no significant independent or additional effect of radiologists or workload of the radiologist on the error rate. While this was somewhat surprising, this may be explained by the fact that all radiologists had extensive experience post-residency. Also, the relatively small number of errors overall limit the statistical power to detect subtle differences.

In summary, in our experience international teleradiology is associated with very low rates of clinically significant errors and the majority of such errors do not have any impact in the acute setting. Even after rigorous correction for possible underreporting, the frequency of misses that have any kind of clinical impact in the acute setting is less than one in 500 cases. Errors that posed any sort of risk to life were virtually non-existent over the large series of over 100,000 cases. Whether this low error rate is actually lower than US-based teleradiology operations is more difficult to judge. While our overall discrepancy rate of 0.26% is favorable compared to a similar study from US-based teleradiologists (1.09% discrepancy rate, reference [6]), there is significant underreporting bias that makes comparison difficult. We speculate that previous studies must also be affected by underreporting bias, being feedback dependent. Therefore, while prima facie there seems to be a benefit in allowing qualified radiologists to read during daytime hours by deployment across time zones based solely on overall low discrepancy rates; this certainly merits additional evaluation. There appears to be very limited feedback for errors of little or no clinical significance and internal review is necessary to correctly estimate such error rates. Encouraging and actively obtaining direct feedback from clients is critical for quality assurance and improvement of any teleradiology program.

Acknowledgments We thank Kiran KC and Naveen Kumar K for maintaining an excellent QA database.

References

1. Kalyanpur A, Weinberg J, Neklesa V, Brink JA, Forman HP (2003) Emergency radiology coverage: technical and clinical feasibility of an international teleradiology model. Emerg Radiol 10:115–118
2. Wagner AL (2004) After-hours coverage: problems and solutions. J Am Coll Radiol 1(5):351–355
3. Bradley WG (2004) Offshore teleradiology. J Am Coll Radiol 1 (4):244–248
4. Gander P, Purnell H, Garden A, Woodward A (2007) Work patterns and fatigue-related risk among junior doctors. Occup Environ Med 64:733–738
5. Ebbert TL, Meghea C, Iturbe S, Forman HP, Bhargavan M, Sunshine JH (2007) The state of teleradiology in 2003 and changes since 1999. AJR 188:W103–W112
6. Wong WS, Roubal I, Jackson DB, Paik WN, Wong VK (2005) Outsourced teleradiology imaging services: an analysis of discordant interpretation in 124, 870 cases. J Am Coll Radiol 2(6):478–484
7. American College of Radiology. ACR standard for teleradiology. Available at www.acr.org. Accessed 11 Jan 2010
8. Lorenz MO (1905) Methods of measuring the concentration of wealth. Publ Am Stat Assoc 9(New Series, No. 70):209–219