5-18-2012

Emulating the volume of pulminary nodules: A Quantitative approach

Meagan Kuhfahl

ESTIMATING THE VOLUME OF PULMONARY NODULES: A QUANTITATIVE APPROACH

by

Meagan P. Kuhfahl

A THESIS

Submitted to
Rochester Institute of Technology
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

Applied and Computational Mathematics

at the

School of Mathematical Sciences
College of Science
Rochester, NY 14623

on

May 18, 2012

Rochester Institute of Technology
School of Mathematical Sciences

Applied & Computational Mathematics Program

Master's Thesis

Applicant's Name: Meagan Kuhfahl
Please Print

Applicant's Signature: _____ **Date:** 05/18/12

Thesis Title: Estimating the Volume of Pulmonary Nodules: a Quantitative Approach

Proposed Defense Date: 05/24/2012

Advisor's Name: Nathan Cahill **Signature:** _____

Committee Member 1: Elizabeth Cherry **Signature:** _____

Committee Member 2: Linlin Chen **Signature:** _____

Committee Member 3: Peter Burns **Signature:** _____

Graduate Programs Director: _____ **Date:** 05/24/12
Dr. Tamas Wiandt

To complete your program, you need to secure all necessary signatures in time for graduation. A one-page summary describing your thesis and an overview of the area of your thesis should be attached. Describe the principal ideas and deliverables of your thesis. Annotated references, books, papers, URLs and previous master's projects or theses should be included.

Abstract

Doubling time for small pulmonary nodules is an important indicator used to diagnose lung cancer, a leading cause of death in the United States. The volume of the nodules is measured using computed tomography (CT) scans. Each volume measurement comes with a degree of uncertainty, which in turn increases the uncertainty for the doubling time measurement. Decisions regarding risky and expensive patient treatment depend on doubling time, so accuracy is important. The volume of nodules is estimated by taking a series of points marked on CT scans by radiologists and connecting these points to make a boundary. This boundary includes whole and partial pixels. By including and excluding partially filled pixels, the estimation errors can be quantified to ensure that a more accurate error estimation is made, allowing clinicians to make a better informed treatment decision. Since this process requires a radiologist to manually mark CT scans, there is a possibility for variation between radiologists, and it is time-consuming. A semi-automated method would be useful for measuring volume because it would reduce variation from radiologists' opinions and time. We can use Gaussian weighted integration to eliminate the need for radiologists to mark points on a scan. Instead, Gaussian weighted integration requires only a square boundary centered at the nodule. A Gaussian mask is applied and volume estimations are made. By simulating two scans per patient, the accuracy of each method is measured by statistical comparison with the original volume calculations, or the ground truth.

Acknowledgements

I would like to thank my advisor, Nate Cahill for his continuous guidance and for putting a smile on my face even on days when I thought my head was going to fall off. For their encourgement and financial support, I thank my parents, Dave and Pat, my brother, Tom, and my sister, Sandie. For his knowledge of English and willingness to help me with mine, I thank

my friend, Tim Shank. I would also like to thank my friends who have a curiosity for my thesis and for being there when I needed a break from it. For always wagging her tail and reminding me when it was time to go for an adventure outside, Velvet. A special thank you to Eli Putzig for helping me get through this year.

I acknowledge the National Cancer Institute and the Foundation for the National Institutes of Health, and their critical role in the creation of the free, publicly available LIDC/IDRI Database used in this study.

1 Introduction

In this section, we will discuss the background information and the problems which arise when estimating the volume and doubling time of pulmonary nodules. Then we will look at the nodule volumes according to the radiologists' input, introduce a semi-automated method for estimating volumes, and finally discuss the purpose of this thesis. MATLAB R2011a was used for the computation of all methods.

1.1 Problem

Abnormal growths in the lungs, or lung nodules, may be cancerous tumors. One useful tool for identifying tumors is doubling time. Computed tomography (CT) scans are used in determining doubling time. Low-dose spiral CT scans are found to be more effective than x-rays, but they also show cysts, nodules, and other growths that may or may not be cancerous. If a lung nodule is spotted, it is possible to determine whether it is cancerous by performing a lung biopsy. However, with each biopsy, there is a risk of pneumothorax (a collapsed lung) or hemorrhage.

In order to mitigate these risks, clinicians typically estimate the volume of

the nodule and then ask the patient to return for another screening. The volume is again estimated and compared to the first volume estimate to calculate the doubling time (DT) according to an exponential growth model given by

$$V = V_1 e^{kt} \tag{1}$$

where V is the projected volume, V_1 is the initial volume and k is the growth rate constant. From this equation, the formula for doubling time, DT, can be derived to be

$$DT = t\frac{\ln(2)}{\ln(\frac{V_2}{V_1})}, \tag{2}$$

where t represents the time between screenings in days, V_1 is the volume estimate for the first scan, and V_2 is the volume estimate for the second scan. Typically, the time between scans is 20-71 days [5]. If the doubling time falls within a certain range indicating that the nodule is likely cancerous, a lung biopsy may be performed. According to [8], this range is "30 days to several hundred days." Figure 1 shows seven more opinions of what this range is. Growths with doubling times outside this range are considered benign.

Like any estimation process, measurement of nodule volumes and doubling times has the potential for error. This margin of error can mean the difference between performing successive x-rays, CT scans or biopsies. Borderline cases are likely to be treated as risky and tested further. Biopsies, CT scans and x-rays come with health risks and are expensive. Therefore, it is important that the nodule volume-estimation is accurate, and it is equally important to understand and quantify the effects of estimation error on the resulting clinical decision process.

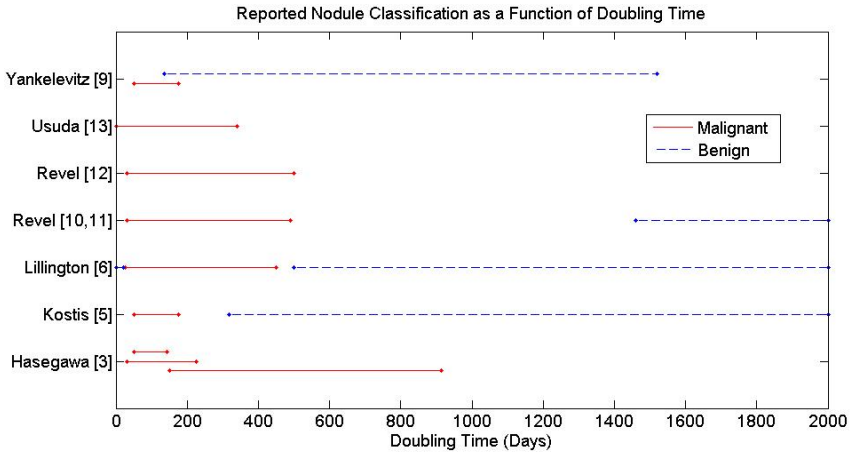


Figure 1: Malignancy and benignity based on doubling time

1.2 Ground Truth

The Lung Image Database Consortium (LIDC) provided by the National Cancer Institute is a database of real patients' CT scans. This data is available to "provide investigators the opportunity to evaluate a wide range of technical parameters and de-identified clinical information within this database that may be important for research applications" [4]. The LIDC currently supplies data from 1,010 patients. The patients' identities have been removed to ensure anonymity. Four radiologists have separately examined these patients' scans and categorized the lesions as a "nodule ≥ 3 mm," "nodule < 3 mm," or "non-nodule ≥ 3 mm" [1]. The radiologists then reviewed their individual markings and together came to a unanimous conclusion concerning the 44.1% where at least one had disagreed. The LIDC contains only nodules ≥ 3 mm.

Since the start of this thesis, scans have been added to the LIDC. Of the patients available currently, we used scans from 264 patients, each with between 1 and 25 nodules. The provided images of the nodules are three-dimensional. Radiologists have marked points on slice images of the CT scans where they

believe the boundary of each nodule is. We used these points to extract two-dimensional slice images, of which the size is the larger of 64 by 64 pixels and twice the largest diameter of the nodule by twice the largest diameter of the nodule. On each slice, we connected the points to create a boundary.

Since the boundary is formed by a relatively small number of points, it is not as accurate as possible. In addition, some nodules may not have distinctive boundaries. This inaccuracy affects the volume estimates. Figure 3 shows a nodule with the points marked by a radiologist connected, which is derived from the larger axial slice shown in Figure 2. The marked points will be used to determine the volume of the nodules, and we will call this volume the ground truth. Given the issues with marking nodules, the ground truth will have inherent uncertainty.

1.3 Semi-Automated Method

One primary goal of this thesis is to construct a semi-automated method to determine the volume of a nodule without requiring that a radiologist mark the boundaries, but only mark a box containing the nodule. This would eliminate variation from radiologists' opinions and eliminate a time-consuming manual step from the process. We will devise an automatic method based on Gaussian weighted integration to estimate the volumes. This method needs to yield results with equal to or better accuracy than the ground truth.

1.4 Organization of the Thesis

In Section 2, we will establish a method for calculating nodule volume using the radiologists' markings. This will include introducing where the data comes from, a scheme for classifying nodules according to their type, the algorithm used, and how manually marking scans can alter results. Section 3 will propose the

Figure 2: Axial slice with a nodule located in the center of the yellow box

semi-automated method based on Gaussian weighted integration for computing the nodule volumes. In Section 4, we will analyze the results from the two methods to determine whether their differences are statistically significant. We will compare the methods and discuss their usefulness. Finally, in Section 5, we will discuss the possibilities for future work.

2 Establishing the Ground Truth

In this section, we will first discuss the LIDC data and then establish a scheme for classifying nodules according to their type. Next, we will explain how vol-

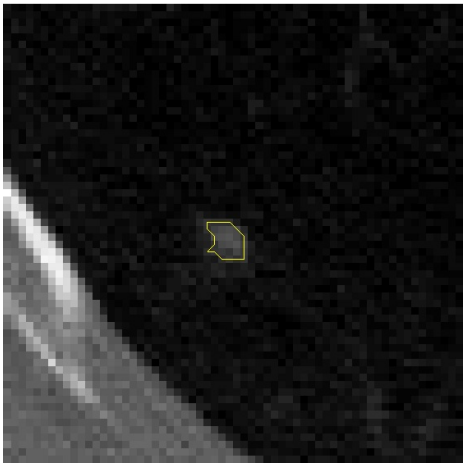


Figure 3: Cropped and marked image of the nodule

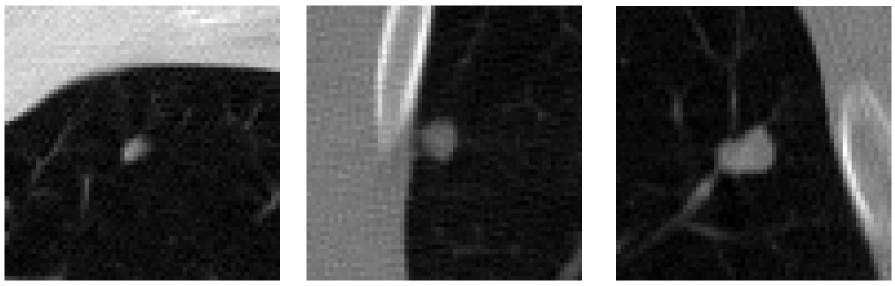
umes are computed from the manually labelled point and regions , and finally we will look at how manually marking scans can alter results.

2.1 Data

The Lung Image Database Consortium (LIDC) provided by the National Cancer Institute is a database of real patients' CT scans. There are 1,010 patients and 244,527 scans available. We use 264 of these patients each with between 1 and 25 nodules. In preprocessing, we made each file represent one nodule by cropping the image around the nodule so that it lies in the center. The size of this region is the larger of 64 by 64 pixels and twice the diameter of the nodule. Tied to each nodule is unique data. The images given are slices of the nodule taken a given distance apart. There may be anywhere from 2 to 40 slice images for each nodule. Corresponding to each slice is an ordered list of three dimensional coordinates determined by the radiologist. They represent the border of the nodule.

2.2 Classifying Nodules

There are three types of nodules. Their classification depends on how the nodules appear in the CT scans. Isolated nodules, as seen in Figure 4(a), are well-circumscribed and have no connections. The boundaries of isolated nodules are the easiest to see. Juxtapleural nodules are attached to the lung wall. As one can see in Figure 4(b), the boundaries of these nodules are more difficult to distinguish from the lung wall. The last type of nodule is juxtavascular. These nodules have vascular connections, and like juxtapleural nodules, the boundaries can be difficult to detect. Figure 4(c) shows an example of a nodule with vascular connections. In our data, there were 127 isolated, 167 juxtapleural and 149 juxtavascular nodules.



(a) An Isolated Nodule (b) A Juxtapleural Nodule (c) A Juxtavascular Nodule

Figure 4: Classification of Nodules

2.3 Algorithm

For each slice of the CT scan, we are given an ordered set of points designated by the radiologist to mark the boundary of a nodule. We will call these click points, and denote them by (x_j, y_j), where $j = 1, 2, .., m$ with m being the total number of points. Note that each click point is located at the center of a pixel, so each click point corresponds to one pixel. First, the extraneous information of the area surrounding the nodule is eliminated. We use the four

most extreme click points ($\min(x_j)$, $\max(x_j)$, $\min(y_j)$, $\max(y_j)$), and cut off the image information extending one pixel beyond the points. Using the click points, we find the centroid of the nodule slice i, c_i where $i = 1, 2, ..., n$ with n being the total number of slices. Then a vector, $\vec{v}_j$, from each point, j, to the centroid, c_i, is created, and also from each successive click point, creating a series of triangles t_k where $k = 1, 2, ..., m$. This is illustrated in Figure 6 where the yellow stars are click points, the red lines are the vectors connecting the center point to the click points, and the grey area represents the inside of the nodule. The area of each triangle, T_k, is found using the *norm* of the cross product of the vectors, $\vec{v}_j$, coming from the centroid, c_i. That is,

$$T_k = \frac{\|\vec{v}_j \times \vec{v}_{j+1}\|}{2} \tag{3}$$

where $j = 1, 2, ..., m-1$ and $k = 1, 2, ..., m-1$. The triangle made from the first and last click points is computed separately by $T_m = \frac{\|\vec{v}_m \times \vec{v}_1\|}{2}$. The area of each slice, S_i can then be calculated by summing the triangle areas.

The volume of the nodule, V, can be computed with the Trapezoid Rule using the slice areas, S_i; i.e.,

$$V = \sum_{i=1}^{n-1} V_i = \sum_{i=1}^{n-1} \frac{(S_i + S_{i+1})}{2} \cdot \Delta z \tag{4}$$

where V_i is the volume between slice i and slice $(i+1)$, $i = 1, 2, ..., n-1$, and Δz is the distance between slices. Typically, the distance between slices is 1.5mm. Figure 5 is an illustration of how the Trapezoid Rule can be used to estimate the volume of a nodule visible in three slices. Algorithm 1 shows the steps of finding the volume of a nodule, V.

Algorithm 1 Method for finding the volume of a nodule, using boundaries determined by a radiologist.

for slice i **do**

Find the centroid, c_i of the click points by finding the mean.

Create t_k by creating vectors, $\vec{v}_j$ connecting c_i and by connecting each consecutive click point.

for each pair of consecutive click points $(x_j, y_j), (x_{j+1}, y_{j+1})$, where $j = 1, 2, ..., m-1$ **do**

Find the area T_k of each triangle using Equation (3)

end for

Find the area T_m of the triangle connecting the first click point to the last click point: $\frac{\|\vec{v}_m \times \vec{v}_1\|}{2}$

Compute $S_i = T_1 + T_2 + ... + T_m$

end for

Compute the volume between each consecutive slice, V_i using Equation (4)

Compute the volume of the nodule: $V = V_1 + V_2 + ... + V_n$

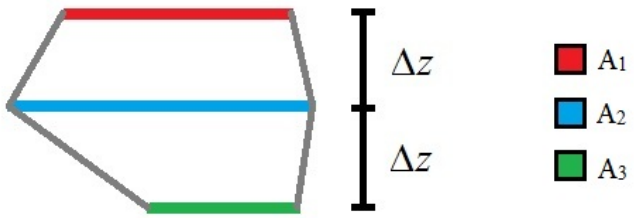


Figure 5: Finding the volume using the Trapezoid Rule

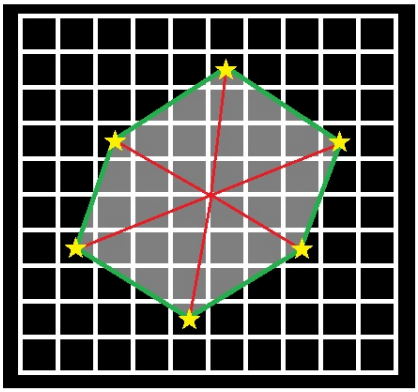


Figure 6: Vectors connect the click points and their centroid to create triangles

2.4 Bounds

Since Algorithm 1 automatically connects the click points with straight lines, the pixels in which click points are located are not fully included in the calculation. If the radiologist marked the next pixel over, what would the difference in volume be? Figure 7 is a model of four pixels with part of a nodule cutting through. The yellow dots are potential click points and the grey region indicates the inside of the actual nodule boundary, represented by the green line. A radiologist would likely choose the points in pixel b and c because they lie close to the boundary. They also might pick the point in a or d. We want to examine what the difference in nodule volume would be when varying click points to nearby pixels. To illustrate this difference in area of one slice, see Figure 8. The blue area represents the portion included in the volume calculation. The yellow stars represent click points, and the green lines are the boundary created when these points are connected.

In this section, we will look at how the pixels on the boundary affect the volume by including and exluding them in the volume estimation. We call the volume which excludes partial pixels the lower bound; likewise, the volume which

includes partial pixels is the upper bound. For example, for lower bound, pixels in which only one, two, or three corners are considered inside the boundary are excluded from the volume estimate, as illustrated in Figure 8(a). In an upper bound estimate, however, these pixels would be included. Figure 8(c) shows an upper bound estimate of a slice, and Figure 8(b) represents the ground truth.

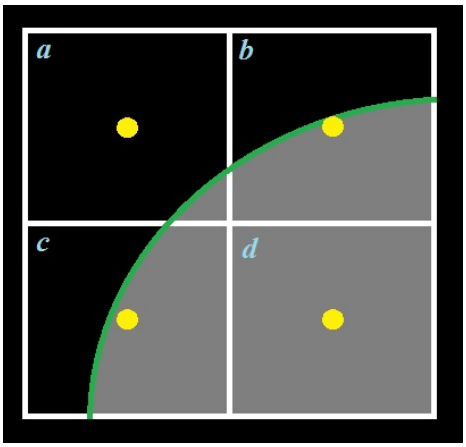
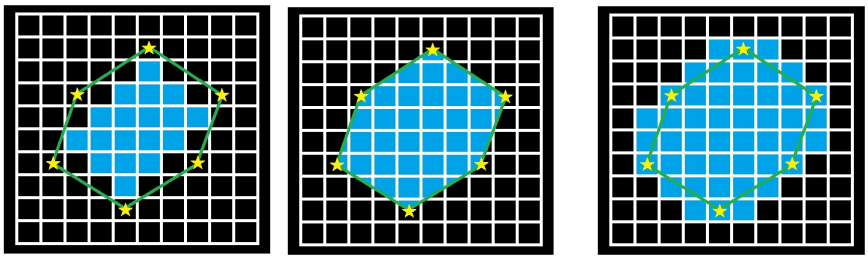


Figure 7: Visualizing how click points are chosen



(a) Lower Bound (b) Ground Truth (c) Upper Bound

Figure 8: The variation in bounds on a slice

Algorithm 2 shows the method used for determining where the pixels are in regards to the nodule boundary. We check each corner of a pixel and label it as "in" if it lies on the inside of the boundary or "out" otherwise. This is used to calculate the upper and lower bounds of the nodule volume. To start,

parametric equations are found to represent the boundary of the nodule, as determined by the click points.

Given two consecutive click points (x_1, y_1) and (x_2, y_2), we can write parametric equations for the line connecting the two click points by:

$$x = (1-t)x_1 + tx_2 \tag{5}$$

$$y = (1-t)y_1 + ty_2 \tag{6}$$

where the point (x, y) is on the line, and t is a real number. Note: if $0 \leq t \leq 1$, then (x, y) is between (x_1, y_1) and (x_2, y_2). Now, we can take the first pixel corner's x-coordinate, x_c, and find where this x-value intersects the parametric equations. We store the y-coordinate of the intersections. Then we add the pixel corner's y-coordinate, y_c, to the list and sort it in ascending order. If y_c appears more than once in the list, then we know that the pixel corner, (x_c, y_c), lies on a parametric line, and therefore lies on the boundary of the nodule. We will say this corner is inside of the nodule. This conclusion is stored in a binary array where each entry represents a pixel corner. Since this corner is "in", it is assigned a 1 in the array. Corners which are "out" are assigned a 0 in the array.

If y_c only appears once in the list, we use the method *Single-shot inclusion* as described in [7] to determine whether or not (x_c, y_c) lies inside of the nodule. This method involves including the location of (x_c, y_c) in the list of intersections. As can be seen in Figure 9, the point x_{in} is located within the form. When a horizontal line is drawn through it, there is an odd number of intersections with the boundary before and after the point. Alternatively, x_{out} is located outside the form, and the line has an even number of intersections with the boundary before and after it. We want to find the points that are located inside the nodule, requiring (x_c, y_c) to have an odd number of intersections surrounding it, and therefore have an even index number, where indices start at one. To find

the volume of the nodules using the lower and upper bound methods, we sum up the pixels in each slice we want to include in the estimate. Then, just like the ground truth, we use the Trapezoid Rule to find the overall volume.

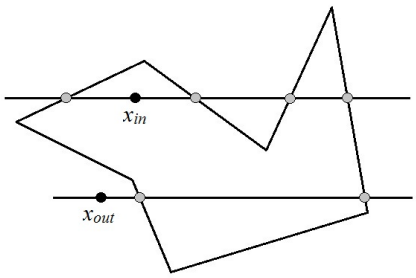


Figure 9: Method for classifying points as inside or outside

3 The Semi-Automated Method for Volume Estimation

In this section, we will look at the algorithm for a semi-automated method for finding relative volumes of nodules. Since doubling time is a function of the ratio of volumes (see Equation (2)), we do not require an absolute volume estimate. We will define the method, Gaussian weighted integration and then adapt and apply this to the CT scans. MATLAB R2011a was used for the computations in this section.

3.1 Mathematical Method

Guassian weighted integration is a method used in image processing for approximating a definite integral in a way that favors information in the center of the image. The method uses a weighted sum of the function, $f(x)$, evaluated at abscissas $x_1, x_2, ..., x_\alpha$ in the domain of integration, $[a, b]$. Typically, x_κ,

Algorithm 2 Method for finding the pixels which are included, excluded and on the boundary of the nodules as defined by the radiologist.

Find parametric equations for lines between consecutive click points to represent the boundary of the nodule in each slice.

for each pixel corner **do**

Use x-coordinate to find intersection with boundary

Store the corresponding y-coordinate of each intersection

Put the pixel y-coordinate in the list of intersecting y-values and sort in ascending order

if the pixel y-coordinate appears more than once in the list of y-values **then**

then the pixel corner is on the boundary

consider the corner inside the nodule

else if the pixel y-coordinate has an *even* index value **then**

the pixel corner is inside the nodule

else

the pixel corner is outside the nodule

end if

end for

$\kappa = 1, 2, ..., \alpha$, are chosen to be equally spaced; i.e., $x_1 = a; x_2 = a + \frac{b-a}{\alpha-1}; x_3 = a + 2\frac{b-a}{\alpha-1}; ...; x_{\alpha-2} = a + (\alpha-3)\frac{b-a}{\alpha-1}; x_{\alpha-1} = a + (\alpha-2)\frac{b-a}{\alpha-1}; x_\alpha = b$ where α is the number of pixels in both directions (since the images are square) in a given slice i. In one dimension, the equation is

$$\int_a^b w(x)f(x)\,dx \approx \sum_{\kappa=1}^{\alpha} w(x_\kappa)f(x_\kappa)\Delta x. \tag{7}$$

where $w(x_\kappa)$ is the weight function, α is the number of points, $\Delta x = \frac{b-a}{\alpha-1}$ and $1 \leq \kappa \leq \alpha$. To extend this to two dimensions, we apply (7) repeatedly, i.e.

$$\begin{aligned}\int_a^b \int_a^b w(x,y)f(x,y)\,dxdy &\approx \int_a^b \left(\sum_{\kappa=1}^{\alpha} w(x_\kappa)f(x_\kappa,y)\Delta x\right) dy \\ &\approx \sum_{\kappa=1}^{\alpha}\sum_{\lambda=1}^{\alpha} w(x_\kappa)w(y_\lambda)f(x_\kappa,y_\lambda)\Delta x\Delta y \\ &= \sum_{\kappa=1}^{\alpha}\sum_{\lambda=1}^{\alpha} w(x_\kappa,y_\lambda)f(x_\kappa,y_\lambda)\Delta x\Delta y. \end{aligned} \tag{8}$$

We can say this is true since $w(x,y)$ is separarable, i.e., $w(x,y) = w(x)w(y)$ because for the weight functions, $w(x_\kappa, y_\lambda)$, we use a two-dimensional Gaussian function:

$$\begin{aligned} w(x,y) &= \exp(-\frac{(x-\frac{b-a}{2})^2 + (y-\frac{b-a}{2})^2)}{2\sigma^2}) \\ &= \exp(-\frac{(x-\frac{b-a}{2})^2}{2\sigma^2})\exp(-\frac{(y-\frac{b-a}{2})^2}{2\sigma^2}) \\ &= w(x)w(y). \end{aligned} \tag{9}$$

This Gaussian function acts as a mask on the CT scans. Resulting is a clearer image isolating the nodule from surrounding noise. The width of the bell varies depending on a chosen σ value. See Figures 10, 11, 12 for examples of varying σ values and how they affect the masks. Larger nodules (in terms of diameter on a slice) would be better suited by a larger σ value to avoid reducing

nodule size, and similarly, smaller σ values would be better suited for smaller nodules.

Let Δx, Δy and Δz be the pixel sizes in the x, y and z directions, respectively. Typically, Δx and Δy are both 0.5mm and Δz is 2.5mm. Since we do not have isotropy in the sampling, we use a two-dimensional mask w for each slice i, where $i = 1, 2, ..., n$, applied in x, y dimensions. Otherwise we could use a three-dimensional mask.

Algorithm 3 Algorithm for applying a Gaussian mask on the slice images.

Load in image data as a three-dimensional array I_i.

Load and store image size and pixel sizes Δx, Δy and Δz

Construct a two-dimensional mask w_i according to Equation 9

for each slice i **do**

  Multiply the slice image by the mask, $w_i \cdot I_i$

end for

After the new masked image slices, $w_i I_i$, are found using Algorithm 3, the array entries can be summed and multiplied by the pixel sizes, $\Delta x \Delta y$, to find the slice area, A_i. We can find the volume, V, by summing each A_i multiplied by the distance between slices, Δz. That is,

$$V = \sum_{i=1}^{n-1} \frac{A_i + A_{i+1}}{2} \Delta z. \tag{10}$$

As you can see in Figures 13, 14 and 15, the mask eliminates background noise, vascular tissue, and even the pleural wall. This method could eliminate the need for a radiologist to mark the CT scans with click points and any error that would result. Instead, we would have the radiologist create a square window around the nodule either 64 by 64 pixels or twice the largest diameter by twice the largest diameter of the nodule, whichever is larger.

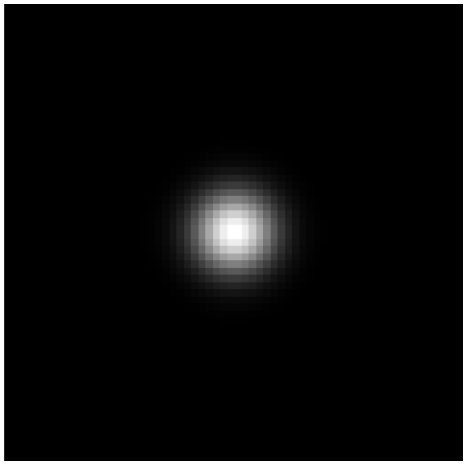


Figure 10: An example Gaussian mask with $\sigma = .06$

4 Results

In this section, we introduce the statistical tests used to analyze the estimated volumes by means of the different methods with the ground truth. We will analyze the isolated nodules, the juxtapleural nodules, and the juxtavascular nodules in order. We are looking to see how well the semi-automated method compares to the ground truth, and how different the lower and upper bounds are from the ground truth. For the semi-automated method, we use Gaussian functions with $\sigma = .01, .06, .11, .16, .21, .26, .31, .36, .41$, and .46 to compare results. We will describe and perform the statistical tests Kruskal-Wallis using Minitab 16 and Wilcoxon Rank-Sum using MATLAB R2011a.

4.1 Statistical Tests

Typically, when a patient is screened once, they are asked to have another screening performed after 20 to 71 days [5]. Since we did not have a second scan for the patients, we simulated a second screening. Considering the isolated, juxtapleural and juxtavascular nodules separately, we randomly selected two

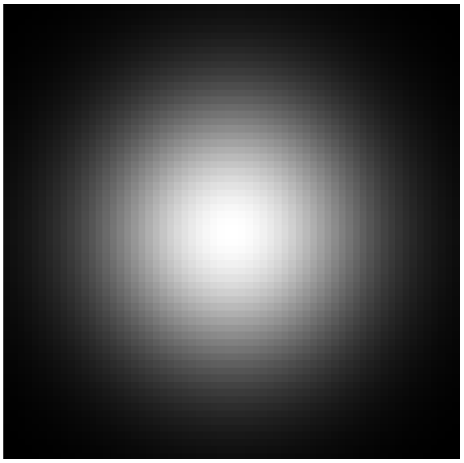


Figure 11: An example Gaussian mask with $\sigma = .21$

nodules from all nodules of the same classification in the LIDC database, and then repeated the process $M = 1000$ times. The volumes of these nodules were then calculated using each method: the ground truth, bounds (lower and upper) and the semi-automated method using ten different σ values.

Across methods, we compared growth rate of the nodules by calculating the *log ratio*, or $\ln(\frac{V2}{V1})$, where V_1 and V_2 are the volumes of the nodule at the first and second screening, respectively. While $\frac{V_2}{V_1}$ gives the growth ratio, the log ratio enables symmetric comparison of growth/decay. For example, if a nodule's volume doubled from $V_1 = 1$ to $V_2 = 2$, then the comparison $\frac{V_2}{V_1} = 2$. However, if we reverse the scenario to be that the nodule's volume halved from $V_1 = 2$ to $V_2 = 1$, then the comparison $\frac{V_2}{V_1} = 1/2$. As one can see, the reverse scenario is not symmetric about zero. In fact, $\frac{V_2}{V_1}$ will never be less than zero. However, $\ln(\frac{V2}{V1}) = \ln\frac{2}{1} \approx .6931$ and $\ln(\frac{V2}{V1}) = \ln\frac{1}{2} \approx -.6931$ which shows the growth versus decay clearly. Note that because we are comparing a function of $\frac{V_2}{V_1}$, we only need relative volume measurements, not absolute.

The log ratio calculation was done with the 13 different volume estimation methods (ground truth, bounds as described in Section 2.4 and ten different σ

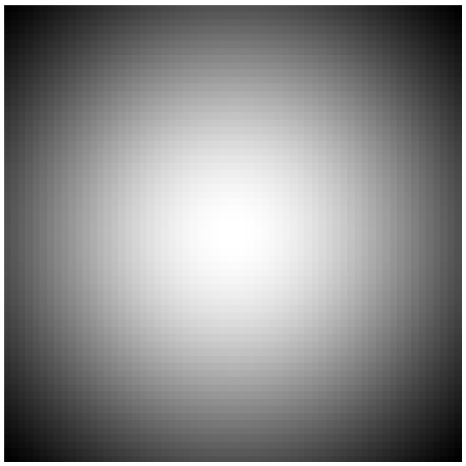
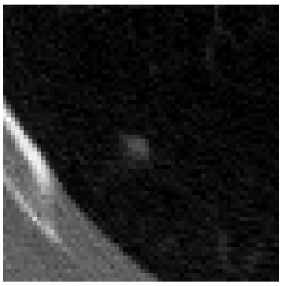


Figure 12: An example Gaussian mask with $\sigma = .41$

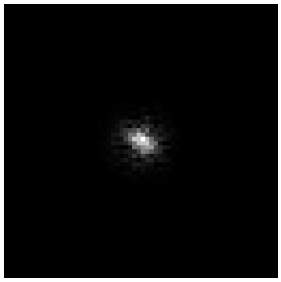
values as described in Section 3 and the beginning of this section). Resulting was a list of log ratios for each estimation method, for each type of nodule, for all $M = 1000$ random samples.

Using Minitab and MATLAB, two nonparametric tests were performed, specifically Kruskal-Wallis and Wilcoxon Rank-Sum. Nonparametric tests were used because data from the log ratio failed normality tests. Of the isolated nodules, the log ratios of $\sigma = .31, .36, .41, .46$ did not fail the normality test whereas the log ratios of all treatments failed the normality test of juxtapleural nodules. The lower bound, ground truth, upper bound, and $\sigma = .01$ were the only treatments to not fail the normality test for the juxtavascular nodules. Figure 16 shows that the ground truth log ratios of juxtapleural nodules fails the normality test, and Figure 17 shows that the ground truth log ratios, denoted by "true," of juxtavascular nodules does not fail the normality test. Since not all of the treatments can be considered normal, we use nonparametric tests.

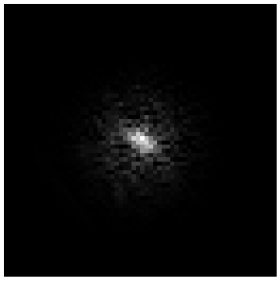
Kurskal-Wallis is referred to as a "distribution-free ANOVA"[2]. Wilcoxon Rank-Sum and Kruskal-Wallis tests assume the data come from similarly shaped



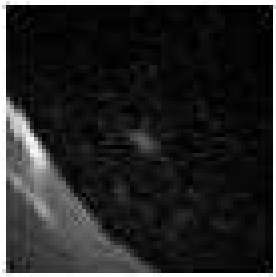
(a) No mask



(b) $\sigma = .06$

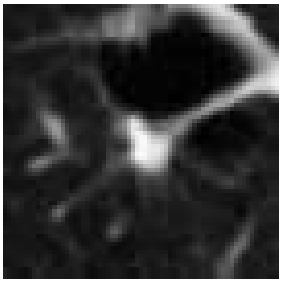


(c) $\sigma = .11$

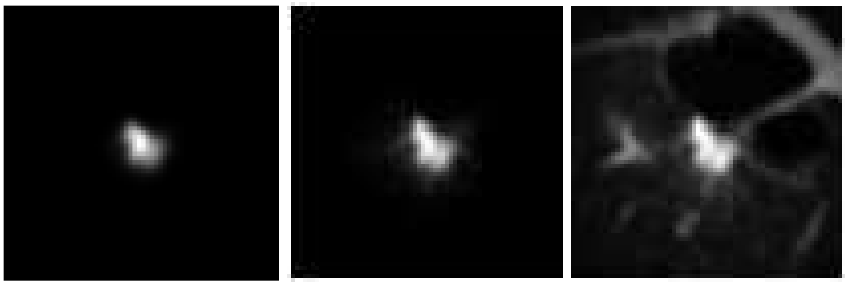


(d) $\sigma = .41$

Figure 13: An isolated nodule with a Gaussian mask at varying σ values



(a) No mask



(b) $\sigma = .06$ (c) $\sigma = .11$ (d) $\sigma = .41$

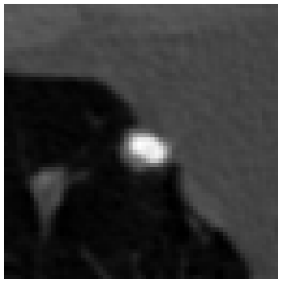
Figure 14: A juxtavascular nodule with a Gaussian mask at varying σ values

distributions. Figure 18, Figure 19, and Figure 20 show that volume estimation methods, or treatments, come from similarly shaped distributions within each type of nodule.

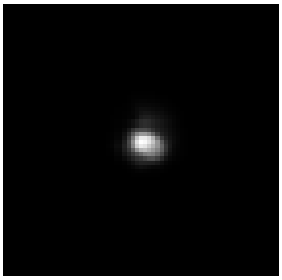
4.1.1 Kruskal-Wallis

Let I be the number of treatments, $i = 1, 2, 3, ..., I$ and J_i be the number of samples in treatment i with an unknown distribution. Then the total number of observations $N = \sum J_i$. Let X_{ij} be the jth observation from the ith treatment where $j = 1, 2, ..., J_i$.

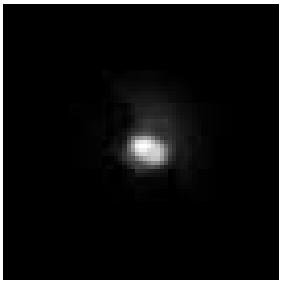
Suppose we rank all observations from all treatments in ascending order. We aim to test the likelihood of rank assignments being equal, given a random sample. Define R_{ij} as the rank of X_{ij}. Then let $\bar{R}_{i\cdot}$ be the average rank in the



(a) No mask



(b) $\sigma = .06$



(c) $\sigma = .11$

(d) $\sigma = .41$

Figure 15: A juxtapleural nodule with a Gaussian mask at varying σ values

ith treatment (when the null hypothesis is not rejected). The test statistic H is computed as follows:

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{I} J_i(\bar{R_{i\cdot}} - \frac{N+1}{2})^2. \tag{11}$$

The null and alternative hypotheses are

H_0: $\bar{R}_{1\cdot} = \bar{R}_{2\cdot} = \bar{R}_{3\cdot} = ... = \bar{R}_{I\cdot}$

H_a: at least one $\bar{R}_{i\cdot}$ is different.

4.1.2 Wilcoxon Rank-Sum

In a Wilcoxon Rank-Sum test, we use two independent random samples, Y and Z, with means μ_1 and μ_2, respectively, and the same shape and spread. We have W = sum of the Y observation ranks in the combined sample where Y has

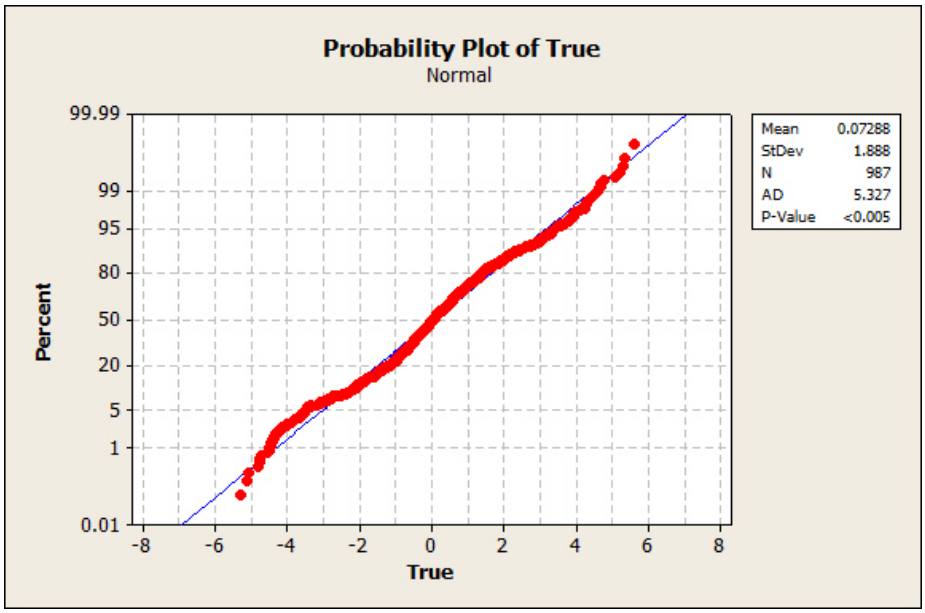


Figure 16: A probability plot shows a data set of log ratios from juxtapleural nodules which cannot be assumed to be normal

at most the same number of elements as Z. The null and alternative hypotheses as described in [2] are

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_a: \mu_1 - \mu_2 \neq 0.$$

This test will be useful for testing each method with the ground truth. Minitab does not offer the Wilcoxon Rank-Sum test, so MATLAB will be used for this portion of the analysis. In the following statistical analysis, we use a level of significance $\alpha = 0.05$.

4.2 Statistical Results

In the following sections, we show and analyze the statistical test results for each type of nodule. We have isolated nodules, juxtapleural nodules, and juxtavascular nodules in that order.

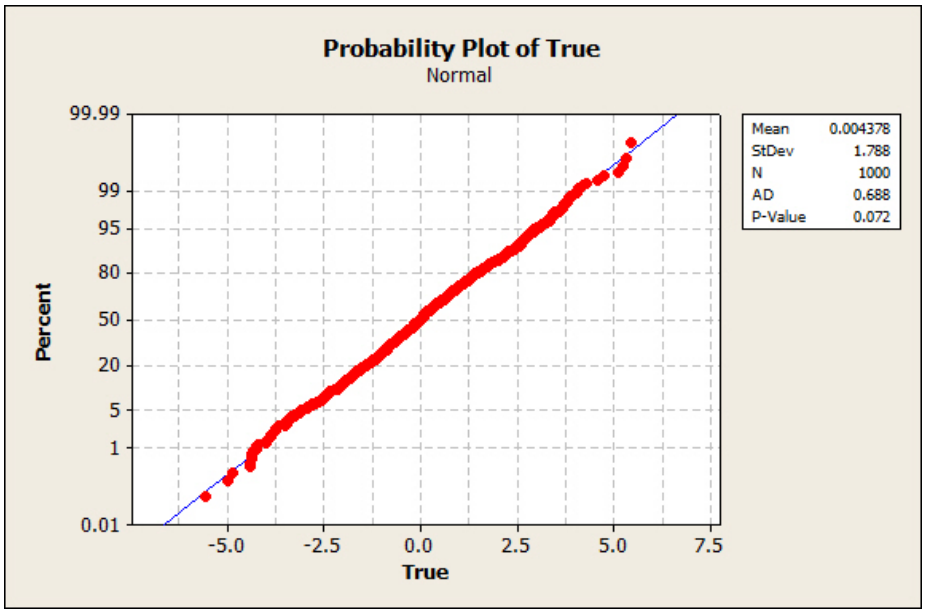


Figure 17: A probability plot shows a data set of log ratios from juxtavascular nodules which can be assumed to be normal

4.2.1 Isolated

Figure 21 shows a scatter plot of the ground truth log ratios with the various σ values. By looking at the leftmost column, we can compare the various σ values to the ground truth and find the best linear correlation by using the R^2 values. One can see that $\sigma = .06$ has the most linearly correlated results with "true." This might suggest that $\sigma = .06$ is the most useful value, but we need to test this further to draw any significant conclusions. The R^2 values in Table 1 show $\sigma = .06$ has the largest R^2 value and is therefore most linearly correlated with the ground truth. This σ value will be tested additionally with the ground truth by a Wilcoxon Rank-Sum test. We will use the Kruskal-Wallis test to see if there are any other σ values whose log ratios yield significant results.

Table 2 shows the results of the Kruskal-Wallis test as performed by Minitab

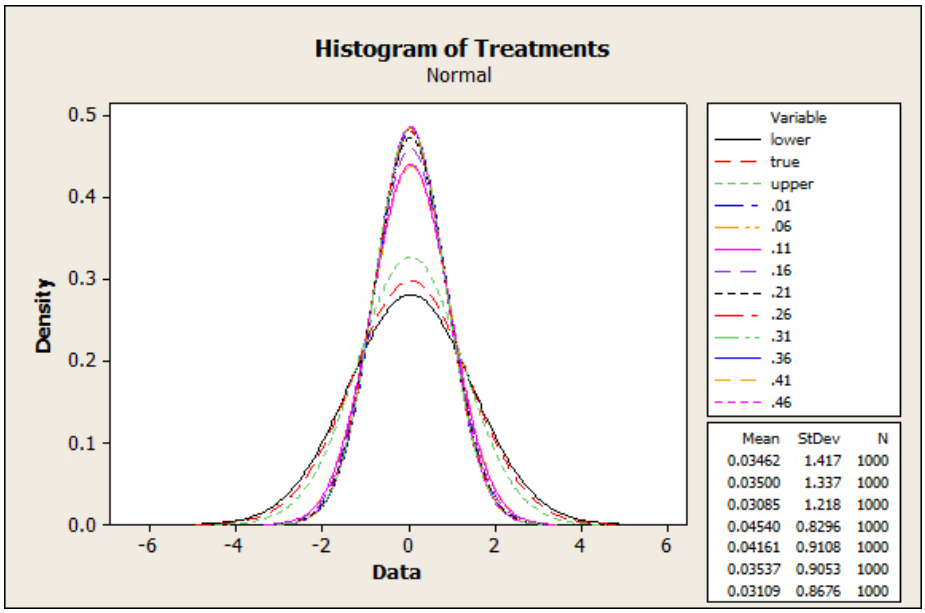


Figure 18: Histogram of log ratios of isolated nodules in all treatments

16, where the different treatments are the ten different σ values, the upper and lower bounds, and the ground truth. The overall average median is $\bar{R}_{ij} = 6500.5$, $H = 1.96$ and $P = 0.999$ with 12 degrees of freedom. This p-value is much greater than α, and in fact is highly significant, so we fail to reject the null hypothesis and conclude that none of the treatments differ significantly.

Notice the median for treatment $\sigma = .11$, $\sigma = .26$, lower bound, ground truth and upper bound are all equal to zero. This could mean they are favorable to the other treatments, however, this is an expected result from our data, so this may be an insignificant coincidence. Additionally, the $\bar{R}_{i\cdot}$'s for the different σ values are all very similar to one another. The same is true among the lower bound, ground truth and upper bound. The R^2 value for $\sigma = .26$ is relatively low, so it is not the best value or σ choice.

We now compare the lower and upper bounds to the ground truth. The Kruskal-Wallis test results are shown in Table 3. The overall average median

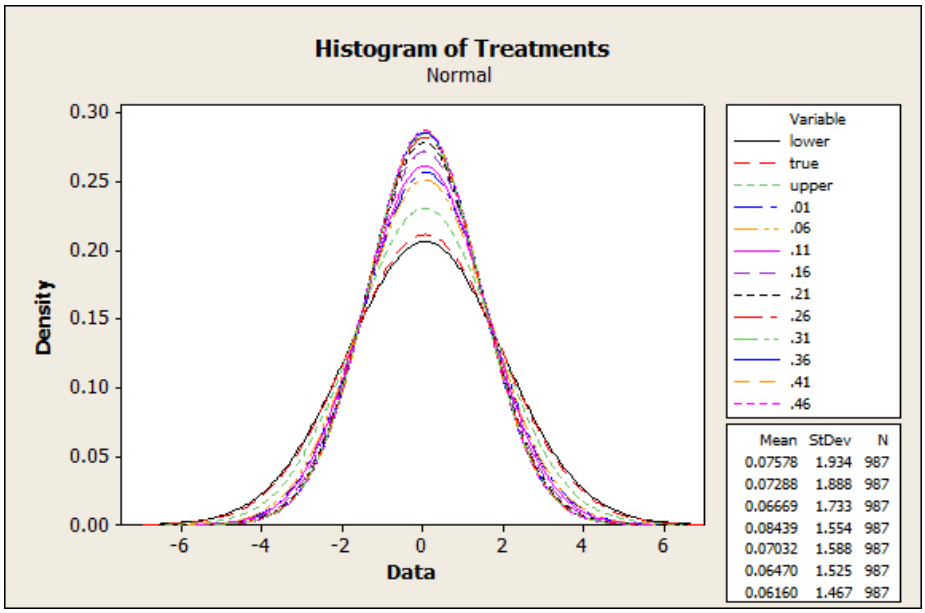


Figure 19: Histogram of log ratios of juxtapleural nodules in all treatments

rank for the lower bound, ground truth and upper bound results in $\bar{R}_{ij} = 1500.5$. The test statistic $H = 0.01$ and the p-value is 0.993 with two degrees of freedom. Since the p-value is greater than α, we can conclude that the mean ranks of the lower bound, ground truth and upper bound do not differ significantly from each other.

Since all of the other tests resulted in not rejecting the null hypothesis, to get a more direct result, we perform the Kruskal-Wallis test ten times to compare the ground truth results to each σ value result. This yielded the results shown in Table 4 where H is the test statistic. Each of the resulting p-values is greater than α, so we again do not reject the null hypothesis and conclude that the mean ranks of the ten σ results do not differ significantly from that of the ground truth result. Notice how the p-value is largest for $\sigma = .11$ and $\sigma = .16$. These values, .621 and .632 respectively, are highly significant. However, many of the other σ values are relatively close, so we cannot conclude anything firm.

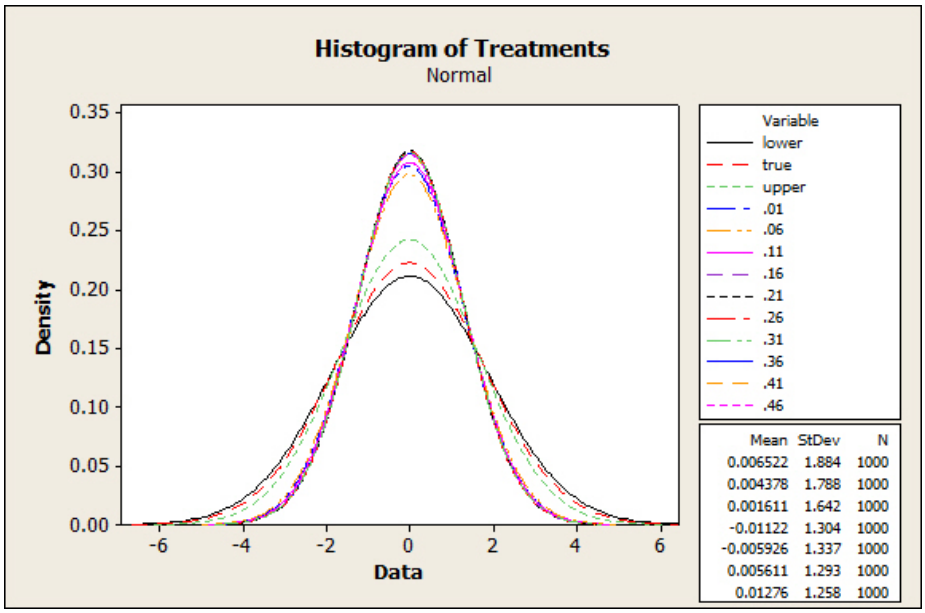


Figure 20: Histogram of log ratios of juxtavascular nodules in all treatments

We investigate some of the treatments that have had significant results so far. There are large p-values for $\sigma = .11$ and $\sigma = .16$, and the scatterplot of $\sigma = .06$ versus the ground truth is the most linearly correlated. We use the two-sample Wilcoxon Rank-Sum test to compare σ values and the lower and upper bounds with the ground truth with results shown in Table 5. We still need to see if there is a significant difference between the bounds and ground truth. All of the p-values are larger than α. Of the σ values, $\sigma = .16$ has the largest p-value, and the upper bound has the larger p-value of the two bounds. However, the R^2 value for $\sigma = .16$ is 45.7 which is not as large as the R^2 value for the other two σ values.

We use a boxplot to further compare the isolated nodule data. In Figure 22, we can see a direct comparison of log ratios resulting from the different methods. The numerical labels (e.g. "0.1") represent the σ values used in Gaussian weighted integration described in Section 3. The ground truth volume esti-

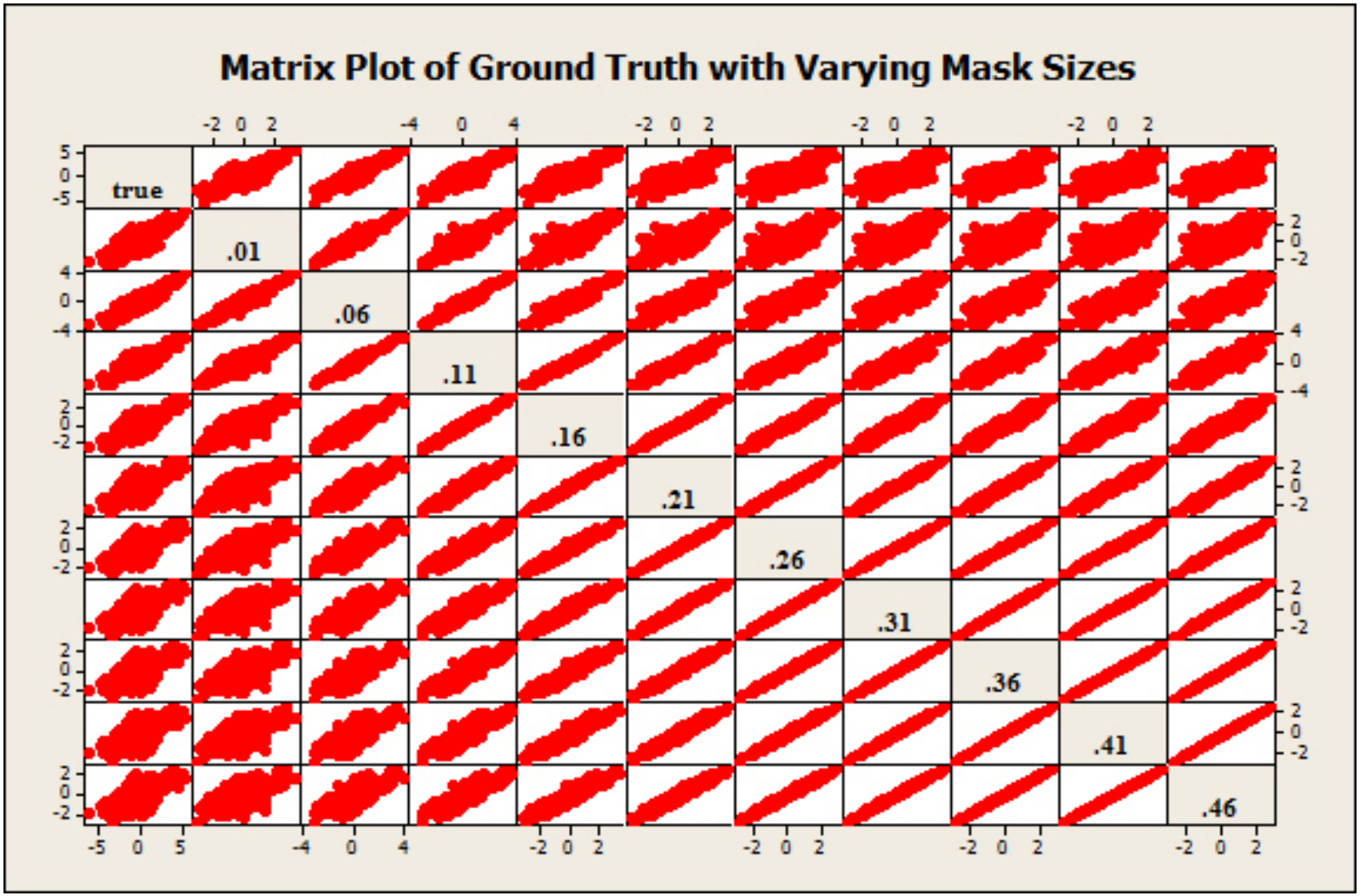


Figure 21: Matrix of scatter plots comparing the log ratio of true volumes to the different sigma values of isolated nodules of isolated nodules

σ	R^2
.01	.502
.06	.686
.11	.576
.16	.457
.21	.373
.26	.322
.31	.293
.36	.275
.41	.264
.46	.257

Table 1: R^2 values for scatterplots of varying σ values with ground truth for isolated nodules

mation is represented by "true." The remaining labels "lower" and "upper" represent the lower and upper bounds. This notation is used throughout the statistical analysis.

The spread of the data for the different σ values is similar. The lower bound, ground truth, and upper bound have a larger, but similar, spread. They also have outliers further away from the median. The location, or median, of the data is very close, as we saw in the Kruskal-Wallis test results. However, the outliers may tell us something about the accuracy of the semi-automated method. The ground truth tells us that there should be many outliers in both the positive and negative directions. The σ log ratios seem to result in fewer outliers that are closer to the median. It is possible the semi-automated method does not give us results as good as or better than the ground truth.

Treatment i	J_i	**Median**	$\bar{R}_{i\cdot}$
.01	1000	0.0219375	6577.7
.06	1000	0.0031910	6536.9
.11	1000	0.0000000	6509.4
.16	1000	0.0026300	6510.2
.21	1000	0.0035965	6515.9
.26	1000	0.0000000	6516.9
.31	1000	0.0045655	6518.2
.36	1000	0.0147765	6518.9
.41	1000	0.0137200	6517.7
.46	1000	0.0090390	6517.5
lower	1000	0.0000000	6414.0
true	1000	0.0000000	6430.2
upper	1000	0.0000000	6422.9

Table 2: Kruskal-Wallis test on all of the treatments for isolated nodules

Treatment i	J_i	**Median**	$\bar{R}_{i\cdot}$
lower	1000	0.000000000	1497.9
true	1000	0.000000000	1501.5
upper	1000	0.000000000	1502.1

Table 3: Kruskal-Wallis test on the lower bound, upper bound, and ground truth for isolated nodules

σ	**p-value**	**H**
.01	.399	.71
.06	.528	.40
.11	.621	.24
.16	.632	.23
.21	.618	.25
.26	.612	.26
.31	.607	.26
.36	.602	.27
.41	.603	.27
.46	.608	.26

Table 4: Individual Kruskal-Wallis test results for ground truth tested with varying σ values of isolated nodules

Treatment	**p-value**
.06	.5279
.11	.6211
.16	.6319
lower	.9277
upper	.9901

Table 5: Wilcoxon Rank-Sum test results for isolated nodules with the ground truth

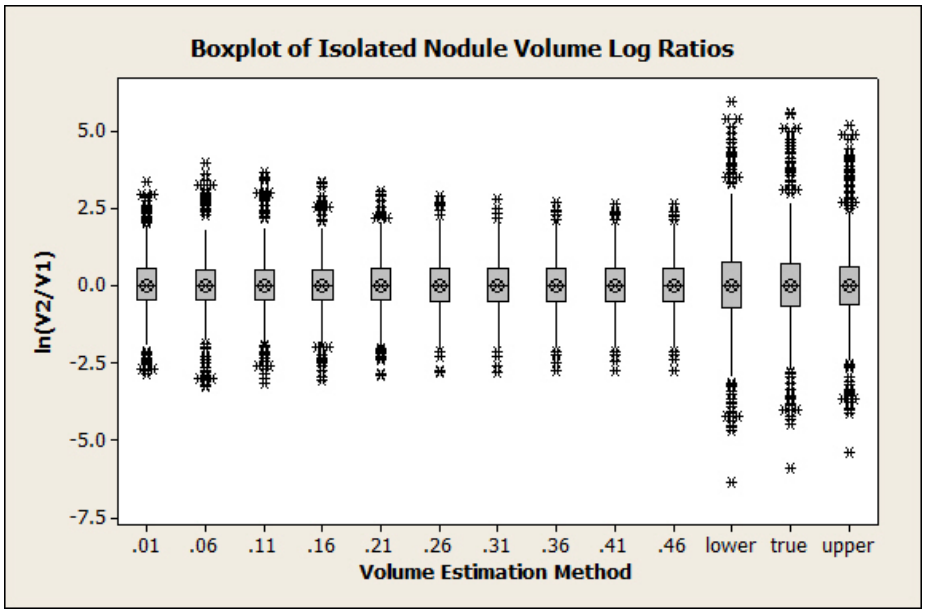


Figure 22: A boxplot comparing the 13 treatments for isolated nodules

4.2.2 Juxtapleural

We now examine the juxtapleural nodules. We perform the same statistical tests as we did for the isolated nodules. We have the scatterplots in Figure 23 and corresponding R^2 values in Table 6. The R^2 values show that the log ratios of $\sigma = .06$ and $\sigma = .11$ are the most linearly correlated with the ground truth. They have R^2 values of .753 and .730, respectively. In fact, all of the σ values had similarly large R^2 values. We will test these two σ values with the ground truth using Wilcoxon-Rank Sum since they had the highest R^2 values..

We have the Kruskal-Wallis test results shown in Table 7. The medians are all similar, especially those of the σ values. The overall average median is $\bar{R}_{ij} = 6500.5$, $H = 0.67$ and $P = 1.000$ with 12 degrees of freedom. This p-value is highly significant, so we can say that there is sufficient evidence that the average medians for each treatment are equal. Unlike the isolated nodules,

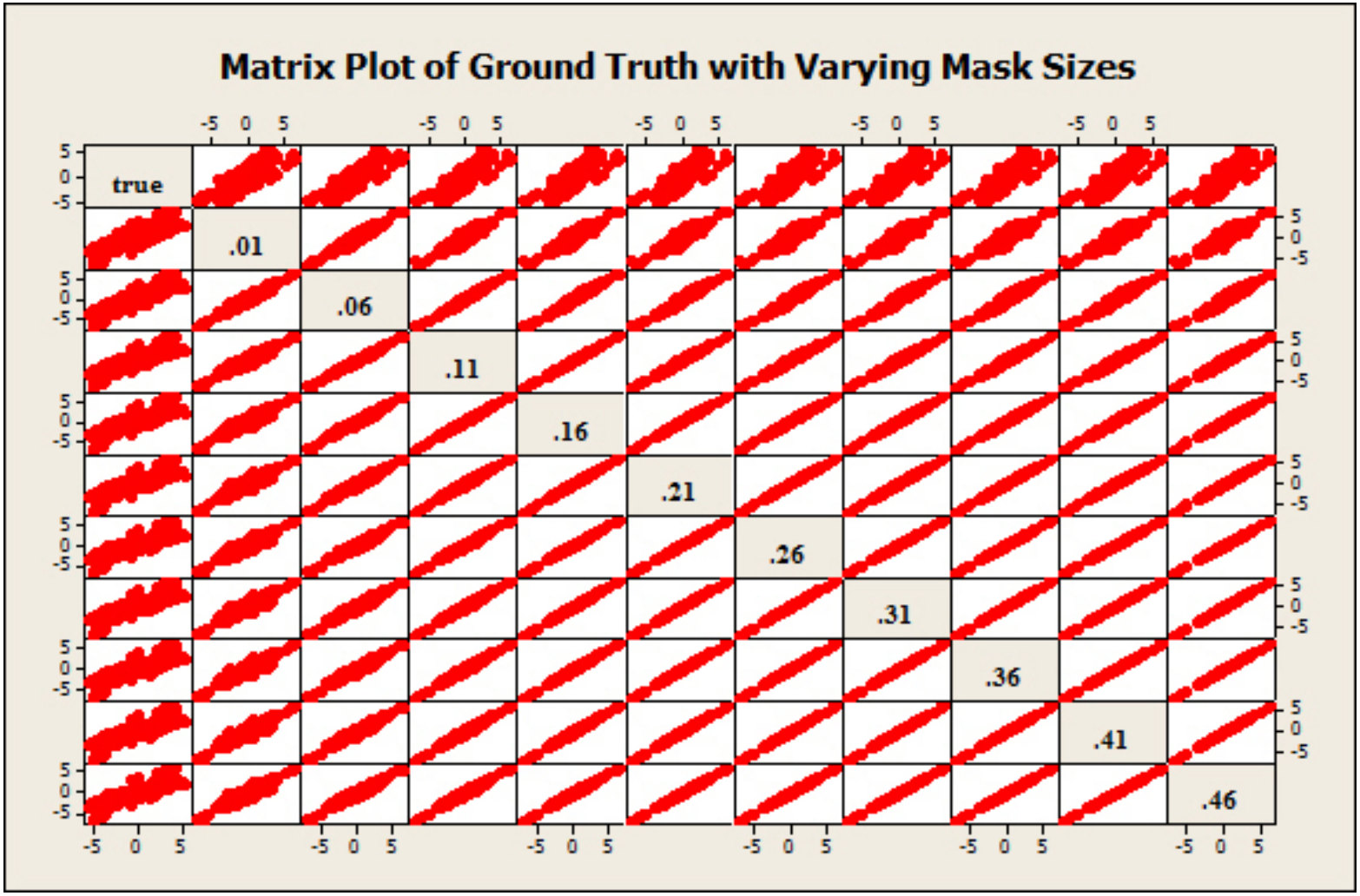


Figure 23: Matrix of scatter plots comparing the log ratio of true volumes to the different sigma values of juxtapleural nodules

σ	R^2
.01	.663
.06	.753
.11	.730
.16	.721
.21	.718
.26	.717
.31	.716
.36	.716
.41	.716
.46	.716

Table 6: R^2 values for scatterplots of varying σ values with ground truth of juxtapleural nodules

none of the medians are zero.

Next, we test the lower bound, ground truth, and upper bound. The results are shown in Table 8. The overall average median is $\bar{R}_{ij} = 1481.0$, $H = 0.02$ and $P = 0.989$. This p-value is highly significant, so we fail to reject the null hypothesis. Thus we can say that there is sufficient evidence that the average ranks of treatments $\bar{R}_{i\cdot}$, $i = 1, 2, ..., 13$, are equal.

We perform the Kruskal-Wallis test on each σ value with the ground truth. The results are shown in Table 9. The p-values for each σ value with the ground truth are large. The table shows that all the p-values are similar, but $\sigma = .01$, $\sigma = .06$, and $\sigma = .11$ have the largest p-values.

Table 10 shows the results of the Wilcoxon Rank-Sum tests on the log ratios of $\sigma = .01$, $\sigma = .06$, and $\sigma = .11$ as well as the lower and upper bounds with the ground truth. All of the tests resulted in large p-values, with $\sigma = .01$ having the

Treatment i	J_i	**Median**	$\bar{R}_{i\cdot}$
.01	987	0.060488	6479.5
.06	987	0.009169	6419.1
.11	987	0.025999	6409.1
.16	987	0.028359	6403.4
.21	987	0.027396	6399.4
.26	987	0.027480	6396.6
.31	987	0.025783	6394.8
.36	987	0.028233	6393.9
.41	987	0.023677	6392.9
.46	987	0.019381	6392.0
lower	987	0.040483	6452.8
true	987	0.041405	6446.4
upper	987	0.021783	6428.2

Table 7: Kruskal-Wallis test results for all treatments of juxtapleural nodules

Treatment i	J_i	**Median**	$\bar{R}_{i\cdot}$
lower	987	0.04048	1483.3
true	987	0.04140	1481.8
upper	987	0.02178	1477.9

Table 8: Kruskal-Wallis test of lower bound, upper bound, and ground truth for juxtapleural nodules

σ	**p-value**	**H**
.01	.900	.02
.06	.881	.02
.11	.824	.05
.16	.805	.06
.21	.776	.08
.26	.764	.09
.31	.760	.09
.36	.761	.09
.41	.757	.10
.46	.758	.09

Table 9: Individual Kruskal-Wallis test results for ground truth tested with varying σ values of juxtapleural nodules

largest of the three σ values. Though the lower bound had the largest p-value, both the lower and upper bound had large p-values.

Figure 24 shows boxplots for all the methods used on juxtapleural nodules. The locations of the various treatments are similar. The spread of the ground truth is similar to that of smaller σ values and the lower and upper bounds. The location of outliers might tell us that the semi-automated method may not give results as good or better than the ground truth. The outliers in the semi-automated method are further from the median than in the ground truth and bounds results. Of all the semi-automated results, the plot of $\sigma = .01$ looks the most similar to the ground truth. The lower bound plot is more similar to the ground truth plot than the upper bound does.

Treatment	p-value
.01	.9017
.06	.8836
.11	.8277
lower	.9634
upper	.9150

Table 10: Wilcoxon Rank-Sum test results for juxtapleural nodules with the ground truth

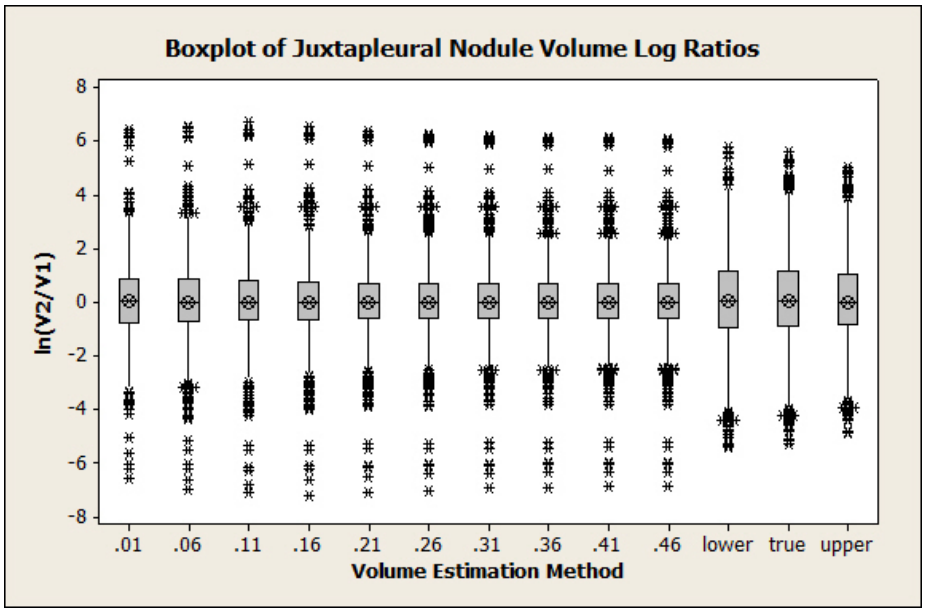


Figure 24: Boxplot comparing all the methods for juxtapleural nodules

4.2.3 Juxtavascular

Finally, we examine the juxtavascular nodules. Figure 25 is a matrix of scatter plots, and Table 11 shows the corresponding R^2 values for the different σ values with the ground truth. The results for $\sigma = .06$ and $\sigma = .11$ have the highest R^2 values at .813 and .818 respectively. We will use Wilcoxon Rank-Sum to examine these further. Note that all of the σ values result in relatively high and similar R^2 values. This could mean that for juxtavascular nodules, any of the given σ values produce log ratios similar to the ground truth.

σ	R^2
.01	.657
.06	.813
.11	.818
.16	.763
.21	.704
.26	.662
.31	.634
.36	.617
.41	.605
.46	.597

Table 11: R^2 values for scatterplots of varying σ values with ground truth of juxtavascular nodules

Table 12 shows the results of a Kruskal-Wallis test of all the treatments with the overall average median $\bar{R}_{ij} = 6500.5$, $H = 0.56$ and $P = 1.000$. This p-value is large, so we can say that there is sufficient evidence that the average ranks of treatments $\bar{R}_{i.}$, where $i = 1, 2, ..., 13$, are equal.

The results of individual Kruskal-Wallis tests of varying σ values with the

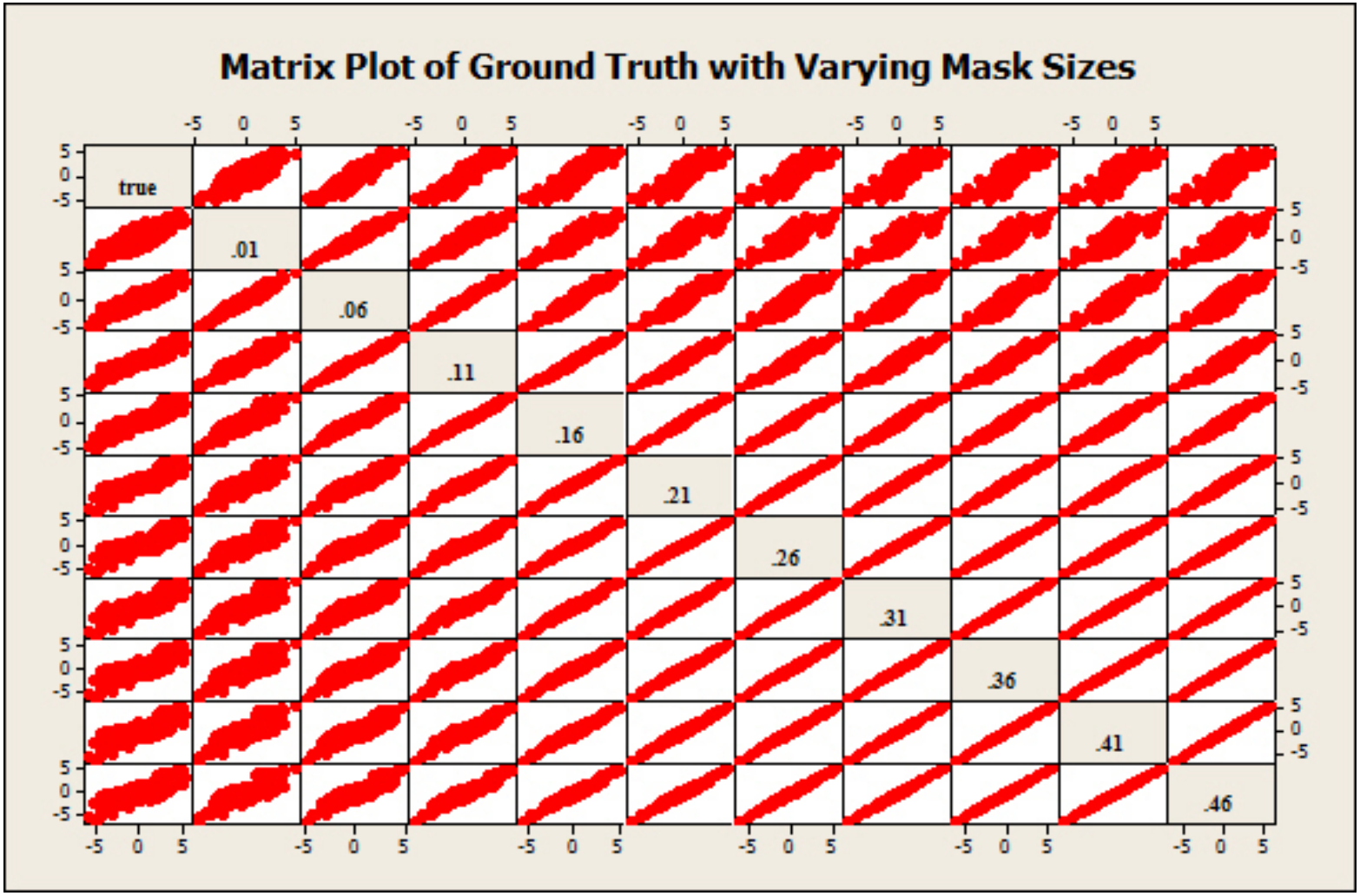


Figure 25: Matrix of scatter plots comparing the log ratio of true volumes to the different sigma values of juxtavascular nodules

Treatment i	J_i	**Median**	$\bar{R}_{i\cdot}$
.01	1000	-1.24000E-04	6449.4
.06	1000	-7.51350E-03	6464.5
.11	1000	0.000000000	6498.2
.16	1000	0.000000000	6523.3
.21	1000	0.000000000	6527.8
.26	1000	0.000000000	6524.7
.31	1000	0.002452500	6521.4
.36	1000	0.005121500	6518.9
.41	1000	0.007277000	6516.0
.46	1000	0.003710500	6514.3
lower	1000	0.000000000	6484.7
middle	1000	-2.14000E-04	6481.8
upper	1000	-3.40150E-03	6481.4

Table 12: Kruskal-Wallis test on varying treatments of juxtavascular nodules

ground truth are shown in Table 13. All ten separate tests resulted in large p-values, with $\sigma = .06$ and $\sigma = .11$ yielding the largest p-value. Table 14 shows the results of Wilcoxon Rank-Sum tests of σ values that had significant results when tested with the ground truth, as well as the tests of both the lower and upper bounds with the ground truth. All of these treatments have highly significant p-values when tested with the ground truth. Thus we can say there is sufficient evidence that each of these treatments has a mean equal to that of the ground truth.

The boxplot shown in Figure 26 shows that the varying σ values produce results similar to the ground truth. The spread is larger for the ground truth, but all treatments have similar locations and outliers. The plots for $\sigma = .06, .11, .16,$ and .21 are the most similar to the ground truth plot, but the plot for $\sigma = .01$ has the most similar spread as the ground truth. The lower and upper bound plots are similar, but the lower bound plot has a larger spread and the outliers are further from the median.

5 Conclusion/Future Work

Based on the statistical results, for the isolated nodules, $\sigma = .06$ is the best choice of σ value. It had the highest R^2 value and a large p-value in the Wilcoxon Rank-Sum test. For the juxtapleural nodules, $\sigma = .06$ is also the best choice of σ value for the same reasons as it is for isolated nodules. For juxtavascular nodules, $\sigma = .11$ barely had the highest R^2 value and a large p-value in the Wilcoxon Rank-Sum test. However, in all three of these tests, $\sigma = .06, .11$ had large R^2 values and large p-values. Therefore, to generalize for all types of nodules, $\sigma = .06$ and $\sigma = .11$ are the best choices to use in the semi-automated method. Individually, these σ values produced significant results.

The lower and upper bound results were determined to be statistically similar

σ	**p-value**	**H**
.01	.874	.03
.06	.938	.01
.11	.926	.01
.16	.803	.06
.21	.792	.07
.26	.808	.06
.31	.821	.05
.36	.833	.04
.41	.841	.04
.46	.846	.04

Table 13: Individual Kruskal-Wallis test results for ground truth tested with varying σ values of juxtavascular nodules

Treatment	**p-value**
.06	.9377
.11	.9262
lower	.9920
upper	.9931

Table 14: Wilcoxon Rank-Sum test results for juxtavascular nodules with the ground truth

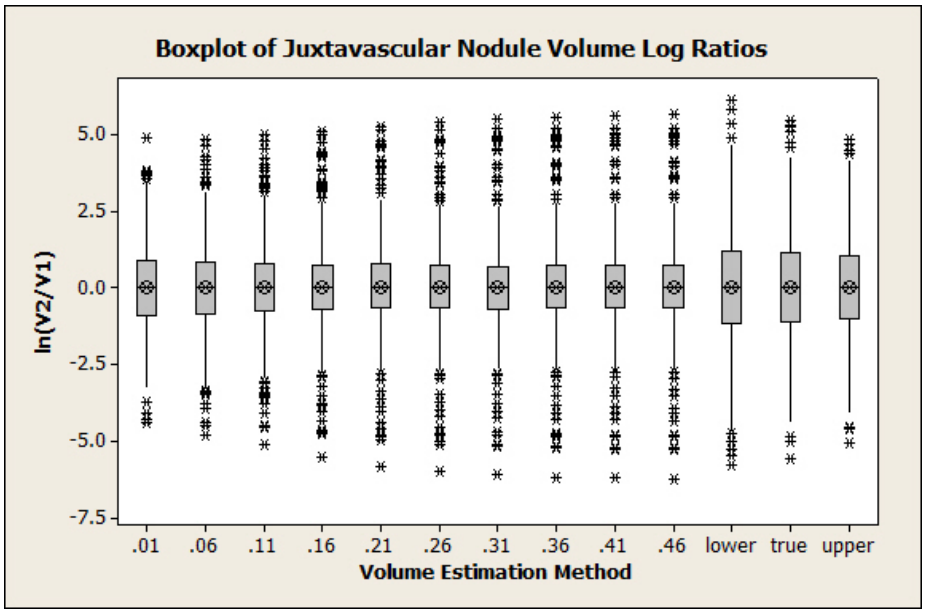


Figure 26: Boxplot comparing all the methods for juxtavascular nodules

to the ground truth by using the Kruskal-Wallis and Wilcoxon Rank-Sum tests. Further testing should be done to determine if this is a correct conclusion. Had all of the data been normally distributed, it could have been analyzed with more rigorous statistical tests such as the two sampe t-test and ANOVA. This would have provided us with more concrete statistical results. Had these tests been available for use, we may have found σ values that were not accurate enough, or the lower and upper bound may not have been considered statistically similar to the ground truth, as one might expect.

There are some other points that need to be taken into consideration when making these conclusions. The statistical tests performed resulted in large p-values, so we failed to reject the null hypothesis in all cases. Though some values yileded larger p-values than others, more rigorous tests should be performed to confirm their accuracy. To determine the accuracy of the different σ values, we have to consider how accurate the ground truth estimation actually is. Since the

CT scans limit the number of image slices and the resolution of the images, the ground truth is only as accurate as these limitations allow. The methods used in this thesis could be improved if more types of scans were available or if there were more slices per scan. In this study, we simulated two CT scans for each patient, but additional scans would allow for more accurate results. A specified algorithm for each type of nodule (isolated, juxtapleural, juxtavascular) would be useful for estimating volumes. For this, an automated or semi-automated method could be used to classify nodules as one of these types and the appropriate volume estimation algorithm would be used.

As technology advances, so will the quality of the images from these scans. With better resolution, the size of voxels changes and allows for more accurate boundary lines and in turn, volume measurements. With isotropy, a three-dimensional mask could be used in a semi-automated method to yield more accurate results. Importantly, the ground truth would be more accurate and a better comparison to use with other methods. Ultimately, a semi-automated method would provide a better option for radiologists to use for measuring the size of pulmonary nodules.

References

[1] S. G. III et al Armato. The lung image database consortium (lidc) and image database resource initiative (idri): A completed reference database of lung nodules on ct scans. *Medical Physics*, 38, February 2011.

[2] J. L. Devore. *Probability and Statistics for Engineering and the Sciences*. Duxbury, seventh edition, 2008.

[3] Sone S. Takashima S. Li F. Yang Z.-G. Maruyama Y. Hasegawa, M. and T. Watanabe. Growth rate of small lung cancers detected on mass ct

screening. *Br. J. Radiology*, 73:1252–1259, 2000.

[4] National Cancer Institute. Lung image database consortium (lidc), 1996. "https://wiki.cancerimagingarchive.net/display/Public/TCGA-GBM".

[5] Reeves A. P. Yankelevitz D. F. Kostis, W. J. and C. I. Henschke. Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical ct images. *IEEE Trans. Med. Imag.*, 22(10):1259–1274, 2003.

[6] G. A. Lillington. Management of solitary pulmonary nodules. *Postgraduate Medicine*, 101(3), March 1997.

[7] F. Preparata and M. Shamos. *Computational Geometry: An Introduction.* Springer-Verlag, 1985.

[8] A. Reeves, A. Chan, D. Yankelevitz, C. Henshke, B. Kressley, and W. Kostis. On measuring the change in size of pulmonary nodules. *IEEE Transactions on Medical Imaging*, 25:1–2, April 2006.

[9] Chan A. B. Yankelevitz D. F. Henschke-C. I. Kressler B. Reeves, A. P. and W. J. Kostis. On measuring the change in size of pulmonary nodules. *IEEE Trans. Med. Imag.*, 25(4):435–450, April 2006.

[10] Bissery A. Bienvenu M. Aycard L. Lefort-C. Revel, M.-P. and G. Frija. Are two-dimensional ct measurements of small noncalcified pulmonary nodules reliable? *Radiology*, pages 231:453–458, 2004.

[11] Lefort C. Bissery A. Bienvenu M. Aycard-L. Chatellier G. Revel, M.-P. and G. Frija. Pulmonary nodules: Preliminary experience with three-dimensional evaluation. *Radiology*, pages 231:459–466, 2004.

[12] Merlin A. Peyrard S. Triki R. Couchon-S. Chatellier G. Revel, M.-P. and G. Frija. Software volumetric evaluation of doubling times for differentiat-

ing benign versus malignant pulmonary nodules. *Am. J. Roentgen.*, pages 187:135–142, 2006.

[13] Saito Y. Sagawa M. Sato M. Kanma K.-Takahashi S. Endo C. Chan Y. Sakurada A. Usuda, K. and S. Fujimura. Tumor doubling time and prognostic assessment of patients with primary lung cancer. *Cancer*, pages 2239–2244, October 2005.