ROC Curve for Extremely Subtle Lung Nodules on Chest Radiographs Confirmed by CT Scan

Hua Shang, MD[1], Yulei Jiang, PhD, Feng Li, MD, PhD, Heber MacMahon, MB, BCh, Jianing Wang, MS

Rationale and Objectives: The aim of the present study was to test the hypothesis that when a radiologist does not perceive an abnormality in images that contain either extremely subtle abnormalities or no abnormalities, the radiologist cannot distinguish these two types of images and the receiver operating characteristic (ROC) curve reflects that performance.

Materials and Methods: This retrospective study was conducted with approval of our institutional review board. Four general radiologists participated in an observer performance study of 100 chest images, each of which had a 5 × 5 cm region of interest (ROI) drawn (50 containing a lung nodule, and 50 did not, based on computed tomography [CT] confirmation). About half of the lung nodules were extremely subtle. The readers reported their confidence that a nodule was present within the ROI, from which empirical and maximum-likelihood "proper" binormal and conventional binormal ROC curves were estimated. The readers also reported whether they saw an abnormality that could be a nodule within the ROI.

Results: Empirical ROC curves deviated from typical ROC-curve shapes, and a portion of the curve leading to the northeast corner of the ROC space had relatively steep and constant slopes. The readers reported not seeing anything suggestive of a lung nodule in this portion of the ROC curve, which also corresponded to cases that either contained extremely subtle nodules or normal cases. The average area under the ROC curves (mean ± standard deviation) was 0.66 ± 0.02 for proper binormal, 0.62 ± 0.02 for conventional binormal, and 0.60 ± 0.03 for trapezoidal ROC curves.

Conclusions: When a radiologist does not perceive an abnormality in images that contain either extremely subtle abnormalities or no abnormalities, the ROC curve (or a portion thereof) is characterized by a straight line, which is not consistent with conventional ROC theories.

Key Words: Receiver operating characteristic (ROC) analysis; observer performance; extremely subtle abnormality; ideal observer; human observer.

INTRODUCTION

Receiver operating characteristic (ROC) analysis is a cornerstone for the evaluation of diagnostic performance in binary tasks (e.g., cancer present vs. cancer absent) (1–5). A fundamental idea of ROC analysis is that the readers could alter, essentially at will, their decision criterion to call a case positive (e.g., cancer-present), thus altering sensitivity and specificity (i.e., ROC operating point) along the readers' ROC curve (3,6). Charles Metz explained in 1978, referring to Bayesian (or ideal observer) decision theory, that: "one can show on theoretical grounds that if the decision maker uses available information in a proper way, the slope of the ROC curve must steadily decrease (i.e., it must become less steep) as one moves up and to the right on the curve" (1). It is widely accepted that an ROC curve, including that of human observers, must have slopes that decrease monotonically as one moves up and to the right on the curve (7–9).

We observed recently that "occult abnormalities" in detection tasks could give rise to human observer empirical ROC curves that deviate from this shape (10). We defined an occult abnormality as one in an image for which, even in retrospect, when informed of its presence and location in the image, a reader is not able to identify it confidently. In radiology literature, these abnormalities are more commonly known as "extremely subtle"—to the extent that they may be not visible at all in the image[1]. We will thus use the term "extremely subtle" here instead of "occult." We showed in Jiang (10) that depending on frequencies of extremely subtle abnormalities and apparently normal cases (which can be considered a counterpart of extremely subtle abnormalities because they, too, do not contain visible abnormalities), human

Acad Radiol 2016; 23:297–303

From the Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, Chicago, IL 60637. Received September 18, 2015; revised November 9, 2015; accepted November 9, 2015. [1]Permanent address: Department of Radiology, The Second Hospital of Hebei Medical University, 215 Hepingxi Road, Shijiazhuang, Heibei Province, China. **Address correspondence to:** Y.J. e-mail: yjiang@uchicago.edu

http://dx.doi.org/10.1016/j.acra.2015.11.014

[1]Some researchers use "subtle" to describe only abnormalities that are "visible" in the image. The use of this term here includes abnormalities that are "not visible" in the image (e.g., lung nodules that are computed tomography-confirmed but not visible in chest images).

observers' ROC curves could have slopes that decrease monotonically (as ROC theory predicts), or remain approximately constant, or even increase—near the northeast corner of the ROC-curve space. In contrast, the ideal observer's ROC curve always has decreasing slopes (10).

We hypothesize that when a radiologist does not perceive an abnormality in images that contain either extremely subtle abnormalities or no abnormalities (i.e., apparently normal images), his or her detection performance as reflected by the ROC curve must be without any capacity to separate the images that contain extremely subtle abnormalities from apparently normal images. This ROC curve (or a portion thereof) is a straight line because any fractional gain in sensitivity must always be accompanied by an equal fractional increase in the false-positive rate. Figure 1 shows a schematic illustration of this hypothesized ROC curve compared to a conventional ROC curve. The straight-line portion of the hypothesized ROC curve near the northeast corner of the ROC space captures images that contain either extremely subtle abnormalities or no abnormalities, whereas the curved portion of the ROC curve captures the radiologist's discrimination performance of visible abnormalities.

Fundamentally, this hypothesis is not compatible with contemporary theories of ROC analysis. The theory predicts that radiologists are able to operate, meaningfully and by personal choice, anywhere along the conventional ROC curve, with commensurate trade-offs between sensitivity and specificity. However, this hypothesis predicts that radiologists can do so, meaningfully and by choice, only along the curved portion of the hypothesized ROC curve but not along the straight-line portion. Furthermore, the area under the conventional and hypothesized ROC curves will likely differ because of differences in the ROC curve shape. Moreover, statistical comparison between the hypothesized ROC curves could differ from statistical comparison between conventional ROC curves and could yield opposite conclusions from the same source data.

To test this hypothesis empirically, we conducted an observer performance study with a large portion of the images either containing extremely subtle abnormalities or being apparently normal images, and investigated the resulting ROC curves.

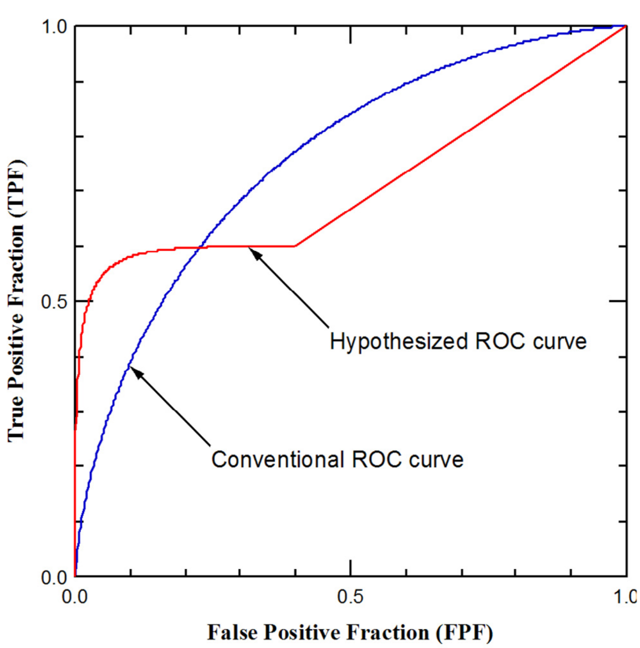

Figure 1. Schematic illustration of the hypothesized ROC curve compared with a conventional ROC curve. ROC, receiver operating characteristic.

MATERIALS AND METHODS

Study Cases

We selected standard chest radiograph (CXR) patient cases from the image database made publically available by the Lung Image Database Consortium (LIDC) (11,12). Of the 1012 cases available in the LIDC database, 290 cases have both standard posterior-anterior CXR and helical chest computed tomography (CT) scans (12). Of these, 99 cases were documented to contain a single CT-confirmed solitary nodule, and 191 cases contain more than one CT-confirmed nodule. We selected 50 nodule-present cases from the 99 cases that contain a single solitary nodule, after first eliminating 36 cases because (i): the nodule size based on CT was less than 5 mm or greater than 25 mm, (ii) the nodule was calcified, or (iii) poor CXR image quality. Of the 50 selected nodule-present cases, half (25 cases) were selected because the nodule was moderately or clearly visible in the CXR image, and the other half (25 cases) were selected because the nodule was extremely subtle or not visible in the CXR image. The LIDC database contains nodule subtlety scores on the CXR images provided by four radiologists (11,12). The ranges of the median and average subtlety scores (a score of 5 indicates that the nodule is conspicuous, and a score of 0 indicates that the nodule is not visible) were 2.0–5.0 and 1.75–5.0, respectively, for the 25 cases of visible nodules, and 0–0.5 and 0–1.25, respectively, for the 25 cases of extremely subtle nodules. One of the 50 nodule-present cases was a case of ground-glass opacity, which was not visible in the CXR image, and the rest were solid nodules.

On each CXR image, we drew a 5×5 cm region of interest (ROI) to enclose the solitary nodule. The purpose of these ROIs was to eliminate any uncertainty in the data collected from the readers with regard to which abnormality they intended to report. This is particularly important for cases in which the CT-confirmed nodule was extremely subtle on the CXR image; without this ROI, different readers could report on different perceived abnormalities and render the ROC analysis not meaningful. In addition, this ROI also allowed us to select cases from the LIDC database to be used in the present study as nodule-absent cases as long as CT showed no nodule within the ROI. We selected a total of 50 nodule-absent cases from the pool of 290 cases with CXR images. Of these, 25 cases had a nodule-like opacity within the ROI: in 13 cases, the opacity was due to normal anatomic structures (7 hilar or peripheral vessels with or without bony structures, 4 calcified cartilage opacities, 1 pericardial fat opacity, and 1 nipple