ORIGINAL ARTICLE

Novel Method to Improve Radiologist Agreement in Interpretation of Serial Chest Radiographs in the ICU

*Denise A Castro, Asad A Naqvi[1], David Manson, Michael P Flavin[1],
Elizabeth VanDenKerkhof[2], Donald Soboleski[1]*

Department of Imaging, University of Toronto, Toronto, [1]Department of Diagnostic Radiology, [2]Department of Anesthesiology and Perioperative Medicine, School of Nursing, Queen's University, Kingston, Ontario, Canada

Address for correspondence:
Dr. Donald Soboleski,
Department of Diagnostic Radiology,
Queen's University, Kingston,
Ontario - K7L 2V7, Canada.
E-mail: daas@queensu.ca

ABSTRACT

Objectives: To determine whether a novel method and device, called a variable attenuation plate (VAP), which equalizes chest radiographic appearance and allows for synchronization of manual image windowing with comparison studies, would improve consistency in interpretation. **Materials and Methods:** Research ethics board approved the prospective cohort pilot study, which included 50 patients in the intensive care unit (ICU) undergoing two serial chest radiographs with a VAP placed on each one of them. The VAP allowed for equalization of density and contrast between the patients' serial chest radiographs. Three radiologists interpreted all the studies with and without the use of VAP. Kappa and percent agreement was used to calculate agreement between radiologists' interpretations with and without the plate. **Results:** Radiologist agreement was substantially higher with the VAP method, as compared to that with the non-VAP method. Kappa values between Radiologists A and B, A and C, and B and C were 46%, 55%, and 51%, respectively, which improved to 73%, 81%, and 66%, respectively, with the use of VAP. Discrepant report impressions (i.e., one radiologist's impression of unchanged versus one or both of the other radiologists stating improved or worsened in their impression) ranged from 24 to 28.6% without the use of VAP and from 10 to 16% with the use of VAP ($\chi^2 = 7.454$, $P < 0.01$). Opposing views (i.e., one radiologist's impression of improved and one of the others stating disease progression or vice versa) were reported in 7 (12%) cases in the non-VAP group and 4 (7%) cases in the VAP group ($\chi^2 = 0.85$, $P = 0.54$). **Conclusion:** Numerous factors play a role in image acquisition and image quality, which can contribute to poor consistency and reliability of portable chest radiographic interpretations. Radiologists' agreement of image interpretation can be improved by use of a novel method consisting of a VAP and associated software and has the potential to improve patient care.

Key words: Chest radiograph, ICU consistency in interpretation, image acquisition, inter-observer agreement, variables in radiodensity and contrast

Access this article online

Quick Response Code:

Website:
www.clinicalimagingscience.org

DOI:
10.4103/2156-7514.161848

INTRODUCTION

Inter-observer variation in interpretation of chest radiographs by clinicians is recognized and has been well documented, particularly in the ICU setting.[1-3] Reliability in interpretation of these studies is crucial to guide appropriate timely patient care.[4-7] It also helps to ensure reproducibility in clinical research studies so as to reduce sample size requirements and allow true-positive trial findings. Utilization of standardized reporting criteria, proctoring, and double readings have shown to improve agreement of image interpretation, helping to address issues such as interpreter experience and fatigue.[5,6] Radiographic image quality and the techniques used to obtain the images also contribute to the physicians' ability to reliably interpret and compare a patients' chest radiographs.[4,5,8-10] Inherent differences in the patients' lung volumes and body habitus alter exposure leading to a variable appearance of the chest radiograph, making it more difficult to assess for changes between studies.[5,6] Numerous technical factors including patient positioning, exposure settings, overlying support apparatus, etc., can also alter the appearance of the radiograph contributing to inter-/intra-variation in interpretation. There are few data addressing these image acquisition difficulties, which will contribute to the poor inter-/intra-observer variation in chest radiograph reporting. We therefore, conducted an observer agreement study with the use of a novel device placed on the X-ray cassette that would help to correct and balance exposure differences between radiographs. The purpose of this study is to determine if the use of this novel device would help to improve radiologists' agreement in interpretation when comparing results.

MATERIALS AND METHODS

The pilot study was approved by the university and affiliated teaching hospitals' Research Ethics Board (REB) and informed consent from the patients was waived by the REB.

Study design

Study location
The study was conducted in the ICU of our tertiary care 456-bed university affiliated teaching hospital.

Patient population
This prospective cohort study included 50 non-consecutive patients (age range: 17 days to 85 years; 24 males, 26 females) in the ICU and NICU, who were undergoing anterior–posterior chest radiographs as part of their routine care. Of note, to take into account similarity in lung volumes, 29/50 of the ICU patients were intubated.

Inclusion criteria
- Patients who had undergone two chest X-rays (CXRs) during their ICU stay with adequate placement of the variable attenuation plate (VAP)
- Patient requisition history was query edema (31/50) with underlying cardiac decompensation, pneumonia (14/50) with underlying chronic obstructive pulmonary disease (COPD), or both edema and pneumonia (5/50). There were four cases of neonates underlying surfactant disease with a query of fluid and/or pneumonia.

Exclusion criteria
- Radiographs that had the VAP completely or partially superimposed by body parts and/or support apparatus
- Radiographs with the VAP not completely included on exposure.

Description of device (VAP)
The VAP is a metal plate made of one or more layers of 0.1-inch-thick brass mounted on a 0.0625-inch aluminum base. For this study, a 2 cm × 2 cm square shape or a 1 cm × 4 cm strip configuration was utilized for the VAP. Both VAP configurations consist of four 1 cm squares. Each square within the VAP is of a different thickness due to the number of layers of brass in each square. Upon radiographic exposure, the VAP will thus be of four different radiodensities correlating with the different thicknesses of the brass layer [Figure 1]. The VAP device is not commercially available and was developed by individuals with the expertise on special request.

Description of software
The VAP software assumes a linear model in the remapping of the radiodensities and contrasts between the two radiographs being compared by using matching densities of the VAP to estimate the linear relationship. It uses the least square minimization of the four radiodensities of the VAP after placement of a region of interest (ROI)
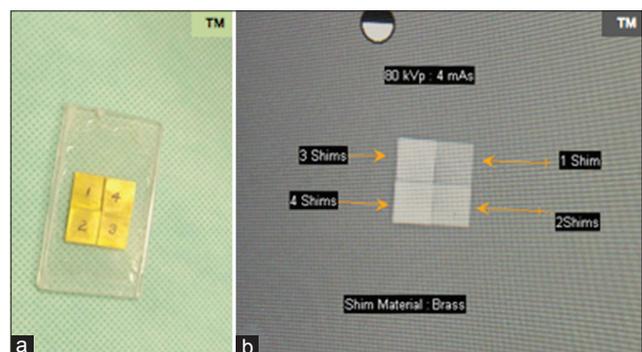
Figure 1: (a) Square VAP mounted on acrylic. (b) Its radiographic exposure demonstrating the four different densities according to the thickness of the VAP.

within each square. The ROI used is 30 × 30 pixels to allow the best average and the least noise. The software evaluates the linear correlation between the two sets of radiodensities in order to validate the assumption of linearity. A high value of correlation (close to one) indicates the radiodensities are distributed linearly and therefore validate the methodology since the calibration of the images was independent of the density used. The software is not commercially available and was developed on special request by individuals with expertise in this field.
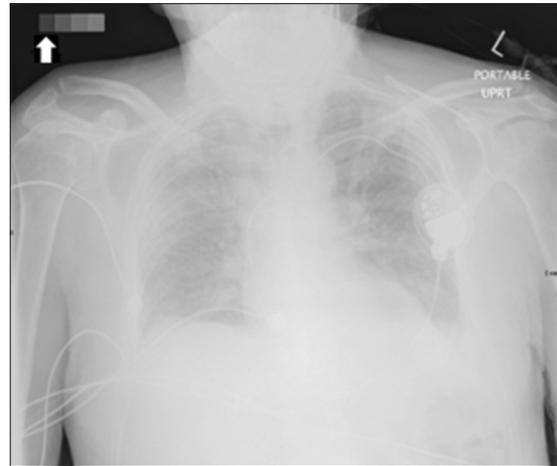
Chest radiograph acquisition

A short training session was conducted for five technical staff to explain the importance of adequate/correct placement of the VAP. The staff were instructed to place the VAP similar to how a right–left marker is placed, but to ensure it does not overlap with body parts or support devices [Figure 2]. A "handle" for the VAP was developed to allow easier transport and placement on the cassette [Figure 3]. The patients in whom the VAP was placed were randomly selected; however, as per inclusion criteria, only patients with a history of query pneumonia ± edema were utilized for this study. The radiographic technique was otherwise unaltered. The radiographs were obtained at 90 kVp and 4 mA in the ICU and at 60 kVp and 1.5 mA for the NICU patients, as per the usual technique in our facility. The chest radiographs were incorporated into the hospital Picture Archiving and Communication System (PACS) and utilized/reported in the usual fashion.

CXR preparation for interpretation

Two serial frontal chest radiographs obtained from the 50 patients, who met the above criteria, were randomly placed in the viewer system on our PACS. The cases were anonymized and labeled 1a/b through 50a/b. The earlier study was placed to the left of a more recent study.

The same 50 sets of frontal radiographs were also randomly placed in a second identical viewer system also on the radiologists' PACS in which the specifically designed software had been added. The software allowed for the equalization of the densities and contrasts of the matching squares on the VAPs between the previous and more recent radiographs, resulting in an equalization of chest radiograph appearance. The software also allowed the synchronization of the windowing of the patients' previous and more recent chest radiographs during radiologist assessment. The software required manual placement of a cursor within each square of the VAP, which was performed by a radiologist not involved with the interpretations [Figure 4].

Figure 2: Anterior–posterior chest radiograph in upright position, with the VAP appropriately placed in the upper right corner of the cassette (Arrow), separate from patient's body and support.

Figure 3: Rectangular VAP with handle.

Figure 4: VAP software demonstrates the method in which each segment of the VAP is matched to the corresponding segment on a comparison radiograph, allowing automatic equalization of density and contrast between the studies.

Interpreters

Three trained thoracic radiologists having between 3 and 30 years experience were recruited to evaluate the studies. The interpreting radiologists had no conflicting commercial interest in the development of the VAP or associated software.

Chest X-ray reporting process

The thoracic radiologists were then instructed to dictate each of the 50 cases in each viewer in their usual fashion

via hand-held recorder. The radiologists were instructed to add a final impression to their dictation, stating whether they thought the more recent radiograph showed improvement, disease progression, or no change compared to the earlier radiograph. No clinical history was provided. The radiologist reported a batch of 25 of the 50 cases randomly placed in one viewer, and then 25 of the 50 cases randomly placed in the other viewer during a single day. The following day, the remaining 25 cases that were randomly placed in each viewer were dictated, beginning in the viewer used last on the previous day. The "dictation tapes" from each radiologist were then sent to a secretary for transcription. The time taken to dictate each case as well as the total time needed to report all the cases were recorded.

Statistical methods

We were interested in finding the level of agreement in the interpretation of the chest radiographs between the three radiologists and whether the VAP allowed for improvement in agreement. The study was designed to assess radiologist agreement as opposed to reliability due to the lack of a Gold Standard, and accuracy, *per se*, was not deemed relevant. Observed agreement between the non-VAP and VAP methods and between the unweighted and linear weighted chance-corrected agreement (kappa) were calculated for each radiologist (intra-observer) and between radiologists (inter-observer). Kappa ranged from 0 to 1 (i.e., 0–100%), with a value of 1 implying perfect agreement among interpreters. It is suggested that values above 0.75 reflect excellent agreement and values less than 0.4 reflect poor agreement.[11]

Frequency of "discrepant" impressions between the non-VAP and VAP groups, i.e., at least one of the radiologists giving an impression of no change versus the other radiologists indicating improvement or disease progression of the condition, and that of "opposing" impressions between the non-VAP and VAP groups, i.e., a radiologist having an impression of disease progression versus the other radiologists indicating an improved condition or vice versa, were calculated for each radiologist and the overall value was also calculated. The number of "discrepant" and "opposing" impressions, after interpretation of all cases with and without using VAP, were calculated between radiologists. Chi square was used to assess for statistically significant differences between discrepant or opposing reports for the non-VAP and VAP groups. Mean time (95% confidence interval) to complete the batch of 50 views was calculated for each radiologist and the overall time was also calculated. Paired *t*-tests were conducted to assess for statistically significant differences in time

to completion. Radiologists were also asked to report on the perceived time to complete each method and their confidence in their impressions using each method. Statistical significance was set at 0.05. SPSS version 21 was used to conduct the statistical analyses.

RESULTS

Of the 50 sets of CXRs, one set of one radiologist was discarded due to the inability of the transcriptionist to adequately hear the report. The intra-observer weighted kappa values suggested moderate to very good agreement between the non-VAP and VAP methods, which ranged from 63% to 86% for all three radiologists [Table 1]. Intra-observer discrepant reports ranged from 12% to 20% with only one radiologist having an opposing impression between methods. The inter-observer weighted kappa for the non-VAP method ranged from 52% to 66%, while the weighted kappa for the VAP method ranged from 71% to 84% [Table 2]. Inter-observer discrepant reports ranged from 24% to 29% when using the non-VAP method and from 10% to 16% when utilizing the VAP method. The total number of discrepancies was 40 for the non-VAP method (vs 108 non-discrepant impressions) and 21 on using the VAP method (127 non-discrepant impressions) ($\chi^2 = 7.45$, $P < 0.01$). There were a total of seven opposing reports when using the non-VAP method, as compared to four when using the VAP method ($\chi^2 = 0.85$, $P = 0.54$).

Table 1: Intra-rater agreement (within radiologist) between interpretations done with and without a VAP		
Radiologist	Number of pairs	% agreement between non-VAP and VAP+ (95% confidence interval)
Observed agreement		
Radiologist A	49	86 (72-94)
Radiologist B	50	74 (59-85)
Radiologist C	50	88 (75-95)
Unweighted chance-corrected agreement (Kappa)		
Radiologist A	49	77 (62-93)
Radiologist B	50	61 (42-79)
Radiologist C	50	81 (67-95)
Linear weighted chance-corrected agreement (Kappa)		
Radiologist A	49	83 (71-95)
Radiologist B	50	63 (44-82)
Radiologist C	50	86 (75-97)
Discrepant reports^		
Radiologist A	49	7 (14.2)
Radiologist B	50	10 (20.0)
Radiologist C	50	6 (12.0)
Opposing reports#		
Radiologist A	49	0 (0.0)
Radiologist B	50	3 (6.0)
Radiologist C	50	0 (0.0)

+Rating scale: 1=Better, 2=No change, 3=Worse, ^Discrepant=Difference of one category, i.e., from better to no change and vice versa or from worse to no change and vice versa, #Opposing=Difference of two categories, i.e., from better to worse and vice versa. VAP: Variable attenuation plate

Although Radiologist B and Radiologist C were significantly slower and faster dictating each particular case, respectively, using the VAP, there were no statistically significant differences when data from all three radiologists were combined (44.6 s vs 44.5 s; P = 0.93) [Table 3]. The time spent speaking into the "mike" for each case was similar. The overall total time (time spent speaking on-mike and the pauses related to manipulating the image on the PACS and deciding what to dictate) required by each of the three radiologists to dictate all 50 cases was less using the VAP method (Radiologist A: 100 min vs 71 min; Radiologist B: 80 min vs 75 min; Radiologist C: 110 min vs 85 min).

Radiologists were asked to self-report any perceived differences in the speed to complete the reporting of the

batch of cases and confidence in their impressions using one method versus the other. Two felt they completed the VAP batch of cases faster, while the third radiologist felt there was no difference in completion time. None of the three radiologists reported feeling more confident in their impressions with one method versus the other.

DISCUSSION

Reliable chest radiograph interpretation remains an essential requirement in the functioning of any radiology department. Multiple studies have illustrated significant inter- and intra-observer variability in the interpretation of chest radiographs, which impacts patient management. In the clinical trial setting, this lack of agreement in image interpretation results in erroneous or misleading findings and decreases the power and accuracy of a research study.[4-6,8-10]

Factors such as interpreter experience, personality type, and time constraints have an effect on the radiologist's interpretation. In order to address this issue, the development of standardized reporting forms, proctoring/pilot-testing of the observers, and double-reading of images have been shown to improve agreement in interpretation of studies.[2]

Radiologists and intensivists involved with ICU reporting have known that image quality and reproducibility have been a limiting factor in the reliability of interpretations, particularly when comparing patients' serial chest radiographs.[4-7] Studies have demonstrated a sub-optimal chest radiograph appearance in up to one-third of cases and poor correlation with autopsy findings.[5,6,12] Patient positioning relative to the cassette, shorter beam distances than recommended, overlying support apparatus, and changes in body habitus are some of the factors resulting in variability of the chest radiograph appearance contributing to the inter-/intra-observer variability.[4] There is little data available addressing potential solutions to this inherent image acquisition problem, which is encountered numerous times daily in most hospital ICUs.[13-18]

In this study, we assessed the inter- and intra-observer variability of chest radiograph interpretation, comparing the usual method with a novel method, which involves placing of a VAP on the cassette. The rationale of the VAP is straightforward. The positioning of the VAP away from the patient and any overlying/equipment allows for an exposure, which is unaltered by other factors. The appearance of the VAP should thus be equal between serial radiographic studies. The associated software within the viewer system allows for each quadrant of the VAP

Table 2: Inter-rater agreement (between radiologists) on interpretation of chest radiographs using the non-VAP and VAP methods

Method	% Agreement (95% confidence interval)[+]		
	Radiologists A vs B (n = 50)	Radiologists A vs C (n = 50)	Radiologists B vs C (n = 50)
Observed agreement			
Non-VAP	65 (50-78)*	71 (57-83)*	68 (53-80)
VAP	82 (68-91)	88 (75-95)	78 (61-89)
Unweighted chance-corrected agreement (Kappa)			
Non-VAP	46 (26-67)*	55 (36-74)*	51 (31-70)
VAP	73 (56-89)	81 (67-95)	66 (47-85)
Linear weighted chance-corrected agreement (Kappa)			
Non-VAP	52 (32-72)*	66 (50-81)*	54 (34-74)
VAP	75 (58-91)	84 (70-97)	71 (53-88)
		n (%)	
Discrepant reports^			
Non-VAP	14 (28.6)*	14 (28.6)*	12 (24.0)
VAP	7 (14.0)	5 (10.0)	9 (16.0)
Opposing reports#			
Non-VAP	3 (6.1)*	0 (0.0)*	4 (8.0)
VAP	2 (4.0)	1 (2.0)	1 (2.0)

[+]Rating scale: 1=Better, 2=No change, 3=Worse, *n=49, ^Discrepant=Difference of one category, i.e., from better to no change and vice versa or from worse to no change and vice versa, #Opposing=Difference of two categories, i.e., from better to worse and vice versa. VAP: Variable attenuation plate

Table 3: Mean time (95% confidence interval) required to review and dictate impressions for non-VAP versus VAP method

Radiologist	Method	Time (seconds)		Paired t-test values	P value
		Mean	95% confidence intervals		
A	Non-VAP	51.7	49.1-54.2	−0.37	0.71
	VAP	52.2	49.2-55.1		
B	Non-VAP	40.0	37.1-42.8	−2.49	0.02
	VAP	45.5	42.0-48.9		
C	Non-VAP	42.2	37.9-46.5	2.96	0.01
	VAP	35.9	33.5-38.3		
Overall	Non-VAP	44.6	42.6-46.7	0.08	0.93
	VAP	44.5	42.5-46.5		

VAP: Variable attenuation plate

that is placed on one radiograph to be "matched" to the corresponding quadrant of the VAP on the radiograph being compared. Both radiodensity and contrast can be equalized resulting in a more similar appearance to the radiographs being compared and likely resulting in less image manipulation by the individual radiologist. It is hypothesized that this image equalization and decrease in image manipulation would make it more likely that the reviewers are basing their interpretation on the same quality/appearance of the radiograph image. The study indicated an improved agreement in interpretations between radiologists when using the new method and more importantly decreased the percentage of "opposing" interpretations, which would, more likely, result in altered patient management.

Limitations

Our pilot study has several limitations. A Gold Standard with which to compare the interpreter's impressions was not possible in this study. As the study was designed to assess the radiologists' agreement with each other, the accuracy *per se* was not regarded as relevant. The variability of a test is measured by its reproducibility. This does not assure the accuracy of the test, as a reproducible test can reproduce imprecise results. The three interpreters were all thoracic trained sub-specialist radiologists, and the inter-observer variability was regarded as a limitation of the exam and not related to radiologic expertise.

A significant limitation is the time-consuming manual placement of the software cursor into each quadrant to allow for image equalization. An automated process to perform this function is believed to be possible. Another limitation was the inadequate placement of the VAP on the cassette by the technician. A short training session explaining the importance of placement resulted in marked improvement in compliance. This issue could be addressed by incorporating the VAP into the cassette.

A certain limitation in our study was the lack of data in regards to the body mass index (BMI) or body habitus. Lack of optimal lung volumes is an inherent feature of many radiographs performed in the ICU and contributes to the poor inter-/intra-observer variability. Both techniques (with and without VAP) were assessed with the identical inherent limitations.

A further limitation is the relative small study numbers. The dictating of "sets" of 25 may have allowed for patient recognition and a bias limiting the study. This was partially minimized by alternating the dictation of VAP and non-VAP "sets" and the random distribution of batches of cases into each viewer system. The study was performed at a single center with trained sub-specialists, which may not be equivalent to other hospitals with a different mix of interpreting staff and differing underlying pathology. Further studies performed on specific age groups or specific conditions (cystic fibrosis, interstitial lung disease, etc.) would be of interest and we are in the preliminary stages of performing these studies. Due to the relative small numbers, the calculations were not stratified into patients with edema, pneumonia, or both. The VAP and associated software allowed for synchronization of image "windowing" when comparing the patients' serial radiographs. This may have contributed to the decrease in time needed by the radiologists to dictate the entire batch of 50 cases using the VAP method.

CONCLUSION

Our study illustrates a relatively simple, straightforward method and device (VAP) to improve interpreters' agreement of chest radiographs by addressing the inherent chest image acquisition difficulties. Improved agreement in image interpretation should allow more confident diagnosis and has the potential to optimize patient care in the ICU. It will also aid clinical endeavors relying on this modality to filter patients into their appropriate study arm.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Lakhal K, Serveaux-Delous M, Lefrant JY, Capdevilla X, Jaber S. AzuRéa network for the RadioDay study group. Chest radiographs in 104 French ICUs: Current prescription strategies and clinical value (the RadioDay study). Intensive Care Med. 2012;38:1787-99.
2. Graat ME, Choi G, Wolthuis EK, Korevaar JC, Spronk PE, Stoker J, et al. The clinical value of daily routine chest radiographs in a mixed medical-surgical intensice care unit is low. Crit Care 2006;10:R11.
3. Meade MO, Cook RJ, Guyatt GH, Groll R, Kachura JR, Bedard M, et al. Interobserver variation in interpreting chest radiographs for the diagnosis of acute respiratory distress syndrome. Am J Respir Crit Care Med 2000;161:85-90.
4. Rubenfeld GD, Caldwell E, Granton J, Hudson LD, Matthay MA. Interobserver variability in applying a radiographic definition for ARDS. Chest 1999;116:1347-53.
5. Tudor GR, Finlay D, Taub N. An assessment of inter-observer

agreement and accuracy when reporting plain radiographs. Clin Radiol 1997;52:235-8.

6. Cleveland RH, Schluchter M, Wood BP, Berdon WE, Boechat MI, Easley KA, et al. Chest radiographic data acquisition and quality assurance in multicenter studies. Pediatr Radiol 1997;27:880-7.

7. Bloomfield FH, Teele RL, Voss M, Knight DB, Harding JE. Inter- and intra-observer variability in the assessment of atelectasis and consolidation in neonatal chest radiographs. Pediatr Radiol 1999;29:459-62.

8. Ioos V, Galbois A, Chalumeau-Lemoine L, Guidet B, Maury E, Hejblum G. An integrated approach for prescribing fewer chest x-rays in the ICU. Ann Intensive Care 2011;1:4.

9. Christiansen F, Andersson T, Rydman H, Qvarner N, Måre K. Rater agreement in lung scintigraphy. Acta Radiolo 1996;37:754-8.

10. Bhagwanjee S, Muckart DJ. Routine daily chest radiography is not indicated for ventilated patients in a surgical ICU. Intensive Care Med 1996;22:1335-8.

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

12. Lichtenstein D, Goldstein I, Mourgeon E, Cluzel P, Grenier P, Rouby JJ. Comparative diagnostic performances of auscultation, chest radiography, and lung ultrasonography in acute respiratory distress syndrome. Anesthesiology 2004;100:9-15.

13. Hause GJ, Pollack MM, Sivit CJ, Taylor GA, Bulas DI, Guion CJ. Routine chest radiographs in pediatric intensive care: A prospective study. Pediatrics 1989;83:465-70.

14. Chahine-Malus N, Stewart T, Lapinsky SE, Marras T, Dancey D, Leung R, et al. Utility of routine chest radiographs in a medical-surgical intensive care unit: A quality assurance survey. Crit Care 2001;5:271-5.

15. Horst HM, Fagan B, Beute GH. Chest radiographs in surgical intensive care patients: A valuable "routine". Henry Ford Hsop Med J 1986;34:84-6.

16. Greenbaum DM, Maeschall KE. The value of routine chest x-rays in intubated patients in the medical intensive care unit. Crit Care Med 1982;10-29-30.

17. Henschke CI, Pasternack GS, Schroeder S, Hart KK, Herman PG. Bedside chest radiography: Diagnostic efficacy. Radiology 1983;149:23-6.

18. Bekemeyer WB, Crapo RO, Calhoon S, Cannon CY, Clayton PD. Efficacy of chest radiography in a respiratory intensive care unit. A prospective study. Chest 1985;88:691-6.