

PMA MEMORANDUM

From: William Sacks

To: Mike Kuchinski
Lead reviewer

Subject: Clinical Review, P030012
ImageChecker-CT CAD Software System, CAD for lung nodules on chest CT scans

Sponsor: R2 Technology, Inc.

Date: December 16, 2003

The intent of this review is to present an overview of the evolution of the thought processes through which FDA went during the discussions concerning a computer-aided **detection** system (CAD) for **actionable solid lung nodules on chest CT scans** obtained for any reason--particularly non-screening reasons, since lung screening is only now being studied as a feasible medical procedure. [An "actionable" lung nodule means simply one judged by the radiologist to require further evaluation, if not resection, and the term "solid" is intended to differentiate these from so-called ground glass opacities (GGOs), a descriptive term for lung abnormalities that contain air and are therefore not "solid," and which this device does not mark.]

The CAD target: Actionable nodule versus cancer

The first issue that FDA addressed was whether an **actionable** lung nodule, as opposed to a **malignant** lung nodule, was a suitable **target** for a CAD, i.e., could render the CAD a clinically useful device. Previous CADs have been approved for the target of breast **cancer**, suspicion of which requires definitive diagnosis through **biopsy**. In contrast, an actionable lung nodule generally (though by no means always) requires only **monitoring** rather than biopsy. Since a decision point also occurs when an actionable lung nodule is discovered, though in this case usually a follow-up CT after a pre-determined time interval rather than an immediate biopsy, it was felt that an actionable nodule might be a suitable target. The alternative target of malignant nodule would require biopsy to determine ground truth. Biopsy is more invasive in the lung than in the breast, carries a higher risk of morbidity/mortality, and was judged to be not feasible for this application.

The gold standard: Expert panel judgment versus biopsy

The second issue that FDA addressed was closely related to the first, and that was what to use as an independent reference standard or **gold standard** for an actionable lung **nodule** during the training of the algorithm and during the clinical study. In the absence of tissue histology--which would not, in general, be available if monitoring through follow-up CT was to be used as the next step in evaluation--the use of the judgment of an **expert panel** as a possible gold standard was considered.

Expert panel judgment has frequently, though not universally, been rejected as a surrogate for tissue histology. Expert panel judgment has been used by the FDA in the evaluation of devices other than those involved with imaging, particularly with in-vitro laboratory products. The frequent rejection in the case of imaging studies has been based on the dependence of the final judgment of the panel of experts on three variable features that have nothing to do with the item being tested: a) the **method used for combining** the judgments of the individual experts comprising the panel (e.g., majority rule versus consensus after discussion versus feedback re-review, etc.--Revesz G, Kundel HL, Bonitatibus M. *Investigative Radiology* 1983;**18**:194-198), b) the choice of individual experts to comprise the panel (**inter**-reader variability), and c) the judgment of each individual expert from one reading to a subsequent reading after memory of the case has faded (**intra**-reader variability).

Despite these problems with expert panel judgment as a gold standard, the use of actionable (versus malignant) lung nodules as a target seemed to us to necessitate the use of an expert panel. This is particularly true since, even conceptually, the notion of an actionable nodule does not lend itself to histological decision, given the inherently judgmental nature of actionability (as opposed to cancer, that has an objective existence apart from anyone's judgment.) However, this requires that the variability in the panel's judgment, i.e., in the gold standard, be taken into account statistically.

Sizing the study: Use of lung quadrants versus subjects

The third issue that the FDA addressed was whether it was adequate to use a given number of lung **quadrants**, rather than a given number of **subjects**, in sizing the study. Since there are statistical methods for taking into account the correlation among quadrants within the same subject, and since these correlations are not great for solid lung nodules, we agreed that this would give acceptable statistical power for a smaller number of subjects than would otherwise have to be enrolled.

No finer discrimination of location was determined than the particular lung quadrant in the particular subject, which means that the CAD gets credit for marking any abnormality in a quadrant, even if it differs in location from that identified by the expert panel. While less than perfect, this method has been used by the FDA in the past in evaluating a chest CAD, largely because of current limitations in theoretical work on this issue, both within the FDA and in the academic community.

Method of analysis: MRMC ROC (Multiple reader, multiple case receiver operating characteristic curve) versus Se/Sp (sensitivity and specificity)

The fourth issue that the FDA addressed was whether the MRMC method as used by the sponsor was sufficient. The two choices that were considered were, for a group of blinded radiologists (different from the unblinded expert panelists) reading the same set of lung CTs, the differences (from unaided to aided) in either a) the **ROC curves** or b) **Se/Sp** at a particular decision (or so-called operating) point. There is still some internal discussion taking place on this issue, though the FDA has been relying more and more on the use of MRMC ROC analysis for decisions concerning **approval**, if not **labeling**.

The company had the blinded test radiologists render a probability, for each lung quadrant, that an abnormality was an actionable nodule (probability of actionable nodule--POAN), and did not have them render an action recommendation for each quadrant, e.g., normal (no follow-up necessary), possible malignancy (recommend follow-up CT), probable malignancy (recommend biopsy), etc. Their reason was that, for any given study size, the statistical power for Se/Sp is significantly lower than it is for MRMC ROC, and they wished to avoid either enlarging the study substantially, in order to regain that statistical power, or paying a statistical price for multiple hypothesis testing.

This still leaves the possibility of estimating unaided and aided Se/Sp at a particular, arbitrarily chosen, value of POAN, but this would not be the same as estimating Se/Sp at a particular decision point between different action recommendations. For a computer algorithm this could indeed be the same thing, but for human image readers it is not the same, due to their variability in judgment from day to day and from one human reader to another, particularly in maintaining a fixed relationship between action decision points and particular values of POAN.