

Quality Assurance of CAD Systems Implemented in Clinical Use

Update from AAPM CAD Subcommittee Group # 3

Zhimin Huo, Ph.D.
Carestream Health Inc.

Acknowledgements

■ Members

- Heang-Ping Chan, Univ. of Michigan
- David Fryd, Riverain Medical
- Jeff Hoffmeister, iCAD, Inc.
- Joseph Lo, Duke Univ.
- Sophie Paquerault
- Marcos Salganicoff, Siemens Medical
- Hiroyuki Yoshida, Harvard Univ.

Our Goals

- Bring attention to issues such as :
 - Limited quality assurance testing performed on the system
 - there is no requirement and procedures that have been developed to assure and analyze the quality of CAD and the use of CAD over time at clinical sites
 - research on QA of CAD systems is limited.
 - the impact of proper QA on clinical performance with CAD is not well studied.
- Inform the opinions of the members from AAPM Subcommittees on CAD QA
 - Propose preliminary recommendation on CAD QA.

Our Goals

- These recommendations are intended to be work items for AAPM task groups that will be formed to address QA and user training issues on CAD in the future.
- The work items may serve as a framework for the discussion and eventual design of detailed QA and training procedures for physicists and users of CAD.
- The eventual standardization of the requirements of QA procedures for CAD will have to be determined through consensus from members of the CAD community (researchers, manufacturers, clinical users, physicists, regulatory bodies)

Our Goals

- Stimulate new ideas and approaches for effective implementation of CAD QA in clinical environment
- Expect that appropriate CAD QA and user training can ensure the effectiveness of CAD in clinical practice, allowing CAD reach its potential

Overview

- **Report the result of the work from group # 3**
 - Discuss the rationale for CAD QA
 - Discuss preliminary general guidelines on CAD QA.
 - Propose minimum CAD QA requirements that are considered to be reasonably easy and practical, and can be implemented immediately by the end users.
 - Propose recommendations that are considered to be “best practice” QA approaches, which could potentially provide more accurate measure of clinical performance of a CAD system and allow users to work with CAD more effectively and efficiently.

Concerns

- However, such “best practice” approaches may require significant effort, additional tools, additional cost and proper training to implement.

Rationale

- CAD products can be accepted in clinical use only if they can provide clinical benefit that outweighs the potential associated risks
- CAD becomes a part of routine clinical practice
- CAD is also rapidly expanding to provide additional functionalities
 - CADE, CADx, Measurement tools
 - Across different diseases/different modalities (2D/3D/4D)

Rationale

- Some level of QA is required
 - Ensure that CAD products function correctly at time of installation and continue to function correctly during it's clinical use
 - Ensure that CAD products perform according to vendor's specification
- Factors affecting the benefit of CAD products directly or indirectly
 - Image quality
 - image processing (including reconstruction)
 - CAD standalone performance
 - Image display quality
 - Radiologists (how they use CAD)

General Guidelines on CAD QA

- When does a QA procedure need to be performed?
- Who should perform the test?
- What are the tools and materials that may be used to perform the QA tests?
- What are the performance measures to be used?
- What QA results should be captured and how should they be reported?
- What are the criteria of success/failure of the tests (or minimal requirements)?
- What should be done if the CAD QA test fails?

Q1. When does a QA procedure need to be performed?

- Performance verification/conformance to specifications at the first installation or subsequent CAD software upgrades, including operating system upgrades
- Replacement or upgrade of imaging system hardware or software components that affect image quality (e.g., x-ray tube, energy source, detector, or image reconstruction algorithm)
- Check for interaction or impact on the CAD software when other software or upgrades are installed on the same workstation on which the CAD system resides
- Any experience of performance deviation from the original specifications
- Specific procedures identified as necessary to be performed on a regular basis (e.g., monthly, quarterly, yearly) to assure performance consistency, such as recommendations by the manufacturer.

Q2. Who should perform the test?

- CAD QA should be performed by designated quality assurance personnel with adequate training to perform the QA procedures, e.g., technologist, medical physicist, radiologist, manufacturer.

Q3. What are the tools and materials that may be used to perform the QA tests?

- Software tools provided and validated by CAD manufacturers or third party providers for:
 - saving and retrieving image data to and from user-designated QA directories,
 - marking, validating and storing appropriate reference standards in a CAD manufacturer-specified format for automated scoring of the CAD output,
 - storing CAD system output, such as CAD mark locations, lesion malignancy estimates, etc. if it is not an option available to users already,
 - reading CAD system output, which may be stored as DICOM objects/files, for any user-selected cases (such a reading tool is important for many purposes, especially QA, and is not commonly available like a DICOM image reader),

Q3. What are the tools and materials that may be used to perform the QA tests?

- displaying CAD results for QA purposes,
- calculating performance measures (see responses/recommendations to Q4).
- A set of clinical images with appropriate known reference standards provided by the CAD manufacturer or collected at the clinical site for the CAD system as described above.
- Phantoms that are validated for a specific CAD QA purpose with a validated test procedure.

Q4. What are the performance measures to be used?

- Performance measures depend on the application of the CAD system, the claims of the CAD systems, and the purpose of the QA tests. For testing CAD system performance, the measures may include, for example:
 - CAD detection mark rate over time and stored in the CAD system (display time-series graph at the user demand for QA)
 - Sensitivity
 - Segmentation accuracy
 - Malignancy/benignity assessment accuracy.

Q5. What QA results should be captured and how should they be reported?

- The following data may be captured, tracked and reported from either visual assessment or automated computer analysis:
 - Log files with:
 - date and reasons to perform the QA tests,
 - personnel who perform the tests,
 - QA procedures or tasks performed,
 - location where the test data and results are stored if electronic storage is employed.
 - Performance measures (see responses to Q4).
 - Screen shots capturing CAD results on the image.

Q5. What QA results should be captured and how should they be reported?

- Noticeable changes in performance over time or when compared with product specifications.
- Failure in communication between CAD systems and other necessary devices (e.g., acquisition system, display workstation, other imaging devices).
- Record of intermediate results that are deemed to be necessary or critical that are part of the commercial product output (e.g., segmentation of the organ, feature values).

Q6. What are the criteria of success/failure of the tests (or minimal requirements)?

- The tolerance limits or criteria of success/failure for each specific QA test will depend on the application, sample size of the test dataset, and require further research and discussion.

Q7. What should be done if the CAD QA test fails?

- If the CAD QA procedure fails, the following actions should be considered:
 - Trouble-shooting of whether a change in hardware or software that may or may not be part of the CAD system (e.g., a display workstation software upgrade or a change in image detector) affects CAD performance,
 - Investigation of whether the site's own test data set is representative of the larger patient population,
 - Investigation to assess impact of changes in performance and alert site users of potential problems,
 - Contact device developer or manufacturer.

Guiding Principles

- The QA procedures should be relatively easy and simple to perform
- The QA procedures should not result in substantial financial cost or burden for the manufacturers or end users

Practical vs. Best Practice Approaches

- Smaller dataset vs. sufficiently large dataset
- Visual review of the results vs. automatic analysis of results

Practical Approach

- Perform acceptant testing at installation and subsequent upgrades independently or with the assistance from manufacturers.
- Use manufactures' test sets or site-collected images, which are expected to be relative smaller number.
- Perform visual assessment of CAD results displayed on the images to spot unusual behavior.
- Screen capture of images with CAD results are recommended.
- With upgrades, the site is encouraged to use same images if applicable and compare results before and after upgrades, although same results are not expected.

"Best Practice" Approach

- Sites are expected to use software tools to collect a sufficiently large dataset over time and collect truths (e.g., cancer location) according to the reference standards specified by manufacturers.
- Sites are expected to initiate automatic QA run on test cases in batch mode and report sensitivity and specificity against the reference standards.
- Sites are encouraged to review results visually as well to assess performances specific to lesion types.
- Use manufacturer-provided software to track the number of CAD marks per image over time to detect any unusual behavior of a detection system.

"Best Practice" Approach

- It requires additional effort, cost and experience from the clinical sites.
- The interpretation of test results (e.g., sensitivity) may be more complex when comparing to the testing with manufacturer's data set.
- The CAD manufacturers will need to provide software tools to collect data and reference truths and calculate sensitivity.

Summary

- Tacking the number of CAD marks per image over time is the most efficient and effective way to detect any unusual behavior of a detection system.
- Practical approach is considered to be the minimum CAD QA requirement.
- Collecting a sufficiently large data set at clinical sites may require significant effort, additional tools, cost and experience and proper training to implement.
- The “best practice” approach is intended for sites which have the need and the capability.

Summary

- The motivation of the additional effort is that it will not only provide a higher level of QA to monitor CAD performance in local patient population, but may also improve the effectiveness of using CAD as end users can better understand the CAD performance in the local population and adjust how they work with CAD accordingly.
- Due to the diversity of CAD system, the QA procedures discussed herein do not encompass all possible tests that are important and necessary.
