QIBA VolCT Group 1B Update WebEx
Thursday, Feb. 12, 2009
12 PM (CST)

Draft Call Summary

In attendance:

Michael McNitt-Gray, PhD
Kristin Borradaile
Charles Fenimore, PhD
Robert Ford, MD
John Lu, PhD
Nicholas Petrick, PhD
Daniel Sullivan, MD
Binsheng Zhao, PhD

RSNA
Susan Anderson
Joe Koudelik

General Discussion

Dr. McNitt-Gray addressed Dr. Gottlieb's question concerning Group 1B project aims and existing studies:

"One major question I pose is: What do we have at the "end of the day" once we complete these "experiments"? Much of this has already been documented. A simple review of existing studies may suffice. The more challenging experiment is to apply RECIST, WHO, perception based (what we are piloting) and volumetric assessments to "real life" patient data sets with documented outcomes (survival, time to progression, path proof, etc.). Can we expedite this transition?"

- Are many of Group 1B's questions already understood?
- Particular references concerning existing studies would be helpful - all group members encouraged to forward references
- Mark-up ideas needed
- Proposed studies may have been done with phantoms - Group 1B to expand on this with patient datasets
- MSKCC Coffee Break experiment has not been done with patients, thus novel dataset here
- Attempting to isolate sources of variability would be a worthwhile first step (i.e., what variable need to be controlled)

Dr McNitt-Gray provided a brief overview of Question 1, the group's first project

- What level of Accuracy and Precision (aka Bias and Variance) can be achieved in measuring tumor volumes in patient datasets
    - Accuracy - the difference between the "true" and measured volume
        - High Accuracy = Low Bias
    - Precision - how reproducible measurements are
        - High Precision = Low Variance
    - Specific Aims
        - To investigate both bias and variance of both readers and algorithm-assisted readers in measuring volumes, diameters and bi-directional diameters of lesions

        - To investigate inter and intra-observer variability in each task
    - Materials and Methods
        - Use LIDC patient datasets (with known lesion sizes; already contoured by four readers)
        - Volumes and contours known
        - Single time points only
        - Still need to identify which nodules to use for study
        - Similar as in BioChange 2008, lesions are identified and coordinates provided to readers
        - These datasets have been provided to Dr Ford (RadPharm)
    - Workflow and Reader Tasks
        - Dr Ford will manually annotate (i.e., circle) lesions and forward to readers
        - Readers to manually mark two diameters (longest diam and perpendicular diam) w/o LIDC marks
        - LIDC longest and perpendicular diameters of LIDC datasets (considered "standard") to be compared to reader measurements
        - Semi-automated, algorithm-assisted contouring performed by readers
        - No reader will see measurements made by other readers (blinded study)
        - Siemens software to assist with semi-automated contouring of lesions to assist with volume determination (w/o LIDC mark-up)
        - Some readers to perform duplicate case reads

- LIDC annotation process described
    - Readings done in a two phase approach
        - Note: A "truth committee" not used here
    - First reading done independently
    - Four readers contoured
    - Four readers make independent decision of other three readers
        - This allows for each reader expression without forcing a consensus
        - Probability maps can be constructed here leading into a possible standard
        - Statisticians may be able to take variability into account here
    - Does a reasonable approximation of "truth" rule exist?
    - LIDC datasets didn't track readers; this will be necessary information for Group 1B

- Need to settle on one metric and defend
    - Group 1B to examine a few cases to determine metric criteria and set standard
    - Threshold value needed for consensus - average contour with standard deviation proposed
    - Volume as a metric - 50th percentile with know variation

- LIDC data is to be used as standard, not "truth"
    - Estimate bias of each reader by comparing:
        - Volume
        - Diameter (manually and assisted)
        - Product of diameters (LD x PD) both manually and assisted
    - Estimate inter-reader variability by comparing:

        - Reader vs. Gold Standard
        - Reader vs. Reader
    - Estimate intra-reader variability
        - Group 1B needs more guidance here
    - QIBA readers vs. LIDC readers
        - Group 1B not looking to test bias of LIDC readers
        - LIDC readers to provide the comparison standard to measure QIBA readers against
        - Establish bias of QIBA readers compared to LIDC readers

Remaining Questions

- Number of cases required
- Number of readers required
    - 5 readers needed for reasonable level of variance determination (i.e. statistical analysis)
    - RadPharm to provide up to 10 readers if needed, 5 to begin the study
    - Website - Nancy Obenchowski - to help determine reader number
- Number of nodules, size and shape (i.e. case composition)
- Enough of each lesion subgroup to perform statistical analysis
    - Reader bias with spherical nodules
    - Reader bias with spiculated nodule
- Are both size and shape subgroup analysis needed?

Next Steps

- Dr McNitt-Gray to contact Dr Gottlieb for more details concerning references to existing studies
- Discuss Question 3
    - What is the minimal detectable level of change that can be achieved when measuring tumors in patient datasets under a “No Change” condition?