Poor interpretation of chest X-rays by junior doctors

Janus Mølgaard Christiansen[1], Oke Gerke[2, 3], Jens Karstoft[4] & Poul Erik Andersen[4]

ABSTRACT

INTRODUCTION: Studies targeting medical students and junior doctors have shown that their radiological skills are insufficient. Despite the widespread use of chest X-ray; however, a study of Danish junior doctors' skills has not previously been performed.
MATERIAL AND METHODS: A total of 22 participants perused a standardised series of ten chest X-rays. The test used a multiple-choice form for each image, and the clinical data and the tentative diagnosis of each image were also made available to the participants. For each image, the participant chose a single primary diagnosis; and for each diagnosis, the participant's confidence in the diagnosis was assessed on a five-point Likert scale. The diagnoses were divided into four groups: normal findings, chronic diseases, acute diseases and hyperacute diseases or conditions.
RESULTS: A total of 22 doctors receiving basic clinical education (BCE) completed the study. Overall, participants correctly established 51% of the diagnoses. The participants' overall confidence in the primary diagnoses was 57.5% on the Likert scale, corresponding to 57.5% confidence in the proposed diagnoses. The sensitivity was calculated to 0.49 (95% confidence interval (CI): 0.41-0.57) and the specificity to 0.55 (95% CI: 0.41-0.68).
CONCLUSION: Based on the results from this study, we conclude that BCE doctors do not meet the minimum requirements for radiological diagnostic skills for the use of chest X-ray that were established for this study.
FUNDING: not relevant.
TRIAL REGISTRATION: not relevant.

Doctors in basic clinical education (BCE) are in the front line in emergency wards and trauma centres. It is expected that they are capable of independently evaluating chest X-rays, even though the Danish medical education does not contain a separate course in radiology with a final examination [1-3].

International studies of medical students and junior doctors show that their radiological skills are insufficient, and these studies emphasize the need for training in radiology [4-6]. To our knowledge and despite the widespread use of chest X-ray, a similar Danish study has never been performed. An assessment of the Danish BCE doctors' diagnostic skills reading chest X-ray therefore seems relevant to medical schools, hospitals, patients and BCE doctors alike.

The minimum radiological diagnostic skills that a BCE doctor should command are the requirements for the doctor of medicine degree as defined at the three Danish medical faculties [1-3].

Specifically, the pathological findings that students must be able to recognize on a chest X-ray are defined in the curricula. In addition, there are more non-specific requirements for doctors, which they should fulfil during their BCE [7].

The purpose of this study was to determine whether BCE doctors are capable of meeting the minimum standards for a BCE doctor's diagnostic skills using thorax X-ray, as required by the Danish medical degree and by the Danish Health and Medicines Authority.

MATERIAL AND METHODS

This prospective, multicentre, controlled, cross-sectional study was conducted after the test material was validated and reference diagnoses were established.

The study was conducted at the following Danish hospitals:
– Odense University Hospital, Odense
– Odense University Hospital, Svendborg
– Hospital of South-west Jutland
– Hospital of South Jutland.

Participants

All 89 BCE doctors affiliated with the above hospitals during the period from 1 April 2011 to 5 May 2011 were invited to participate.

The study period coincided with the first part of the doctor's BCE term. Long work hours and mandatory courses at all hospitals, except at Hospital of South-west Jutland, thus prevented many BCE doctors from participating. The 22 BCE doctors who did participate formed a random group, mainly from Hospital of South-west Jutland, whose work schedule allowed them to participate in the study. The age and the sex of the participants were not registered, but both men and women participated.

The inclusion criteria were: Persons holding a medical degree from a Danish University and currently working as a BCE doctor.
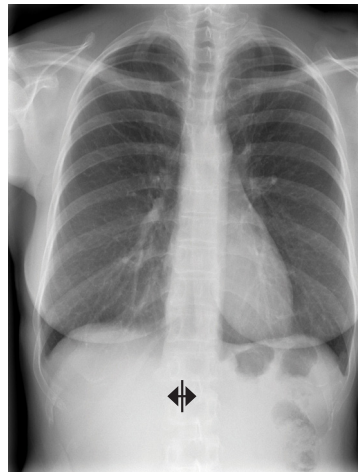
The exclusion criteria were: Former position at a radiological department or other training in radiology.

Normal chest X-ray, used in the study.

visually clear and not be concealed by other structures
– The image should be of good quality without artifacts
– A correct clinical reference and a tentative diagnosis should be provided with the image.

Observance of these criteria is tantamount to adherence to the recommendations of radiological images suitable for studies [9]. This study included ten primary diagnoses in the image material. The images were exported in DICOM format so as to maintain the original image quality. For each image, the original reference text and a tentative diagnosis were copied directly from the RIS/PACS system for use in the multiple choice form.

Questionnaire
The test used a standardised multiple-choice form for each image. The form was designed for maximum clarity and minimised the risk of errors during the subsequent data processing [10]. For each picture, the clinical data and the tentative diagnosis were provided. This was done to simulate the conditions in BCE doctors' daily work and to provide optimal conditions for diagnostics [11]. For each picture, the participant should choose a single primary diagnosis out of 37 possible diagnoses, and for each diagnosis the participant's confidence in the diagnosis was assessed on a five-point Likert scale [12]. The diagnoses were divided into four groups: normal findings, chronic diseases, acute diseases and hyperacute diseases or conditions.

Validation
To validate the material used in the study [13], radiographs and questionnaires were tested by three experienced radiologists under the same conditions as were later used in the study. There was no inter-observer variation, and there was consistency between the original radiological findings and the primary diagnoses chosen by the radiologists.

Statistical analysis
Danish BCE doctors' ability to interpret chest X-rays was, to the best of our knowledge, unknown prior to this study, and no other studies were available for comparison. Thus, no formal sample size computations were performed at the planning stage. The actual number of participants was determined by practical limitations, but as many participants as possible were invited to participate in the study.

Data were presented and described according to data type: categorical variables were analysed using frequency tables, continuous variables were examined with descriptive statistics such as mean, median and standard

The framework of the study
Participants evaluated a standardised series of ten chest X-rays at a radiological workstation with a Radiology Imaging System and picture archiving system (RIS/PACS). For each image, the participants chose a single primary diagnosis in a multiple choice form. Participants also recorded their assessment of how confident they were with regard to their chosen diagnosis. Measuring participants' confidence in their diagnoses provided a tool for assessing whether BCE doctors are aware of the limitations of their own abilities.

The trial took place at various hospitals, and it was estimated that differences in the quality of monitors and the physical environment could affect the results and lead to differences in the BCE doctors' measured performance [8]. To counteract this, the trial was assigned to hospital radiology departments with displays of a diagnostic standard, arranged in rooms with proper backlighting. To further standardise conditions between hospitals, the images were shown with an external DICOM image module (K-PACS. V.1.6.0.)

The primary diagnoses included in the study only comprised conditions and diagnoses that are commonly seen in emergency wards. Thus, the trial reflects the BCE doctors' daily work. The images used in the study were chosen according to the following criteria:

– The image should contain a single primary diagnosis without additional secondary findings that could create doubt as to which diagnosis was the primary one
– The primary diagnosis had been verified by series of X-rays, or other modalities, such as computed tomography or magnetic resonance imaging
– The primary diagnosis should not contain atypical visual elements
– The findings of the primary diagnosis should be

deviation. Where data were not normally distributed, non-parametric methods were used. Point estimates were supplemented by corresponding 95% confidence intervals (CI). Because the participants all scrutinised the same ten pictures, clustered sandwich estimators were used for the computation of the 95% CI. In addition, the participant's confidence in the proposed diagnoses was calculated and expressed as the sum of scores from the Likert scale.

Primary endpoints were a) the total percentage of correct primary diagnoses and b) the percentage of correct primary diagnoses for every single image.

Secondary analyses focused on the sensitivity and specificity of the BCE doctors' diagnostic ability. To that end, the primary diagnoses were divided into two groups. The categories 1 and 2, i.e. "Normal" and "Chronic changes" were combined to form the category "Healthy". These findings had no immediate consequences to the patients, regardless of the BCE doctors' diagnoses. Moreover, the categories 3 and 4, i.e. "acute conditions" and "hyperacute conditions" were combined to form the category "Sick" since these findings require immediate action and, therefore, are highly dependent on the BCE doctors' judgment. Thus, category 1 and 2 findings that were assessed as category 3 or 4 findings were "false positives". Correspondingly, category 3 and 4 findings that were assessed as category 1 or 2 findings were "false negatives". Assessing sensitivity and specificity in this way provides information on BCE doctors' ability to distinguish conditions that require instant treatment.

All data were entered into Microsoft Excel 2007 spreadsheets and calculated with Stata/MP 12.1 (StataCorp LP, College Station, Texas, USA).

Trial registration: not relevant*.*

RESULTS
Primary diagnoses
A total of 22 BCE doctors completed the study (**Table 1**). The result for every participant (**Table 2**) and every single image in the study was calculated (**Table 3**). Overall, participants chose correctly in 51% of the diagnoses (95% CI: 0.43-0.58) with the greatest number of correct answers being provided for picture six, which was normal and without pathological findings.

Confidence
The figures for the participants' confidence levels for each diagnosis were rated on a five-point Likert scale corresponding to 0-100% (Table 3). The participants' overall confidence in the primary diagnoses was an accumulated average of 33 for all ten pictures, but with large variation between the different pictures (standard deviation 7.8). This gives an overall confidence level of 3.3 on the Likert scale, corresponding to a 57.5% confidence in the proposed diagnoses.

Sensitivity and specificity
The assessment of sensitivity and specificity of the BCE doctors' diagnoses allows us to assess whether BCE doctors can distinguish relatively benign conditions from serious conditions that require immediate treatment. We found a sensitivity of 0.49 (95% CI: 0.41-0.57) and a specificity of 0.55 (95% CI: 0.41-0.68) (**Table 4**)

DISCUSSION
The minimum standards for BCE doctors' diagnostic skills using chest X-ray are established by the faculties at the Danish medical schools and by the Danish Health and Medicines Authority. In order to assess whether these goals have been met, a number of factors must be taken into account.

First of all, uncertainty is associated with any diagnostic procedure. This uncertainty is difficult to quantify due to the nature of the material and the method used. Diagnostic precision varies much, even among trained radiologists. Studies attempting to measure this uncertainty indicate that the number of errors made by trained radiologists fall in the 11-30% range [9, 14]. Other studies of inexperienced doctors' radiological diagnostic abilities show errors in diagnostic accuracy ranging from 28% to 45% [5, 6]. It is therefore evident that the acceptable maximum number of diagnostic er-

---

▤ **TABLE 1**

Participants by hospital[a].

	Participants	
Hospital	**n**	**%**
Odense University Hospital	6	27
Odense University Hospital, Svendborg	3	14
Hospital of South Jutland	2	9
Hospital of South-west Jutland	11	50
Total	22	100

a) The age and sex of the participants were not registered.

▤ **TABLE 2**

Number of correct primary diagnoses for each participant.

	Sum of correct answers											
	0	1	2	3	4	5	6	7	8	9	10	total
Frequency, all groups	–	1	–	2	5	4	8	1	–	1	–	22
Estimated % of correct primary diagnoses, all groups (95% confidence interval)												0.51 (0.43-0.58)

TABLE 3

Number of correct primary diagnoses for each picture.

Picture no.	Estimate (95% confidence interval)
1: Stasis pulmonum magna gradu	0.59 (0.37-0.81)
2: Effusio pleura dexter magna gradu	0.41 (0.19-0.63)
3: Abscess/empyema	0.5 (0.27-0.73)
4: Fracturae costae sequelae sinister	0.23 (0.04-0.42)
5: Emphysema pulmonum	0.63 (0.42-0.85)
6: Normal picture, no pathology	0.77 (0.58-0.96)
7: Pneumonic infiltrate sinister	0.05 (0-0.14)
8: Malignant infiltrate sinister	0.55 (0.32-0.77)
9: Pneumothorax dexter	0.73 (0.53-0.93)
10: Pneumothorax dexter	0.59 (0.37-0.81)

TABLE 4

Specificity and sensitivity.

Category	Specificity (95% CI)	Sensitivity (95% CI)
1 & 2: normal and chronic conditions (i.e. "healthy")	0.55 (0.41-0.68)	–
3 & 4: acute and hyperacute conditions (i.e. "sick"')	–	0.49 (0.41-0.57)

CI = confidence interval.

rors allowable for the study cannot be clearly defined, and it is therefore necessary to set a theoretical minimum level .

The study uses optimal radiological material, is based on well-defined diagnoses and radiographs which were analysed under optimal conditions. During their education, BCE doctors are examined in the particular diseases and analysis of radiological findings is part of their curriculum. During the test, they had access to the relevant clinical findings and a tentative diagnosis. The threshold value chosen for this study is that in this well-defined context and working under optimal conditions, BCE doctors ought to score high on diagnostic certainty corresponding, as a minimum, to the lowest certainty reported for radiologists. Therefore, an error rate of up to 30% of the primary diagnoses was deemed acceptable.

Only 22 BCE doctors participated, but the results of the study were still statistically significant given the chosen theoretical threshold value. The participants were deemed representative of all three Danish medical faculties. BCE doctors participating in the survey had an overall accuracy of 51% (95% CI: 0.43-0.58) correct primary diagnoses. These figures largely confirm the results of international studies of junior doctors and medical

students [5, 6]. It is important to note that these earlier studies included all overlooked radiological findings, whereas our study was based on well-defined findings and related to their clinical significance. A shared characteristic of the primary diagnoses is that the estimates as well as the upper limits of the corresponding 95% CI fall below the minimum threshold value defined in this study.

A sensitivity of 0.49 (95% CI: 0.41-0.57) and a specificity of 0.55 (95% CI: 0.41-0.68) also indicate a risk that BCE doctors make mistakes when assessing the radiological findings. This can lead to delay in life-saving treatment or cause patients to be treated too aggressively which may lead to complications.

The participants' confidence in the chosen diagnoses was 57.5%. This, combined with the diagnostic accuracy of 51% observed in our study, suggests a risk that BCE doctors may misdiagnose a patient without consulting a radiologist or experienced colleague.

CONCLUSION

Based on this study, we conclude that BCE doctors do not meet the established minimum requirements for radiological diagnostic skills for the use of chest X-ray.

Every working day, BCE doctors make independent decisions on the basis of paraclinical test results. These decisions can have serious consequences for patients; for example, an overlooked pneumothorax may be life-threatening.

In the new united emergency wards (FAM) and in hospital departments, BCE doctors are in the frontline. Although BCE doctors often have the opportunity to consult colleagues, this does not change the fact that a BCE doctor may misdiagnose an X-ray and decide not to consult others because of misplaced confidence in their own abilities.

In the medical world, there is a belief that young doctors are capable of acquiring the needed skills through work experience and thereby become more competent in making clinical decisions without receiving structured training. One study [7] demonstrated that there was no improvement in the radiological skills of junior doctors over a period of six months, unless they received structured training.

On this basis and keeping in mind the results of the present study, we recommend that increased time and resources be allocated to training of medical students and BCE doctors in basic radiology.

CORRESPONDENCE: *Janus Mølgaard Christiansen*, Thulevej 242, 6715 Esbjerg, Denmark. E-mail: barberkirurg@gmail.com

LITERATURE

1. Studieordning for kandidatuddannelse i medicin. Aarhus: Det Lægevidenskabelige Studienævn, Aarhus University, 2011.
2. Studieordningen kandidatuddannelsen i medicin, Syddansk Universitet. Odense: University of South Denmark, 2006.
3. Kandidatstudieordning for medicin, Københavns Universitet. Copenhagen: University of Copenhagen, 2009.
4. Eisen LA, Berger JS, Hegde A et al. Competency in chest radiography. A comparison of medical students, residents, and fellows. J Gen Intern Med 2006;21:460-5.
5. Vincent CA, Driscoll PA, Audley RJ. Accuracy of detection of radiographic abnormalities by junior doctors. Arch Emerg Med 1988;5:101-9.
6. McLauchlan CA, Jones K, Guly HR. Interpretation of trauma radiographs by junior doctors in accident and emergency departments: a cause for concern? J Accid Emerg Med 1997;14:295-8.
7. Sundhedsstyrelsens målbeskrivelse for den kliniske basisuddannelse. Copenhagen: The Danish Health and Medicines Authority, 2009.
8. Buls N, Shabana W, Verbeek P et al. Influence of display quality on radiologists' performance in the detection of lung nodules on radiographs. Br J Radiol 2007;80:738-43.
9. Garland LH. On the scientific evaluation of diagnostic procedures. Radiology 1949;52:309-28.
10. Juul S. Take good care of your data. Aarhus: Aarhus University, Department of Public Health, 2006.
11. Doubilet P, Herman PG. Interpretation of radiographs: effect of clinical history. AJR Am J Roentgenol 1981;137:1055-8.
12. Maurer, Todd J, Heather R. A Comparison of Likert scale and traditional measures of self-efficacy. J App Psychol 1998;83:324-9.
13. Brealey S, Scally AJ, Thomas NB. methodological standards in radiographer plain film reading performance studies. Br J Radiol 2002;75:107-13.
14. Berlin L. Reporting the "missed" radiologic diagnosis: medicolegal and ethical considerations. Radiology 1994;192:183-7.