Methods toward improving consistency in interpretation of chest radiographs

Presented by:
Dr. D.A. Soboleski
Associate Professor
Queen's University
Kingston, Ontario

Queen's UNIVERSITY

Numerous studies have documented the poor inter/intra-observer agreement on interpretation of chest radiographs

1668

Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction[1]

Arriving at a medical diagnosis is a highly complex process that is extremely error prone. Missed or delayed diagnoses often lead to patient harm and missed opportunities for treatment. Since medi- ing is a major contributor to the overall diagnostic process, potential source of diagnostic error. Although

Chest. 2012 Feb;141(2):545-558. doi: 10.1378/chest.10-1302

Interpretation of plain chest roe

Author information

Abstract

Plain chest roentgenogram remains the most comm... lower sensitivity demands greater accuracy in interpre... to an optimal and organized approach to interpretation... chest radiograph (CXR) as normal. Clinicians can only ... they read hundreds of normal CXRs. An individual shoul... each time. All clinicians must make a concerted effort to re...

Raoof S[1], Feigin D[2], Sung A[3], Raoof S[4], Irugulpati L[5]

CHEST Postgraduate Education Corner

CHEST IMAGING REVIEW

Interpretation of Plain Chest Roentgenogram

Suhail Raoof, MD, FCCP; David Feigin, MD, FCCP; Arthur Sung, MD; Sabiha Raoof, MD, FCCP; Lavanya Irugulpati, MD; and Edward C. Rosenow III, MD, Master FCCP

Plain chest roentgenogram remains the most commonly ordered screening test for pulmonary disorders. Its lower sensitivity demands greater accuracy in interpretation. This greater accuracy can be achieved by adhering to an optimal and organized approach to interpretation. It is important for clinicians not to misread an abnormal chest radiograph (CXR) as normal. Clinicians can only acquire the confidence in making this determination if they read hundreds of normal CXRs. An individual should follow the same systematic approach to reading CXRs each time. All clinicians must make a concerted effort to read plain CXRs themselves first without reading the radiologist report and then discuss the findings with their radiology colleagues. Looking at the lateral CXR may shed light on 15% of the lung that is hidden from view on the posteroanterior film. Comparing prior films with the recent films is mandatory, when available, to confirm and/or extend differential diagnosis. This article outlines one of the many systematic approaches to interpreting CXRs and highlights the lesions that

limitations of CXR is also included.

Contents lists available at ScienceDirect

Clinical Radiology

journal homepage: www.clinicalradiologyonline.net

ELSEVIER

Measuring performance in the interpretation of chest radiographs: a pilot study

Y. Chen [a,*], J.J. James [b], L. Dong [a], A.G. Gale [a]

[a] Applied Vision Research Centre, Loughborough University, Loughborough, UK
[b] Nottingham Breast Institute, Nottingham University Hospitals, Nottingham, UK

ARTICLE INFORMATION

Article history:

AIM: To develop a system to assess the image interpretation performance of radiologists in identifying signs of malignancy on chest radiographs.

MATERIALS AND METHODS: A test set of 30 chest radiographs was chosen by an experienced

Image Compression and Chest Radiograph Interpretation: Image Perception Comparison Between Uncompressed Chest Radiographs and Chest Radiographs Stored Using 10:1 JPEG Compression

Douglas P. Beall, Phillip D. Shelton, Thomas V. Kinsey, Maria C. Horton, Brian J. Fortman, Steffen Achenbach, Vadim Smirnoff, Daniel L. Courneya, Bill Carpenter, and John T. Gironda

We have assessed the effect of 10:1 lossy (JPEG) compression on six board-certified radiologists' ability to detect three commonly seen abnormalities on chest radiographs. The study radiographs included 150 chest radiographs with one of four diagnoses: normal (n = 101), pulmonary nodule (n = 19), interstitial lung disease (n = 19), and pneumothorax (n = 11).

technique used routinely at those sites is JPEG 10:1 lossy compression. The images are interpreted and are accessible as uncompressed data before long-term storage, where the images are compressed and stored on optical discs.

Competency in Chest Radiography

A Comparison of Medical Students, Residents, and Fellows

Lewis A. Eisen, MD,[1] Jeffrey S. Berger, MD,[2] Abhijith Hegde, MD,[2] Roslyn F. Schneider, MD[2]

[1]Division of Pulmonary and Critical Care, Beth Israel Medical Center, New York, NY, USA; [2]Department of Internal Medicine, Beth Israel Medical Center, New York, NY, USA.

BACKGROUND: Accurate interpretation of chest radiographs (CXR) is essential as clinical decisions depend on readings.

physicians,[11] and anesthesiologists.[12] Faulty interpretations change management in up to 11% of cases.[10] Most studies,

Measuring and managing radiologist workload: Application of lean and constraint theories and p... to planning radiology services in ...

Sharyn LS MacDonald,[1] Ian A Cowan,[1] Richard Floyd,[2] Stu... and Richard Hamilton[2]

[1]Radiology Department, and [2]Business Development Unit, Christchurch Hospita...

Radiology. 2000 Nov;217(2):456-9.

Measuring performance in chest radiography.

Potchen EJ[1], Cooper TG, Sierra AE, Aben GR, Potchen MJ, Potter MG, Siebert JE.

Author information

Abstract

PURPOSE: To use a standardized set of chest radiographs to quan... basis for comparing the diagnostic performance of physician...

MATERIALS AND METHODS: A standardized set of ... participants. Each participant reviewed the ra...

Chest Radiograph Interp...

D. R. JEFFREY*, P. R. GODDARD*, M. P. CALL...

*Department of Clinical R...

...ccuracy of interpretations ...confidence levels.

...Maythue FE[1], Rust DD, Aldag JC, Jen...

Author information

...ergency department radiog...

J Med Imaging (Bellingham). 2018 Jul;5(3):031402. doi: 10.1117/1.JMI.5.3.031402. Epub 2017 Dec 11.

Influence of radiology expertise on the perception of nonmedical images.

Kelly B[1], Rainford LA[2], McEntee MF[3], Kavanagh EC[4].

Author information

Abstract

Identifying if participants with differing diagnostic accuracy and visual search behavior during radiologic tasks also differ in nonradiologic tasks is investigated. Four clinician groups with different radiologic experience were used: a reference expert group of five consultant radiologists, four radiology registrars, five senior house officers, and six interns. Each of the four clinician groups is known to have significantly different performance in the identification of pneumothoraces in chest x-ray. Each of the 20 participants was shown 6 nonradiologic images (3 maps and 3 sets of geometric shapes) and was asked to perform search tasks. Eye movements were

The Development of Expertise in Radiology: In Chest Radiograph Interpretation, "Expert" Search Pattern May Predate "Expert" ... Diagnostic Accuracy for ...horax Identification[1]

To investigate the development of chest radiograph interpretation skill through medical training by measuring both diagnostic accuracy and eye movements during visual search.

Radiology

Interpretation of Emergency Department Radiographs: A Comparison of Emergency Medicine Physicians with Radiologists, Residents with Faculty, and Film with Digital Display

OBJECTIVE. We determined the relative value of teleradiology and radiology resident coverage of the emergency department by measuring and comparing the effects of physician specialty, training level, and image display method on accuracy of radiograph interpretation. MATERIALS AND METHODS. A sample of four faculty emergency medicine physicians, four emergency medicine residents, four faculty radiologists, and four radiology residents participated in our study. Each physician interpreted 120 radiographs, approximately half containing a clinically important abnormality. Radiographs were interpreted using the original films and high-resolution digital monitors. Accuracy of radiograph interpretation was measured as the area under the physicians' receiver operating characteristic (ROC) curves. RESULTS. The area under the ROC curve was 0.15 (95% confidence interval [CI], 0.10-...

John Eng[1], William K. Mysko[2], Gregory E. R. Weller[1,3], Regis Renard[1,4], Joseph N. Gitlin[1], David A. Bluemke[1], Donna Magid[1], Gabor D. Kelen[2], William W. Scott, Jr.[1]

...ICAL PRACTICE

Clinical Medicine 2013, Vol 13, No 4: 349–52

E. James Potchen, MD

...ng the accuracy and certainty in interpreting chest X-rays ...edical division

...ugha, A Bibi, R Ahmed, S Mellor and F Zaman

ABSTRACT – The chest X-ray (CXR) is an important diagnostic tool in diagnosing and monitoring a spectrum of diseases. Despite our universal reliance on the CXR, our ability to confidently diagnose and accurately document our findings can be unreliable. We sought to assess the diagnostic accuracy and certainty of making a diagnosis based on 10 short clinical histories with one CXR each. We conclude from our study that specialist registrars (StRs) and consultants scored the highest marks with the highest average certainty levels. Junior trainees felt least certain about making their diagnosis and were less

reported, but this may not be available at the time ... medical review on the medical assessment unit. Thi... uncommon scenario in other NHS trusts due to staff ... highlighted by the Royal College of Radiologists rep... *radiology: a workforce in crisis.*[4] Therefore, for all other CXRs ti... local policy states:

In accordance with the requirements of the The Ionising Radiation (Medical Exposure) Regulations (I(R)MER) under local agreement, the reporting responsibility for this examination has been delegated to the referring practitioner whose

Commentary

The Medical Image Perception Society and the Future of Imaging Research[1]

The Medical Image Perception Society (MIPS) is an organization created to promote a better understanding of the role of ... perception. Drs Krupinski, Kundel, Judy, and Nodine provide a succinct overview of some issues currently being addressed by the society. I concur with the ... radiological service. Quality control in diagnostic imaging requires attention to the entire chain of events that results in a process seeking to diminish uncertainty with regard to a patient. This process includes the generation of an image, the interpretation of that image, the interpretation of that perception, and the communi... Figure was derived from those discussions. It serves to point out that the current variability (room for improvement) in image generation is considerably less than that in image perception and communication. From our studies we have seen that variation in image quality is considerably less than the varia...

Reliability in interpretation is crucial to guide appropriate timely care particularly in the very sick population

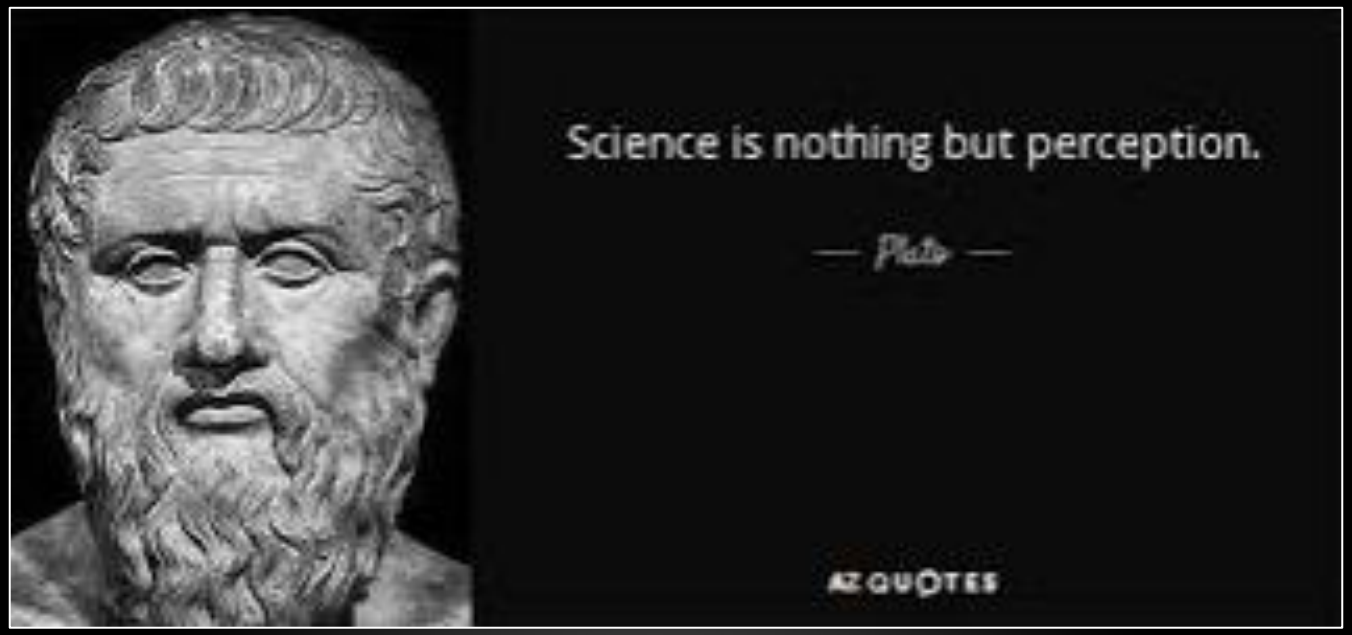
An effect of inconsistent interpretation often overlooked

Many research study protocols rely on the radiographic interpretation to determine which arm of the study a patient may proceed upon.

Reliability helps ensure reproducibility in clinical research studies allowing for a reduced sample size requirements and allow true-positive findings

Why the poor inter/intra-observer agreement?

Is it due to our differences in perception?

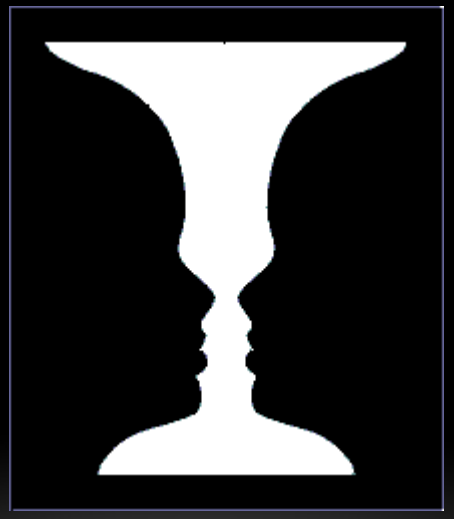


Definition:

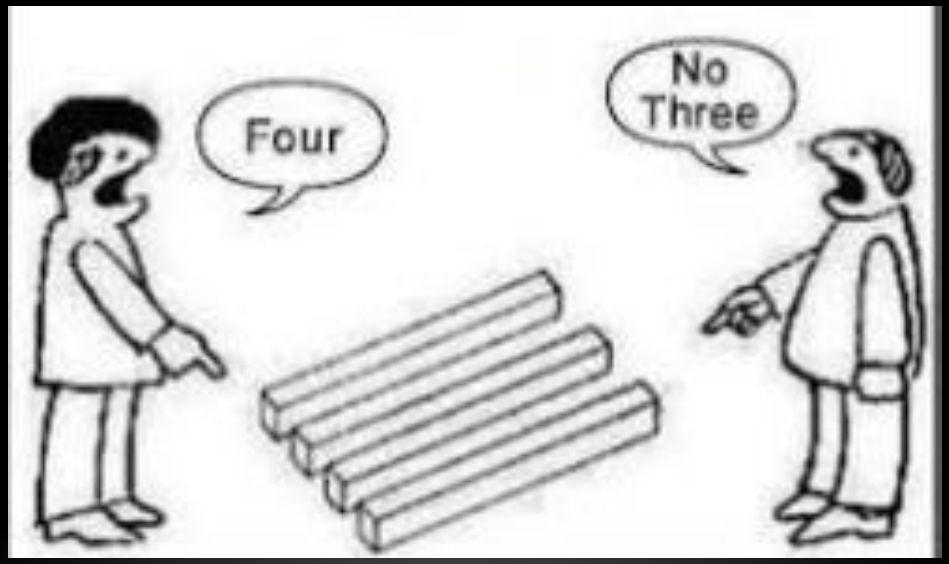
Perception is the process of selecting and interpreting the information we receive thru our senses to produce a meaning and plays a significant role in the interpretation of images

How can we improve perception and consistency?

SOLUTIONS



1. Ensure we and our colleagues are looking at the same image.



Effect of PACS image manipulation on the agreement of chest radiograph interpretation in the NICU

Collaborators

D.Castro- pediatric radiologist

M.Flavin – neonatologist

M.Clarke – neonatologist

J. Flood – thoracic radiologist

J.Gammon – neonatologist/pediatrician

Prospective cohort study- the population

- 60 patients

- gestational age 26-32 weeks

-1 day – 3 months of age

- all with history of surfactant defcy disease

Prospective cohort study

- 2 chest xrays on each pt. performed on different days included

- the 120 xrays (60 pts.x 2) anonymized and numbered

- randomly placed as acquired in two identical viewers

Prospective cohort study

- 3 radiologists - 3 neonatologists

  Radiologists – median length of expertise 14 yr (5-25)

  Neonatologists – median length of expertise 16 yr (2-25)

Prospective cohort study

- Reviewed 2 consecutive xrays on same pt.

  Once in usual manner with ‘windowing’/ image manipulation allowed

  Once again without image manipulation

Effect of PACS image manipulation on the agreement of chest radiograph interpretation in the NICU

only interpretation

is disease appearance

- better
- worse
- unchanged

Results

Assessment of sixty sets of frontal chest radiographs 'without' versus 'with' the ability to manually manipulate the images

	Nonmanipulated (n=60) (%)	Manipulated (n=60) (%)	Relative risk (95% CI)	χ2 (P)
Perfect agreement				
Combined (n=6	8 (13)	3 (5.0)	2.9	2.50 (0.11)
Radiologists (n=3)	25 (42)	21 (35)	1.3	0.56 (0.46)
Neonatologists (n=3)	18 (30)	8 (13)	2.8	4.91 (0.04)

Results

Assessment of sixty sets of frontal chest radiographs 'without' versus 'with' the ability to manually manipulate the images

	Nonmanipulated (n=60) (%)	Manipulated (n=60) (%)	Relative risk (95% CI)	χ2 (P)
Opposing agreement				
Combined (n=6	29 (48)	37 (62)	0.6	2.16 (0.14)
Radiologists (n=3)	12 (20)	17 (28)	0.6	1.12 (0.29)
Neonatologists (n=3)	11 (18)	24 (40)	0.3	6.82 <0.01)

How can we improve perception and our consistency in reporting?

SOLUTION

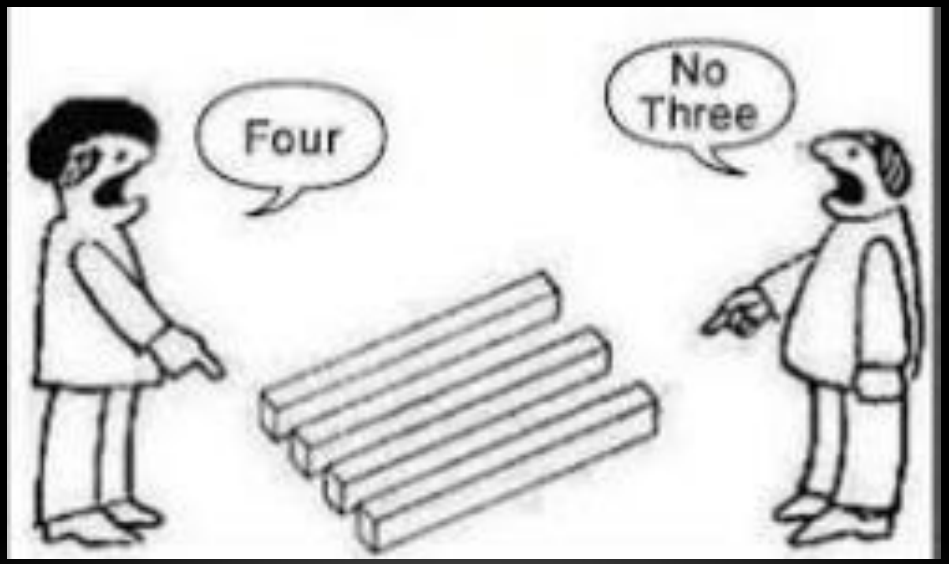
1. Do not allow image manipulation outside imaging the imaging department

How can we improve perception and consistency?

SOLUTIONS

1) Ensure we and our colleagues are looking at the same image

2) Ensure we are looking at the same image



VALUE OF A NOVEL DEVICE AND METHOD ALLOWING FOR IMAGE EQUALIZATION AND SYNCHRONIZATION OF MANUAL WINDOWING, WHEN COMPARING RECENT CHEST RADIOGRAPHS WITH PREVIOUS STUDIES

Collaborators

Denise Castro, pediatric radiologist

S.Salahudeen, thoracic radiologist (asst.prof)

R.Nolan,thoracic radiologist (prof)

E.VanDenKerkhof,epidemiologist,PhD (prof)

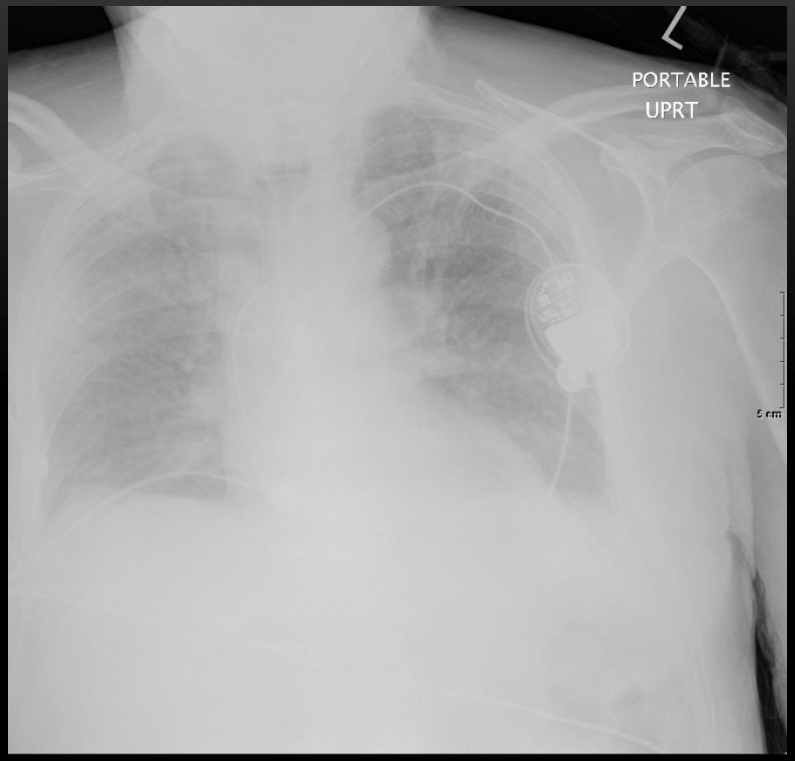
Dr.Flood,thoracic radiologist

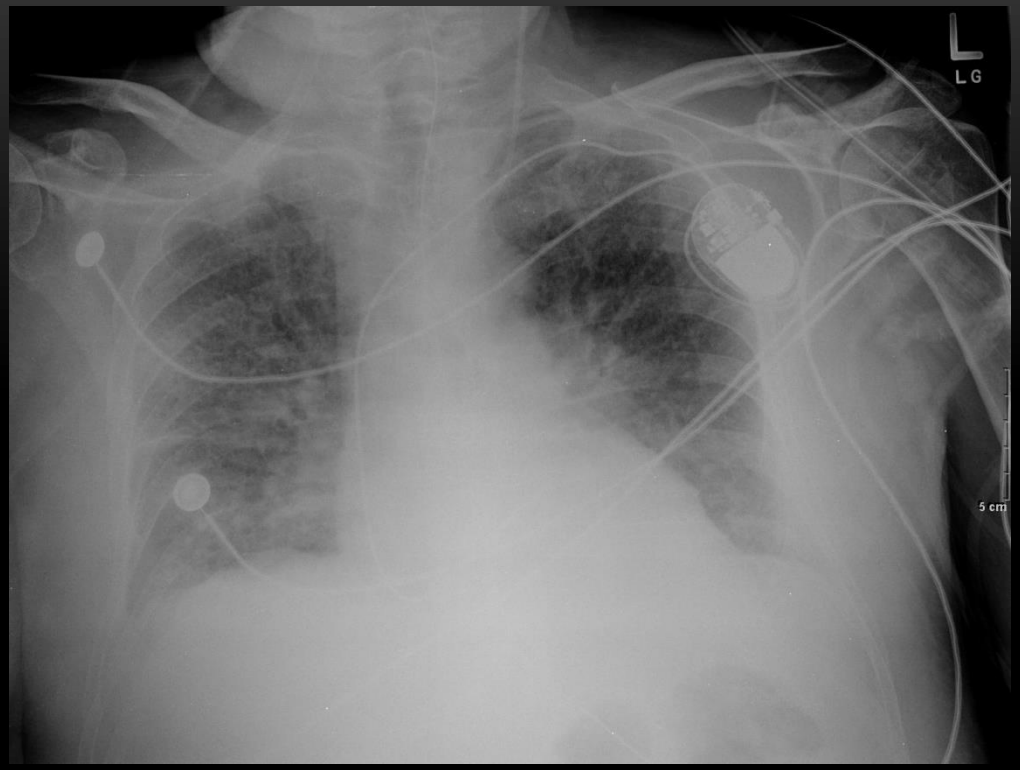
Purpose

to determine whether a novel **DEVICE** and **METHOD**, that equalizes chest radiographic appearance and allows for synchronization of manual windowing with comparison studies, would improve consistency in interpretation and dictating efficiency.

What prompted the study?

Previous studies have demonstrated a sub-optimal chest radiograph appearance in up to one-third of cases and have shown a poor correlation with autopsy findings





We were spending way to much time and effort agonizing over potential changes on the radiograph

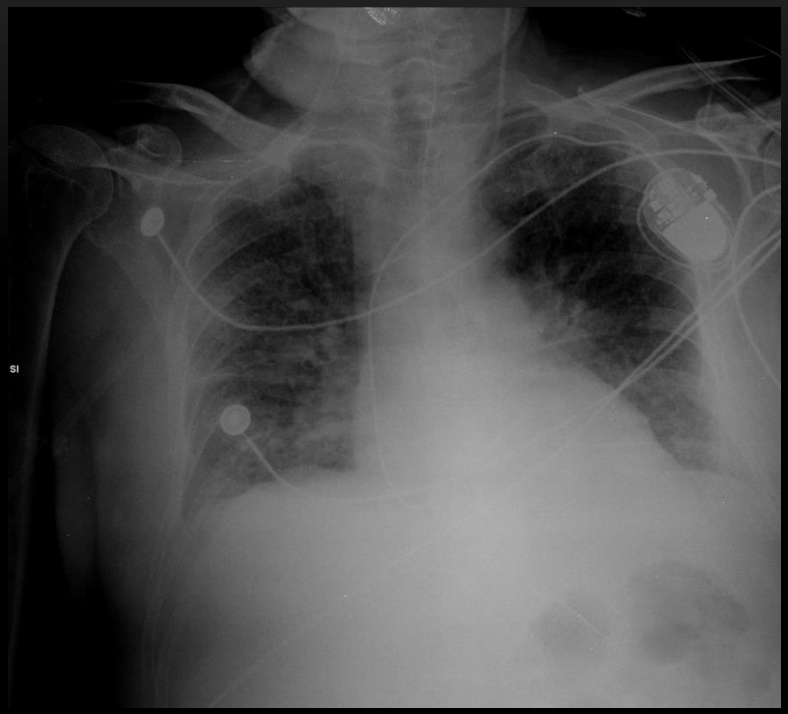
- It can be extremely difficult for the technicians to reproduce exact positioning and exposure techniques particularly in the ICU/NICU

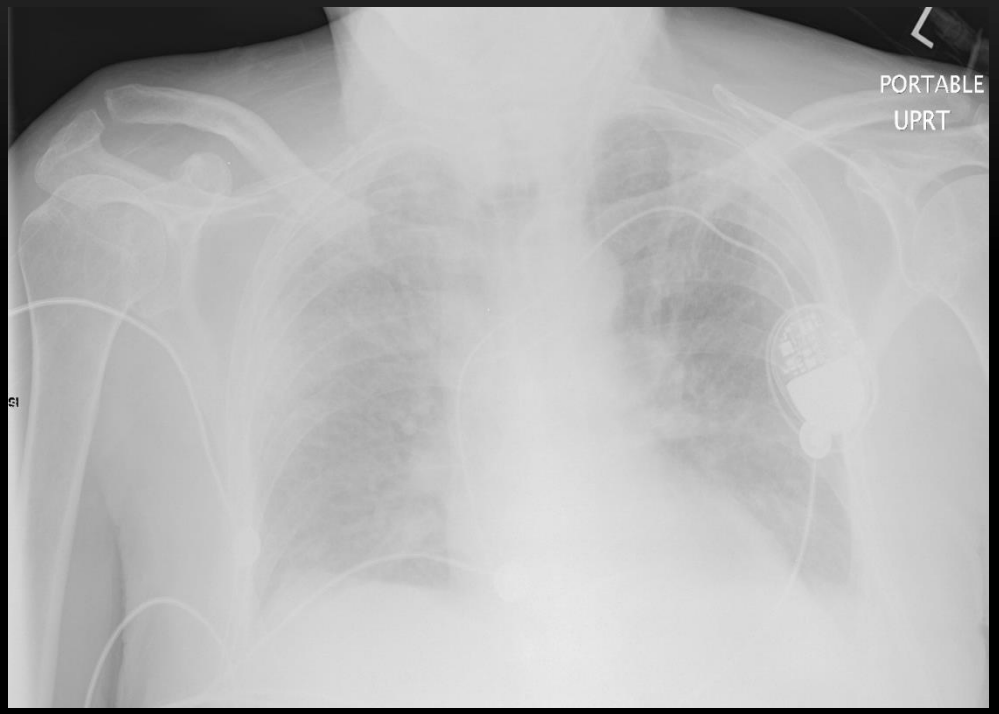
  Patient habitus may also change between studies
  ie. Weight gain or loss , surgery etc.

  Support apparatus may change exposure as well as prevent proper positioning

  shorter beam distances than recommended in some pts

  all contribute to that 1/3 of cases which are suboptimal





The clinical question often breaks down to

- is it better or worse

The Novel Device

- consists of a VAP (variable attenuation plate) composed of variable thicknesses of different metals ie. brass, Al

0.1 inch brass mounted on 0.0625 inch aluminum

The Novel Device

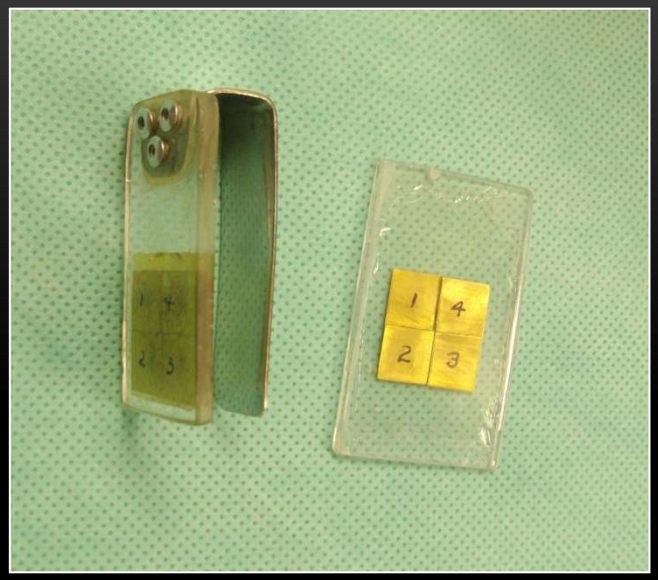
- plate can be varying shapes, with or without ‘handle’

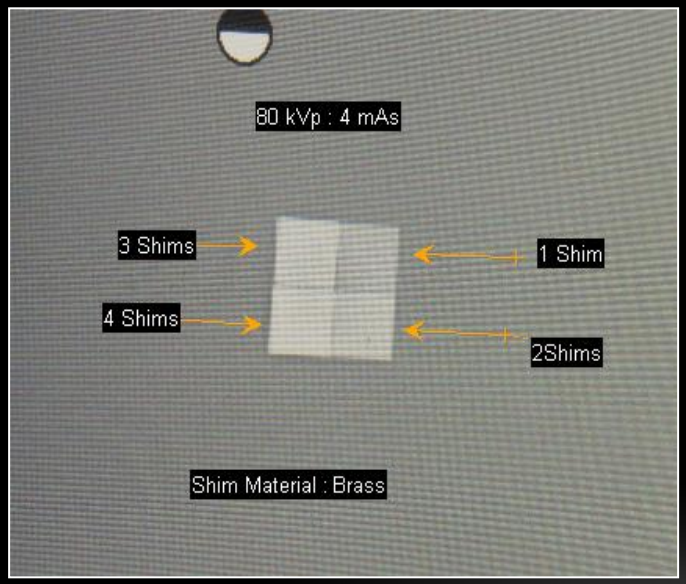
  2cm by 2 cm square or 1 cm by 4 cm strip

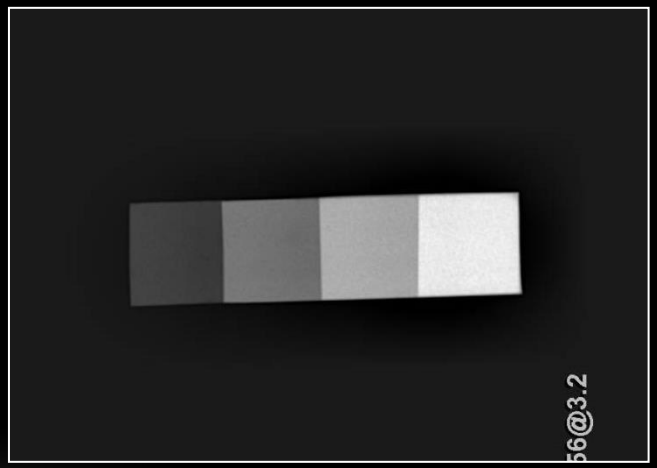
  each square a different thickness based on no. of layers

The Novel Device

- upon exposure 4 differing density quadrants

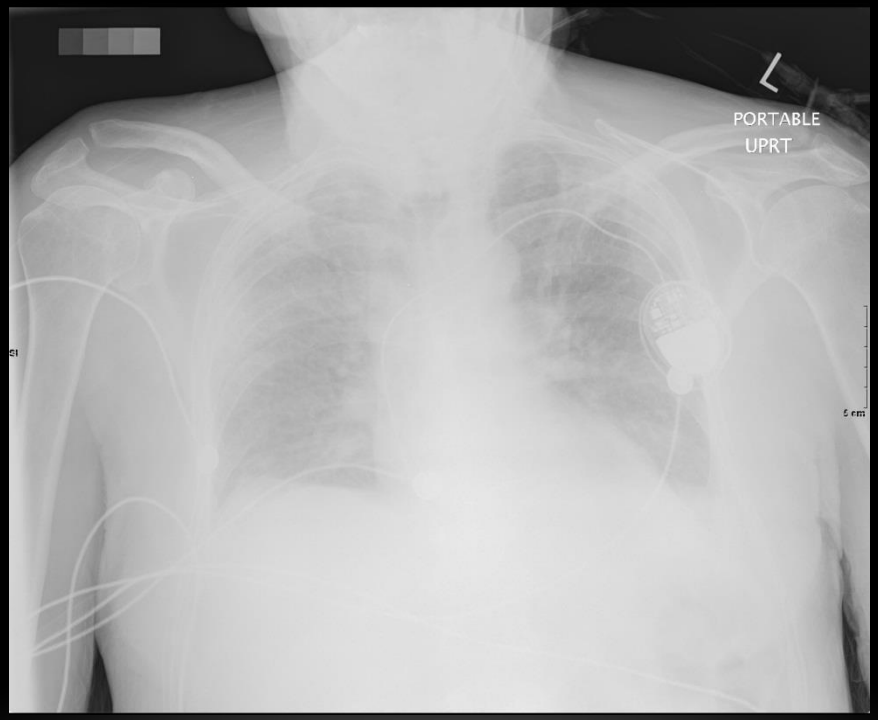




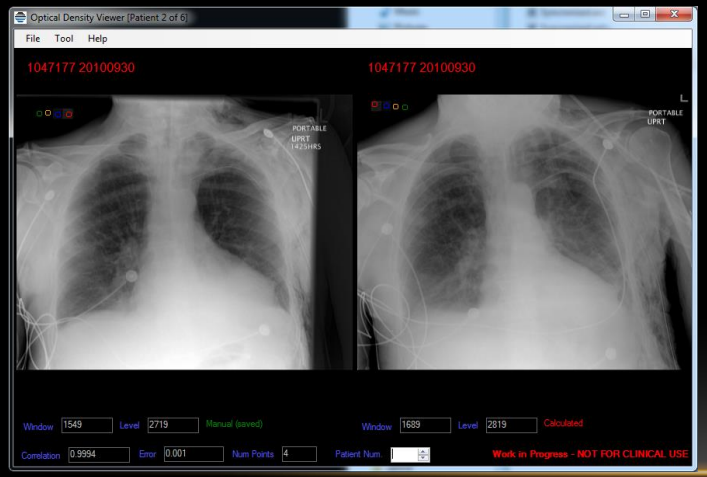


The Novel Method

consisted of software specifically designed to ‘match’ each quadrant on the VAP with its corresponding quadrant on a separate image



An high value of correlation (close to 1) indicates that the intensities are distributed linearly, therefore validating the methodology **because the calibration of the images is independent on which intensity have been used.**



Locating the Markers

Synchronized Window Level

Study methods

- Conducted at a an ICU in a tertiary care hospital

- 50 non-consecutive patients

-17 days to 85 yrs of age (24 males)

-29/50 intubated

Study methods

- The 50 patients undergoing CXRs as part of their routine care
- each patient had frontal radiographs (on different days) in which the VAP had been placed on the cassette
- Important - the VAP needed to be away from the patient and support apparatus

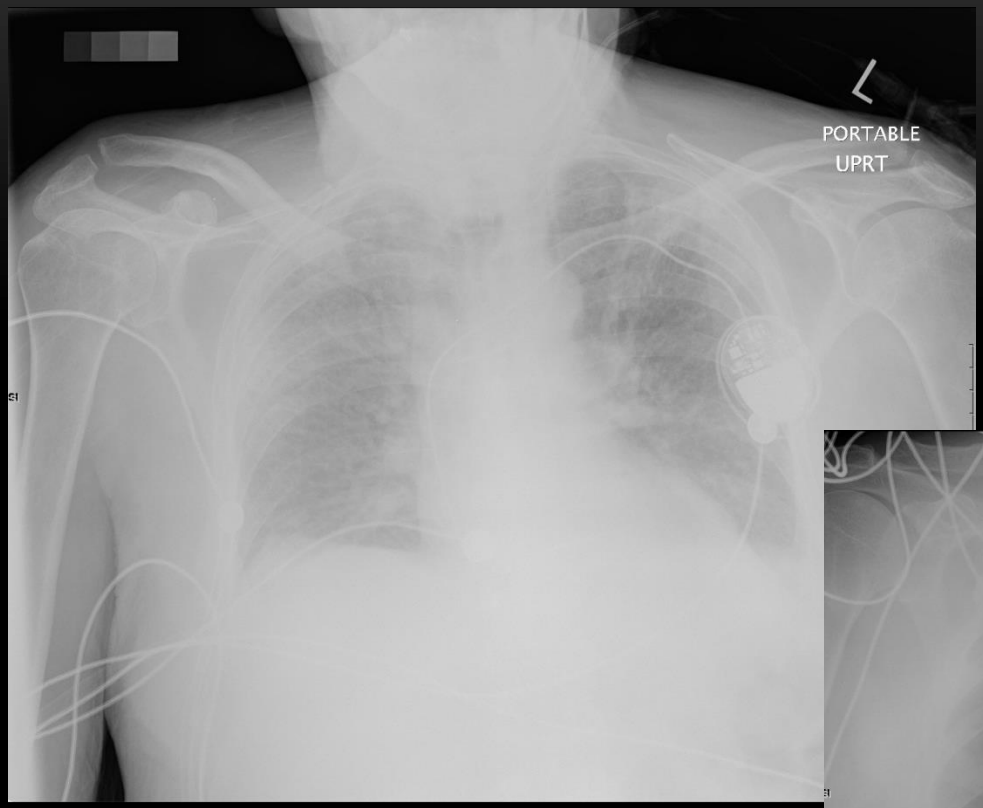
Study methods

- Short training session to 5 technical staff

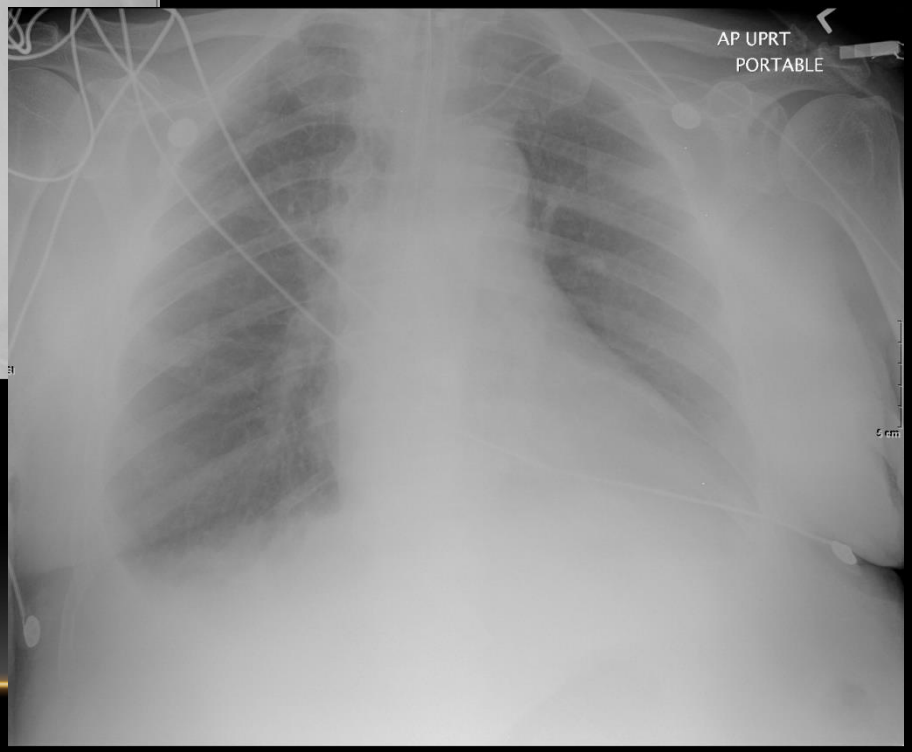
  - explain importance of positioning of VAP

  - instructed to place similar to their rt \ lt markers

  - technique otherwise unaltered
    90kVp and 4mA (60kVp and 1.5mA nicu)



Included



Excluded

Study Method

- the 100 radiographs (50pts x 2 CXRs) were then randomly placed as acquired, in 2 identical viewers on our PACS, with the previous study to the right of the more recent CXR

  cases were anonymized and identified by number

  3 thoracic radiologists then reviewed and interpreted the 100 cases (50 paired cases in each viewer)

  experience 3 – 30 years- no conflicting commercial interest

Study Method

- dictation of the 50 cases was conducted in the radiologists usual manner using one viewer with windowing as deemed appropriate
- on the other viewer, dictation was conducted with use of the specifically designed software utilizing the VAP allowing for equalization of appearance and synchronization of windowing when comparing recent to previous CXRs
- sequence of dictation was randomly alternated between those utilizing the VAP and those not

Study Method

- each report included an impression of either
  - WORSE
  - NO CHANGE
  - IMPROVED
- Dictation time per case and total was calculated by an observer
- Report impressions were compared both between radiologists and individually between methods (with and without use of the VAP/software)

		Radiologist A		Radiologist B		Radiologist C	
			VAP		**VAP**		**VAP**
CASE	# 1	X	X	X	O	X	X
	#2	O	O	☺	O	☺	O
	#3	O	O	O	O	O	O
	#4	X	X	X	X	☺	X
	#5	☺	O	☺	O	☺	O
	#6	X	O	O	X	X	X
	#7	X	O	O	O	O	X
	#8	X	☺	☺	☺	O	O
	#9	O	X	O	X	O	X
	#10	O	O	O	O	O	O

X = worse O = no change ☺ = improved

Results

- Each radiologist dictated all 50 cases in each viewer. (One case discarded due to dictation problem)
- There was a statistically significantly difference in agreement on case impression between the two methods

Results

- Kappa values between Rads A and B, A and C, B and C

without VAP	46%	55%	51%
with VAP	73%	81%	66%

Results

- Intra-observer weighted Kappa values between non-VAP and VAP methods

  - ranged 63% - 86%

  - suggested moderate to very good agreement between methods

Results

Discrepant report impressions

- 1 rad stating improved vs one or both other rads stating unchanged or worse
- without VAP range of 24 – 28 %
- with VAP range of 10 – 16% ($p<0.01$).

Results

Opposing report impressions

- 1 rad stating improved vs one or both other rads stating worse
- without VAP - 12 % of cases
- with VAP - 7 % of cases

Results

- the mean time to dictate each case was 44 seconds for both groups

- however, the mean total time to dictate the whole batch of radiographs was 20 minutes faster (97 min. vs 77 min.) using the VAP method

(includes the actually dictation 'mic' time as well as the time the radiologist spent deciding what to say in their report)

The Novel VAP Method

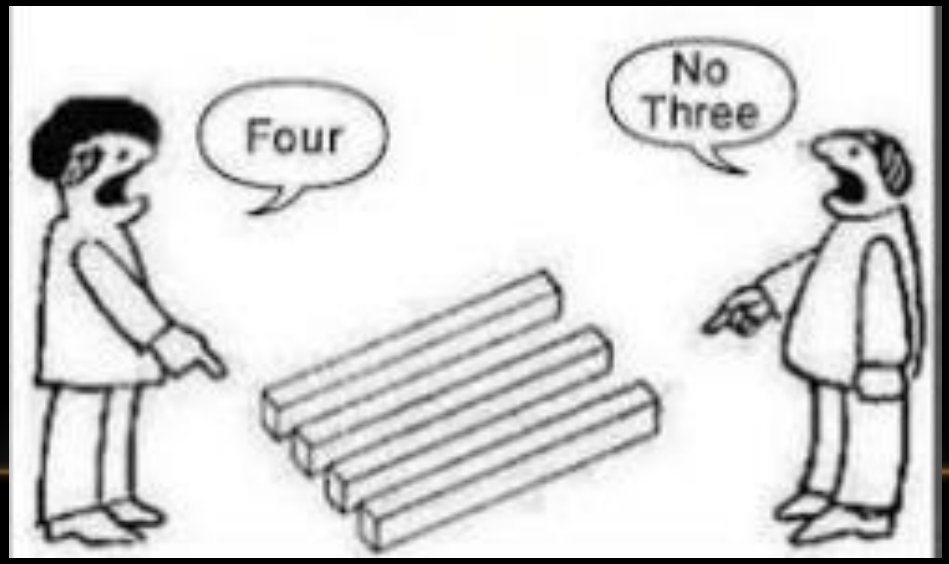
Conclusion

- **IMPROVED CONSISTENCY** in report impressions
  which we believe can result in improved patient care
- **EFFICIENCY** - the novel method allowed
  for an **~ 20 %** decrease in overall reporting time in our study

How can we improve perception and our consistency in reporting?

SOLUTION

2. Equalize the appearance and allow manual window synchronization of recent CXRs with their previous study

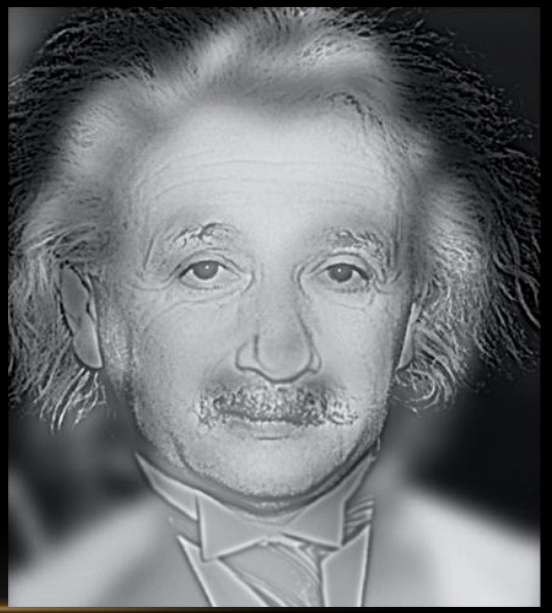


Why the poor inter-observer agreement?

Problem solved?

If perceptual? Not just technical solutions

inherent



How many faces can you perceive?

Definition of diagnostic error

Diagnostic error is defined as a diagnosis which is missed, delayed or wrong as determined by a subsequent definitive exam or test

interpretive vs perception

Perception error up to 80 %

First described by Garland in 1949

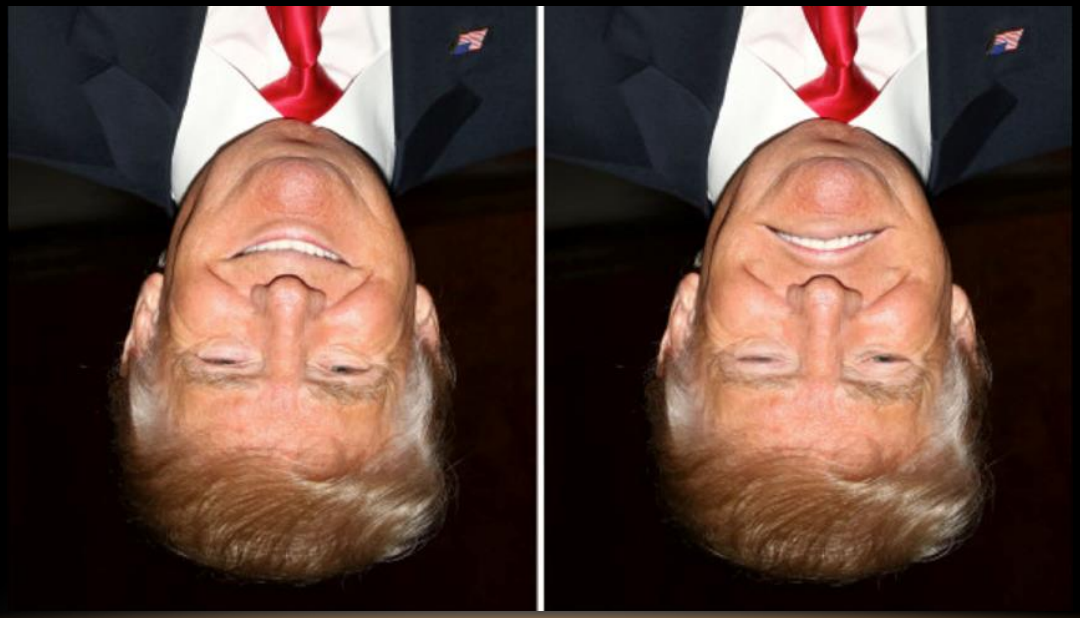
- perception error in 20-30% of chest cases in a study on TB

More recent perception error studies

- across multiple modalities
- across multiple countries and
- across multiple subspecialties

NOT changed significantly since Garland first described

A medical image perception society (MIPS) now exists with a goal to improve the understanding of imaging perceptual factors and foster research.



attempts to improve perception/decrease perception error and thus improve consistency in interpretation of all imaging modalities

- structured reporting
- Double/triple reading
- Improved luminance
- Changing focal zones

Gaze trackers

- limited success so far

Recent submission

Factors affecting perception error in Sonography

Collaborators

D. Castro – pediatric radiologist

E. Sauerbrei – radiologist

M. Kolar – pediatric surgeon

W. Hopman - biostatician

Purpose

- determine if experience
- knowledge base
- amount of time spent

correlated with ability to see a normal appendix

Study population

- 343 pts referred to imaging dept. with a clinical concern of appendicitis

- pts with a normal appendix or non-visualized appendix with subsequent discharge and normal follow-up included

Sonograms performed in the usual manner by

- one of 15 trained sonographers
  (3- 23 yrs experience)

- one of 8 radiology residents
  (4 in PGY 2/3; 4 in PGY 4/5)

Result – overall adult population

- no effect knowledge base
- no effect experience level
- no effect time spent

If patient could have been triaged to the 4 sonographers with the best perception in identifying the normal appendix

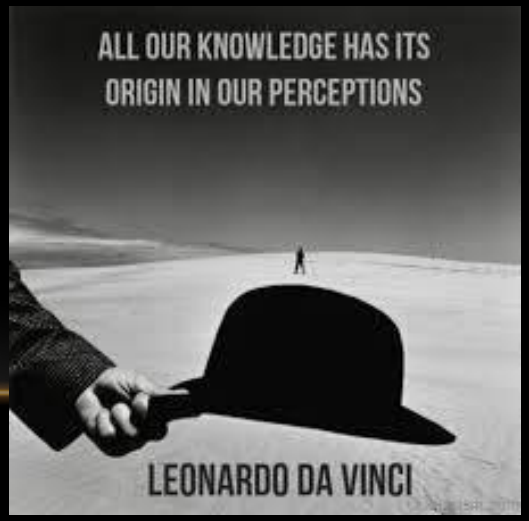
success rate would increase to 75 % from 27%

would have resulted in a 48% decrease in the number of CT’s ordered to further evaluate these patients

Take home message

1. new and novel methods are needed to improve our consistency in interpretation of imaging studies

2. despite our advancements perception error will always play a role in our daily lives



Thank-you for your attention

Dr. D.A. Soboleski