QIBA

CT Volumetric Group

Project (1a): Inter-scanner / Inter-clinic Comparison of Reader Nodule Sizing in CT Imaging of a Phantom

Version: 2.0

Date: 19 September 2011

Review and Approval

Role	Name, Affiliation	Signature	Date
Principal Investigator	Mike McNitt-Gray		
Biostatistician	Hyun Jung Kim		

Revision History

Version	Changes	Compiled by	Date
1.0	Initial Version.	Hyun Jung Kim	18-Sep-11
2.0	First Revision.	Mike McNitt-Gray	19-Sep-11

SCOPE

This document describes an interim status report regarding several aspects of (1a) Inter-scanner / Inter-clinic Comparison of Reader Nodule Sixing in CT Imaging of a Phantom study. These include both imaging protocol and statistical design aspects of the study. The study design in reading order to fit our current data and a revised power calculation in testing equivalence of the level of15% in relative bias has been discussed.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This document describes interim report of both the imaging protocol and statistical support in 1a study. Investigators on the study requested that the statistical design be revised to consider that an equivalence test is appropriate for showing the QIBA profile of 30% variation in CT volumetric measurements.

The following have been completed

1) The imaging protocol was finalized for both arms of the study: the ACRIN 6678 arm as well as the Quality Performance Arm. These were completed prior to this reporting period. NOTE: Imaging has also been completed at all sites using both protocols. All image data has been collected and curated.

2) A revised excel document that described the reading randomization order has been delivered. The revision was to take into account the number of images, number of sites and number of readers

3) The results of the power analysis showed that using 462 combinations (66 images *7 readers) with 0.10 correlation

For the differences in relative bias within 12% and standard deviation being less than 20, we shall have more than equal to 80% power to show the equivalence within 15% threshold under 1a data set.

4) In subgroup analysis for nodule >5mm, 308 (44 images *7 readers) with 0.10 correlation

For the differences in relative bias within 12% and standard deviation being less than 15, we shall have more than equal to 80% power to show the equivalence within 15% threshold under 1a data set.

REPORT

This report describes some additional details for the interim report of 1a:

A. Design for Reader Order

Reader study to characterize uncertainty in volume and other reader-based sizing of nodules in CT imagery collected on scanners from several vendors. The imaging protocol is to include an ACRIN 6678 branch and an image quality-based, device-independent branch.

Relative Bias: (measured size –true size)/true size*100

- 5 site / device**:** 4 manufactures will be a contributor for differences in device effect.
- 2 imaging protocol factors; ACRIN6678, imaging quality-based, device-independent branch
- The 6 phantom nodules (-10HU, 3 spherical and 3 spiculated in size of 5, 10, 20mm) scanned at thin slice thickness
- 7 Readers

The Reading in stratified randomized order has been revised, since one site has same protocol in ACRIN6678 and imaging quality-cased. Thus, 66 HRCT images from the 6 phantom scanned under the two protocols at 5 sites and only under imaging quality-based protocol at one site at thin slice thickness, will be read by radiologists (66 images=6 phantoms*5 sites*2 protocols + 6 phantoms*1 site*1 protocol).

The Figure 1 is a snap shot of reading order in stratified random order

Figure1. Reading Order for 7 readers

reader,Volumetric,phantom ID (1,2,3,4,5,6), scanner (A,B,C,D,E, F*),-protocol (-1 -2); F* has only one protocol as listed -2

reader1	reader2	reader3	reader4	reader5	reader6	reader7
1V4C-2	2V6D-1	3V3E-2	4V6C-2	5V4C-1	6V4D-2	7V2B-1
1V5B-2	2V3C-1	3V4A-1	4V1C-2	5V3D-1	6V3C-1	7V3C-2
1V3B-1	2V5C-1	3V6E-2	4V2C-2	5V1C-2	6V2C-1	7V6E-2
1V6B-2	2V1B-1	3V2B-1	4V4C-1	5V6E-1	6V1D-2	7V5B-2
1V4F-2	2V2E-1	3V1D-1	4V5B-2	5V4F-2	6V6B-1	7V4E-2
1V1E-1	2V5F-2	3V5C-1	4V3B-2	5V5A-1	6V5C-2	7V3F-2
1V2E-2	2V4A-2	3V6C-2	4V1B-2	5V2E-2	6V3C-2	7V1C-2
1V3A-2	2V6A-1	3V1B-2	4V6A-1	5V4D-1	6V2A-2	7V4C-2
1V1F-2	2V1A-1	3V3B-2	4V4C-2	5V5B-1	6V6A-1	7V6F-2
1V5A-2	2V5C-2	3V2E-2	4V3C-1	5V6D-2	6V1A-1	7V3E-2
1V2E-1	2V4D-1	3V4B-2	4V2E-2	5V3A-2	6V4C-2	7V2A-1
1V1C-2	2V3E-2	3V5B-1	4V5E-1	5V1A-1	6V2F-2	7V6A-2
1V4B-2	2V2B-2	3V2D-1	4V6D-1	5V2A-1	6V5E-2	7V5C-1
1V6A-1	2V5B-2	3V1E-2	4V1F-2	5V6F-2	6V6F-2	7V1D-2
1V5C-2	2V2B-1	3V3E-1	4V3D-2	5V4E-2	6V4B-1	7V6C-1
1V4E-1	2V6F-2	3V4D-2	4V2A-1	5V2B-2	6V1C-2	7V1C-1
1V3E-2	2V4D-2	3V6B-2	4V5D-2	5V3E-1	6V3E-1	7V2C-1
1V2D-2	2V3A-2	3V3F-2	4V4D-2	5V6E-2	6V6E-2	7V5B-1
1V1D-2	2V1C-2	3V5E-1	4V1E-1	5V1D-2	6V5A-1	7V3B-2

B. Reconsideration in Statistical Analysis and Power Calculation

B.1 Rationale and Revised Hypothesis

After the recent profiles in CT Volumetric Group from QIBA for the 30% threshold as a measurement volumetric variation in a target lesion, this rational of testing the threshold has been rationally considered as a new interest hypothesis. Instead of showing difference effect of sites, the group interest is to show a *similarity* from readers' measurements in this study within a 15% (half of 30%). Thus, equivalence test was set to this primary interest. [Blackwelder 1998, Chow, S.C. 2003, Devroye, Luc. 1986]

The revised primary endpoint is that a difference of intra scanner/intra clinic effects in relative bias from volume is within 15%.

The null hypothesis (Ho) and alternative hypothesis (Ha) is listed below:

H_o: *The relative bias is greater than 15% across devices and protocols.*

vs.

H_a: *The relative bias is within 15% across devices and protocols.*

In mathematical form, ***H_o:*** $\mu 1-\mu 2 \leq -\delta$ *or* $\mu 1-\mu 2 \geq \delta$ *vs.*

H_a: $\mu 1-\mu 2 \geq -\delta$ *and* $\mu 1-\mu 2 \leq \delta$

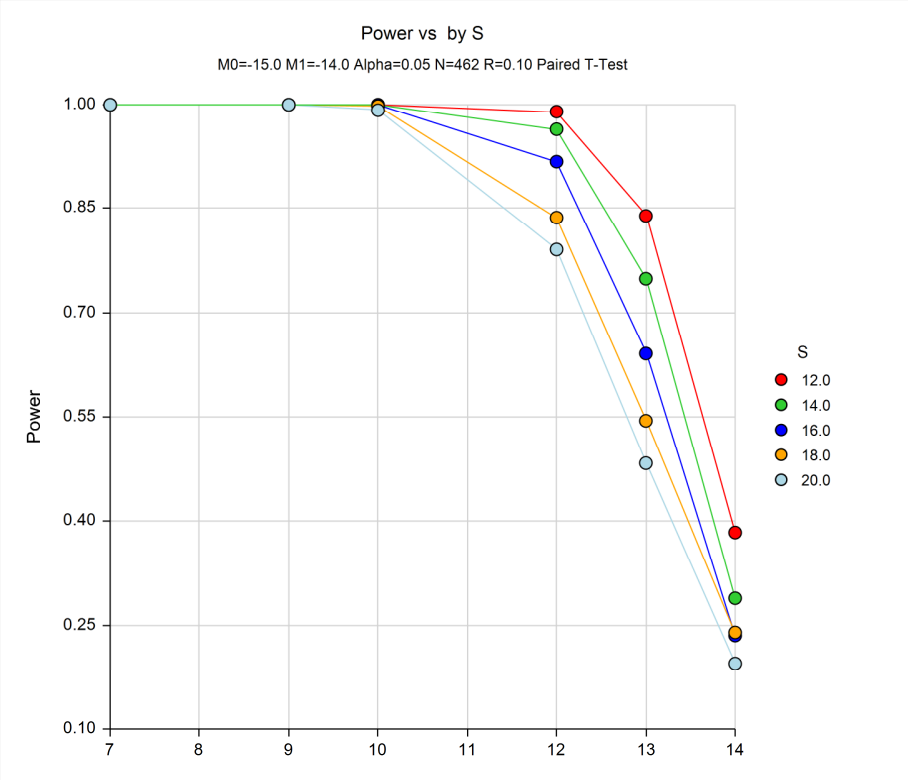
, where $\mu 1$*= relative bias,* $\mu 2$*=0, and* δ*=15%*

The revised primary endpoint is that a difference of intra scanner/intra clinic effects in relative bias from volume is within 15%. Considering the 30% is a claim for the CT volumetric group, the approximately the half of 30% (i.e. 15%) shall be tested as a mean relative bias from this study.

In the previous 1A study under single scanner from one site, the standard deviance in relative bias from 6 readers' volumetric measurement in 0.8mm slice thickness was 13% with mean of 0.5%. In our study, we expect the standard deviance will increase due to multi-center and multi-device effect. The readers were measuring the 6 nodules CT images scanned under different protocols and site; we expect that there will be a correlation the repeated measurement. Since the power of equivalence test proportionally increased by the correlation, we set a correlation 0.10 conservatively.

B.2 Summary of Result

Power vs. Difference in relative bias where N=462 (66 images *7 readers) with 0.10 correlation by varying SD.



As long as we have differences in relative bias within 12%, we have >80% power to show the equivalence within 15% up to standard deviation of 20.

Summary Statements

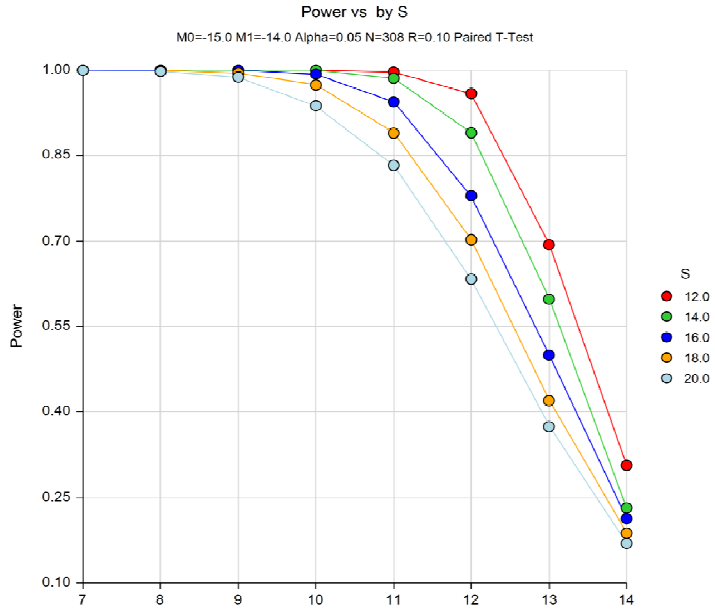
A sample size of 462 pairs with a correlation of 0.10 achieves >80% power to detect equivalence when the margin of equivalence is from -15.0 to 15.0 and the actual mean difference is ranged from 7% to 12%. The significance level (alpha) is 0.050 using two one-sided Paired T-Tests. These results are based on 2000 Monte Carlo samples from the null distribution: Normal(M0 S) -Normal(M0+15,S) and the alternative distribution: Normal(M0,S) - Normal(B, S), where S is ranged from 12 to 20 and B is ranged from 7% to 12% as indicator of actual differences.

B.3 Secondary and Exploratory Testing of Small Lesions

The relatively small lesion size of 5mm is included in this study, whereas the nodule size of the previous study were 10mm, 20mm, and 30mm and the lower limit of revised RECIST 1.1 is suggested to be greater than 5mm.

A subgroup analysis will be performed only including greater than 5mm (i.e. 10mm and 20mm) under the same hypothesis of 15% equivalence limit. The data set will 308 images (44 images *7 readers). Power calculation were performed under the same criteria as the primary analysis.

Power vs. Difference in relative bias where N=308 (44 images *7 readers) with 0.10 correlation by varying SD.



As long as we have differences in relative bias within 12%, we have >80% power to show the equivalence within 15% up to standard deviation of 15.

Summary Statements

A sample size of 308 pairs with a correlation of 0.10 achieves >80% power to detect equivalence when the margin of equivalence is from -15.0 to 15.0 and the actual mean difference is ranged from 7% to 11%. The significance level (alpha) is 0.050 using two one-sided Paired T-Tests. These results are based on 2000 Monte Carlo samples from the null distribution: Normal(M0 S) -Normal(M0+15,S) and the alternative distribution: Normal(M0,S) - Normal(B, S), where S is ranged from 12 to 20 and B is ranged from 7% to 12% as indicator of actual differences.

Additional exploratory analysis shall be tested:

For volume

- Equivalence of Pair t-test on measured in volume per each nodule,

Also for 1-D and 2-D measures from 3D segmentation

- Equivalence of Pair t-test on relative error: (measured size –true size)/true size,
- Secondary outcome: Equivalence of Pair t-test on measured volume, for each nodule,
- Other Equivalence of Pair t-test is possible (e.g. on log transformed data).

C. Finding and Discussion

For the differences in relative bias within 12% and standard deviation being less than 20, we shall have more than equal to 80% power to show the equivalence within 15% threshold under 1a data set. The revised hypothesis shall be contributed as a finding for the profile for QIBA CT volumetric group.

REFERENCES

Blackwelder, W.C. 1998. 'Equivalence Trials' In Encyclopedia of Biostatistics, John Wiley and Sons. New York. Volume 2, 1367-1372.

Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.

Devroye, Luc. 1986. Non-Uniform Random Variate Generation. Springer-Verlag. New York.
Matsumoto, M. and Nishimura,T. 1998. 'Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator.' ACM Trans. On Modeling and Computer Simulations.