




2018

USING THE QBEST EQUATION TO EVALUATE ELLAGIC ACID SAFETY DATA: GENERATING A QNOAEL WITH CONFIDENCE LEVELS FROM DISPARATE LITERATURE

Cynthia Rose Dickerson

University of Kentucky, crdi223@uky.edu

Author ORCID Identifier:

 <https://orcid.org/0000-0002-3572-9374>

Digital Object Identifier: <https://doi.org/10.13023/etd.2018.394>

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

Recommended Citation

Dickerson, Cynthia Rose, "USING THE QBEST EQUATION TO EVALUATE ELLAGIC ACID SAFETY DATA: GENERATING A QNOAEL WITH CONFIDENCE LEVELS FROM DISPARATE LITERATURE" (2018). *Theses and Dissertations--Pharmacy*. 94.

https://uknowledge.uky.edu/pharmacy_etds/94

This Doctoral Dissertation is brought to you for free and open access by the College of Pharmacy at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Pharmacy by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Cynthia Rose Dickerson, Student

Dr. Robert A. Lodder, Major Professor

Dr. David Feola, Director of Graduate Studies

USING THE QBEST EQUATION TO EVALUATE ELLAGIC ACID SAFETY DATA:
GENERATING A QNOAEL WITH CONFIDENCE LEVELS FROM DISPARATE
LITERATURE

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the College of Pharmacy at the University of Kentucky

By Cynthia Rose Dickerson
Lexington, Kentucky

Director: Dr. Robert A. Lodder, Professor of Pharmacy, Chemical Engineering
Lexington, Kentucky

2018

Copyright © Cynthia Dickerson 2018

ABSTRACT OF DISSERTATION

USING THE QBEST EQUATION TO EVALUATE ELLAGIC ACID SAFETY DATA: GENERATING A QNOAEL WITH CONFIDENCE LEVELS FROM DISPARATE LITERATURE

QBEST, a novel statistical method, can be applied to the problem of estimating the No Observed Adverse Effect Level (NOAEL or QNOAEL) of a New Molecular Entity (NME) in order to anticipate a safe starting dose for beginning clinical trials. The NOAEL from QBEST (called the QNOAEL) can be calculated using multiple disparate studies in the literature and/or from the lab. The QNOAEL is similar in some ways to the Benchmark Dose Method (BMD) used widely in toxicological research, but is superior to the BMD in some ways. The QNOAEL simulation generates an intuitive curve that is comparable to the dose-response curve. The NOAEL of ellagic acid (EA) is calculated for clinical trials as a component therapeutic agent (in BSN476) for treating Chikungunya infections. Results are used in a simulation based on nonparametric cluster analysis methods to calculate confidence levels on the difference between the Effect and the No Effect studies. In order to evaluate the statistical power of the algorithm, simulated data clusters with known parameters are fed into the algorithm in a separate study, testing the algorithm's accuracy and precision "Around the Compass Rose" at known coordinates along the circumference of a multidimensional data cluster. The specific aims of the proposed study are to evaluate the accuracy and precision of the QBEST Simulation and QNOAEL compared to the Benchmark Dose Method, and to calculate the QNOAEL of EA for BSN476 Drug Development.

KEYWORDS: Non-parametric Multivariate Statistical Analysis, Mahalanobis Equation, No Observed Adverse Effect Limit, Discriminant Cluster Analysis

Cynthia Dickerson
6/27/2018

USING THE QBEST EQUATION TO EVALUATE ELLAGIC ACID SAFETY DATA:
GENERATING A QNOAEL WITH CONFIDENCE LEVELS FROM DISPARATE LITERATURE

By Cynthia Rose Dickerson

Robert A. Lodder
Director of Dissertation

David Feola
Director of Graduate Studies

08/21/2018

DEDICATION

To Legal Aid of the Bluegrass, the Amanda Center for Domestic Violence Prevention, Sharon Martin and Ross Lovely. I could not have finished this without your support. Your work makes the world a better place every day. I hope this work does the same. I hope it paves the way for calculations that will enhance technology in the modern world, allowing for leaps and bounds in space science, robotics, and medical analysis. But most of all, I hope you get the satisfaction of knowing that you enabled it. I hope I can repay my debt to you by making the world a better place.

ACKNOWLEDGEMENTS

I am grateful to the PhRMA Foundation, who provided funding for this project.

Furthermore, I would like to thank Dr. Robert Lodder, who has been the best PI that a student could have asked for. His support and understanding made this work possible. I aspire to his level of professionalism and expertise.

My lab mate, Ville Tiitto, has been invaluable to me these last three years. His commentary, help, and support have made this research immeasurably stronger.

My previous lab mates have also contributed significantly to my research. Alexander Williams has pushed me to push the envelope with new statistical ideas. His commentary on how the QBEST equation might be improved is included in the conclusion of this work. Mark Ensor's work laid the groundwork for mine. Amy Banfield was an amazing resource when I first joined the lab. Her direction helped me to learn the ropes and to begin organizing my research documents from the beginning.

I appreciate the discussions that Joseph Vanghelof had with me; he was a great help in articulating statistical concepts and in comparing the pros and cons of various methods.

I am grateful to my dissertation review committee members for their time and consideration as well as all of their commentary, which has made this work considerably stronger: Dr. Robert Lodder (Committee Chair), Ms. Mary Davis Esq., Dr. David Feola (Director of Graduate Studies), Dr. Joseph Fink, Dr. Kristen Platt, and Dr. Steven Van Lanen.

I am grateful to Jurgen Rohr for helping with my enrollment at the University of Kentucky. Thank you to Matt McErlean, Ashley Arlinghaus, Steven Van Lanen, and all of the Van Lanen/Rohr lab for making me feel welcomed at the University of Kentucky and to Catina Rossoll who has helped with every part of the graduate school process.

Finally, I would like to thank all of my family for their continuous support.

TABLE OF CONTENTS

Acknowledgements.....	iii
List of Tables.....	x
List of Figures.....	xi
CHAPTER 1: INTRODUCTION.....	1
What is the Quantile Bootstrap Error Statistical Test?.....	1
What is Multivariate Variation, in Layman’s Terms?.....	1
Pharmaceutical Applications of Multivariate Data Analysis.....	4
Why a Statistical Method is Needed Beyond Simple Standard Deviation.....	5
What is the Mahalanobis distance Equation?.....	6
Why is the Mahalanobis distance Insufficient?.....	12
The Mahalanobis equation is Inaccurate When Handling Asymmetrical and Irregular Data Clusters.....	12
The Mahalanobis equation is Mathematically Incapable of Calculating Distances For Datasets with More Variables Than Observations.....	14
What This Means in a Pharmaceutical Context.....	14
The Mahalanobis Requires Too Much Computing Power.....	14
What is the Quantile Bootstrap Error Statistical Test in Detail?.....	15
Why is QBEST Superior to the Mahalanobis equation?.....	15
Observations vs. Variables.....	16
What This Means for Pharmaceutical Studies.....	16
Computational Cost.....	16
What is the QBEST?.....	17
“A Rubber Ruler with a Nail in the Center”.....	17
How Does QBEST Work?.....	17
What is Discriminant Cluster Analysis?.....	18
How Can Discriminant Cluster Analysis be Applied to Drug Development?.....	18
Detecting Contamination.....	18
No Observed Adverse Effect Limits.....	18
What is new about QBEST?.....	19
Nonparametric Method.....	19
Unique Handling of Skewed Data.....	19
Discriminant Cluster Analysis Applied Non-Traditionally.....	20
How do you use QBEST?.....	20
The Variables.....	20
The Training Set.....	20

Radfrac.....	21
When is it appropriate to use a large/small radfrac?.....	21
Number of Bootstrap Replicates.....	21
Number of Variables or Dimensions.....	21
When Is QBEST Insufficient?.....	21
CHAPTER 2: A NOVEL STATISTICAL APPROACH TO NOAEL: QBEST APPLIED TO DOSING OF ELLAGIC ACID AND THE QNOAEL VS. BMD FOR POINT OF DEPARTURE.....	23
Publication and Copyright Information.....	23
The QNOAEL vs. BMD for Point of Departure.....	24
Abstract.....	24
Specific Aims.....	24
Evaluate the accuracy and precision of the QB Simulation and QNOAEL compared to the Benchmark Dose Method.....	25
Calculate the QNOAEL of Ellagic Acid for BSN476 Drug Development.....	25
Strategy.....	25
Significance.....	25
Method Significance.....	25
Application Significance.....	32
Innovation.....	34
Approach.....	35
Credit.....	37
Reference List.....	37
Appendix.....	37
CHAPTER 3: ESTABLISHING EDI FOR A CLINICAL TRIAL OF A TREATMENT FOR CHIKUNGUNYA.....	38
Publication and Copyright Information.....	38
Establishing EDI for a Clinical Trial of a Treatment for Chikungunya.....	39
Abstract.....	39
1 Introduction.....	39
1.1 Compound.....	39
1.2 Chikungunya.....	39
1.3 Metabolism.....	40
1.4 Use of EDI.....	41
2 Assessment of EA Use.....	42
3 Food Consumption Survey Data.....	43
3.1 Survey Description.....	43

3.2 Methods.....	44
3.3 Food Data.....	44
3.4 Food Survey Results.....	44
4 Conclusions.....	48
Support.....	49
References.....	49

CHAPTER 4: CALCULATING THE ACCEPTABLE DAILY INTAKE OR NOAEL WITH QBEST (QNOAEL).....	50
Calculating the Acceptable Daily Intake or NOEL with QBEST (QNOEL).....	51
Introduction.....	51
Current Best Practice.....	51
Discriminant Cluster Analysis and QNOAEL.....	52
Ellagic Acid.....	53
Methods.....	53
Literature Search.....	53
Literature Search Methodology.....	53
Cochrane Protocol.....	54
Converting Enzymatic Activity Units.....	55
Application of QBEST to Determine QNOAEL.....	55
Results.....	57
Discussion.....	59
Appendix: Search Methodology & Cochrane Protocol.....	61
Literature Search Methodology.....	61
Corporate Literature Provision.....	61
Web-Based Search Terms.....	61
Sources Cited by Previous Entries.....	61
Cochrane Protocol.....	62

CHAPTER 5: VALIDATING THE ACCURACY OF QBEST ON SYNTHETIC DATA: ROUND THE COMPASS ROSE.....	63
Validating the Accuracy of QBEST on Synthetic Data: Round the Compass Rose... 64	64
Introduction.....	64
Materials.....	65
System Specifications.....	65
Methods.....	65
Standard Circular Data.....	65
Standard Elliptical Data.....	66

Increasing Dimensions.....	68
Nested Iterative Loops.....	68
Memory Usage Tests.....	70
Results.....	70
Memory Usage.....	70
Relative Error or Bias.....	71
Effect of Radfrac & Bootstrap Replication on Average Relative Error.....	72
Effect of Training Spectrum Size and Dimension on Relative Error.....	74
Effect of Compass Point on Relative Error.....	76
Relative Standard Deviation.....	78
RSD is Dependent on Training Set.....	81
Number of Training Points Does Not Significantly Affect RSD.....	83
Relative Standard Deviation as TNSPEC and Dimension Increase.....	85
Mahalanobis Comparison.....	87
Discussion.....	90
Equal Precision for Small and Large Datasets.....	90
Influence of Bootstrap Replication.....	90
Influence of Dimension.....	91
Influence of Radfrac.....	91
Relative Error.....	91
Relative Standard Deviation (RSD).....	91
Influence of Training Spectrum.....	92
CONCLUSION.....	93
Summary of Chapters.....	93
Chapter 1.....	93
Chapter 2.....	93
Chapter 3.....	94
Chapter 4.....	94
Chapter 5.....	95
Future Applications.....	96
Applications in Pharmaceutical Science.....	98
Big Data in Health Insurance.....	98
Big Data in Personalized Medicine.....	101
Big Data in Safety Analysis.....	101
Applications in Computer Science and Robotics.....	102
Weaknesses of QBEST and Further Work that Must Be Done.....	102

APPENDICES.....	105
Appendix A: Equations.....	105
The Mahalanobis equation.....	105
Appendix B: Computer Programs.....	106
QBEST Algorithm.....	106
REPLICA.....	106
SOB.....	107
Estimated Daily Intake Programs.....	108
EDI.....	108
searchfoodcodes.....	116
searchfoods.....	117
searchnzero.....	118
searchsizes.....	119
ADI Programs.....	120
Generate_ADI_translation_figure.....	120
International Units from King-Bodansky Units.....	121
Round the Compass Rose Experiment Programs.....	122
Round the Compass Rose (Simplified).....	122
Regression Codes.....	130
RSD_Regression_2.....	130
Plotting Codes.....	133
RTCR_Defined_TNSPEC_Size.....	133
Plotting Relative Standard Deviation.....	140
Run TNSPEC Version 2.....	141
Plotting_Code_Increase_Radfrac_Bootstrap.....	148
Plotting_Code_Increase_TNSPEC_Dimension.....	152
Plotting_Code_Increase_TNSPEC_Dimension.....	165
Plotting_Code_Increase_TNSPEC_Dimension.....	170
RSD_as_Bootstrap_Radfrac_Increase.....	175
RSD_as_TNSPEC_Dimension_Increase.....	178
Figures_Code.....	183
Graphing_Code_2.....	185
RTCR_Point_Bias2.....	187
REFERENCES.....	189
Chapter 1 References.....	189
Chapter 2 References.....	190
Chapter 3 References.....	191

Chapter 4 References.....	193
Sources Cited By Company Files.....	193
ToxNet Sources.....	195
UK Library Search: "Ellagic Acid NOAEL"	196
Google Scholar Search: "ellagic acid noael"	197
Citations Made By Papers.....	199
Other References (Not Studies Included in Meta-Analysis).....	209
Chapter 5 References.....	209
Conclusion References.....	209
 VITA.....	 211

LIST OF TABLES

Table 1: Advantages and Disadvantages of the NOAEL, BMD, and QNOAEL.....	26
Table 2: Studies Showing Adverse Effects After Administration of Ellagic Acid.....	58
Table 3: The Eight Compass Points of the Standard Ellipse.....	68
Table 4: Memory Usage of QBEST.....	71

LIST OF FIGURES

Fig. 1: Sample Spectrum.....	2
Fig. 2: Graphing a Sample Spectrum.....	3
Fig. 3: Graphing Increasingly Multivariate Data.....	3
Fig. 4: Clusters Spatially Separated.....	4
Fig. 5: An Irregular Cluster.....	4
Fig. 6.1: A Sample Distribution with Known Center.....	7
Fig. 6.2: Centering the Rubber Ruler; Flipping the Ruler Out into Hyperspace.....	8
Fig. 6.3: Determining Standard Deviation Unit Size Along the First Axis.....	9
Fig. 6.4: Stretching the Rubber Ruler to Fit the Second Axis.....	10
Fig. 6.5: Determining Standard Deviation Unit Size Along the Second Axis.....	11
Fig. 6.6: The Standard Deviation Unit Size is Reflective of the Spread of the Data Along Each Axis.....	12
Fig. 7: Three Asymmetrically-Distributed Data Clusters with Mahalanobis Rubber Drawn.....	13
Fig. 8: QBEST is a “Rubber Ruler with a Nail”.....	17
Fig. 9: Linear BMD Model for THC in Hemp Seed.....	29
Fig. 10: THC Doses Showing Effect or No Effect Graphed in Multivariate Space.....	30
Fig. 11: Determining 98% Confidence Levels.....	31
Fig. 12: Determining the QNOEL of THC.....	32
Fig. 13: Map of Chikungunya Outbreaks.....	33
Fig. 14: QBEST Distance to a Test Point for Various Clusters.....	34
Fig. 15: The V-model for the Systems Engineering Process	35
Fig. 16: A Pharmacokinetic Study Can be Conducted Below the EDI of EA.....	42
Fig. 17: Ellagic Acid Mean EDI.....	45
Fig. 18: Ellagic Acid 90th %-tile EDI.....	46
Fig. 19: Ellagic Acid Mean Weight-Based EDI.....	47
Fig. 20: Ellagic Acid 90th %-tile Weight-Based EDI.....	48
Fig. 21: Distribution of Ellagic Acid Doses and the Corresponding Alkaline Phosphatase Levels.....	55
Fig. 22: Translating Clusters of Data to Attain the 98% Confidence Level of Difference.....	57
Fig. 23: Standard Circle.....	66
Fig. 24: Stylized Standard Ellipse.....	67
Fig. 25: Standard Ellipse with Compass Points.....	67
Fig. 26: Average Percent Error As Bootstrap Replications and Radfrac Increase.....	72
Fig. 27: Effect of Training Spectrum Size and Dimension on Relative Error.....	75
Fig. 28: Average Relative Error As Bootstrap Replicates Increase by Compass Point...77	77
Fig. 29: Relative Standard Deviation as Bootstrap Replicates and Radfrac Increase.....	79
Fig. 30: Relative Standard Deviation as Bootstrap Replicates Increase: Effect of Training Sample and Random Bootstrap Replicates.....	82

Fig. 31: Relative Standard Deviation as Bootstrap Replicates Increase: Effect of Training Set Size and Random Bootstrap Replicates.....	84
Fig. 32: Relative Standard Deviation as Training Set Size and Dimension Increase.....	85
Fig. 33: Standard Circle Used for Mahalanobis vs. BEST Comparison.....	87
Fig. 34: Comparison of Mahalanobis and BEST to Actual Distance when Dimension is Low and Number of Samples is High.....	88
Fig. 35: Accuracy and Precision of the BEST and Mahalanobis Metrics Using an $N(0,1)$ Synthetic Data Set.....	89
Fig. 36: Run Time of Mahalanobis and QBEST as Dimension Increases.....	90
Fig. 37: An Example of a Poor Dataset for Application of QBEST.....	103
Fig. 38: An Example of Coordinate Space Redefined for Purposes of Calculating the Center of a Data Cluster.....	104

CHAPTER 1: INTRODUCTION

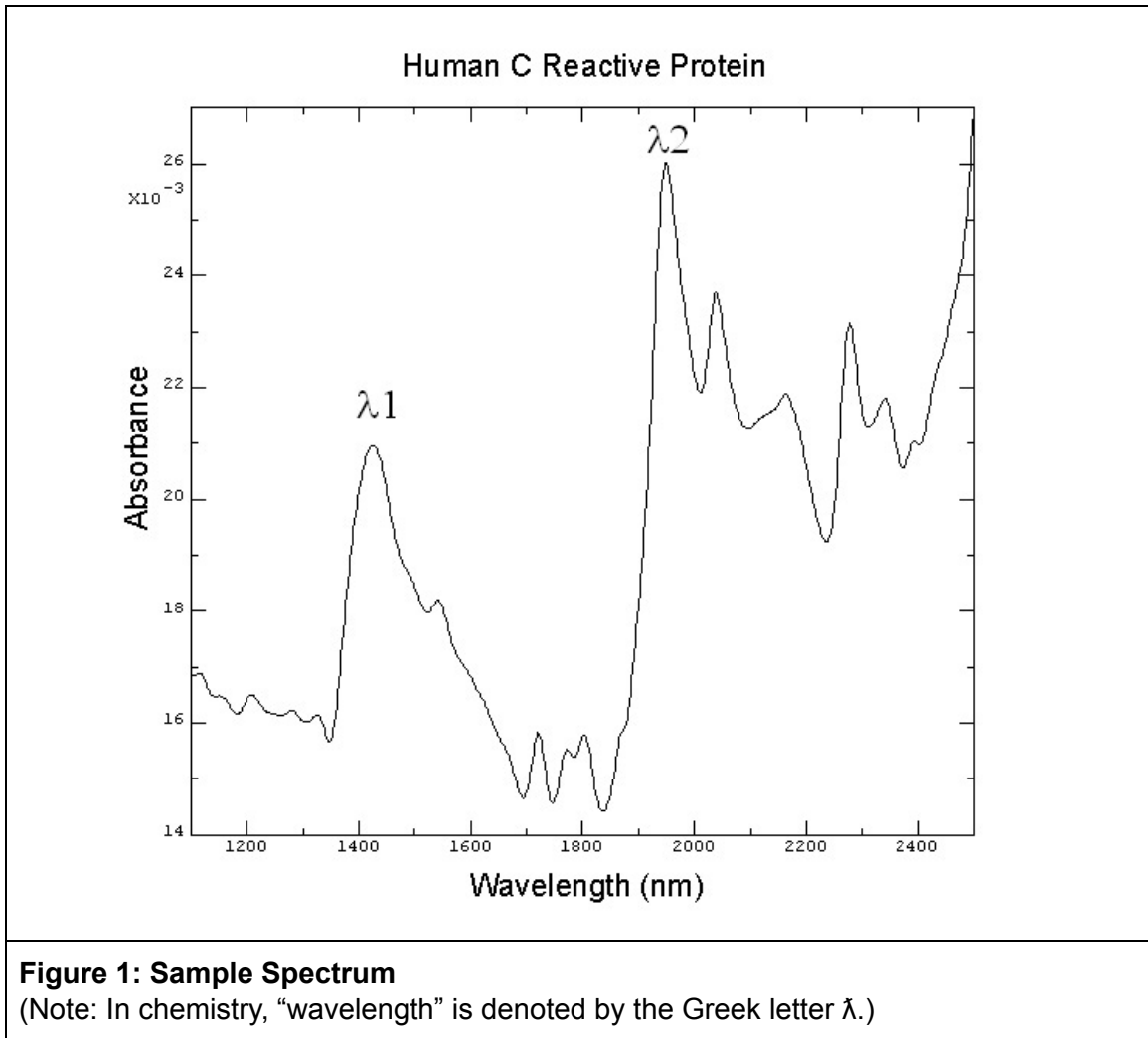
What is the Quantile Bootstrap Error Statistical Test?

The Quantile Bootstrap Error-adjusted Statistical Test (QBEST) is a novel statistical method for the nonparametric analysis of multivariate data. It is similar in many ways to the Mahalanobis test, but represents an important mathematical improvement over the Mahalanobis methodology, as the QBEST is not forced to use a symmetric standard deviation (SD) to analyze a data cluster even when the data cluster is skewed. Instead, QBEST measures a distance in asymmetric nonparametric central 68% confidence intervals (equivalent to SDs for a normal distribution) from the center point of a cluster out to the test sample point^{1, 3-9}. In other words, the QBEST metric represents an asymmetric nonparametric central 68% confidence interval for the mean. This property allows QBEST to competently handle abnormally-shaped clusters of data for which the standard deviation along opposite vectors of an axis from the center ought to vary^{3,5-9}. Furthermore, QBEST relies on simple mathematical calculations of Euclidean distance; its calculation is fairly simple, and requires far less memory than the Mahalanobis test for large matrix problems^{3,5-9}. QBEST makes calculations that are currently technologically impossible to complete on supercomputers using the Mahalanobis equation possible to be run at home on a standard PC laptop. QBEST is the multivariate analysis equation of the future and will allow for the processing of ever-increasing stores of complex Big Data.

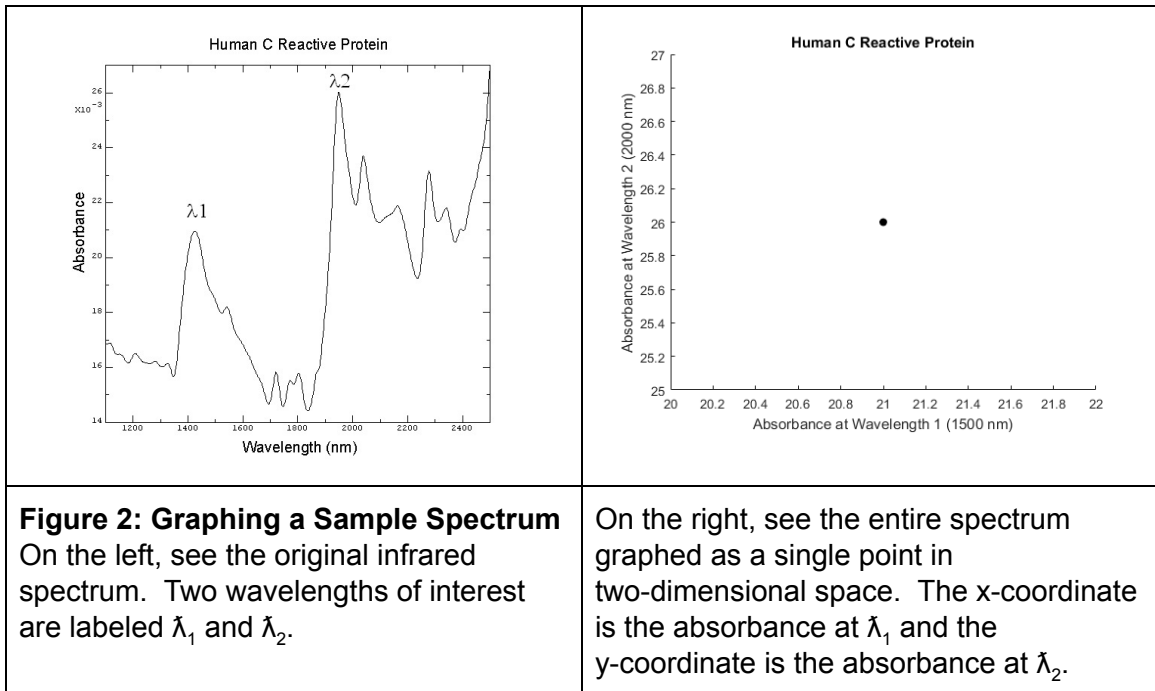
What is Multivariate Variation, in Layman's Terms?

Data are considered to be multivariate when two or more independent variables describe a single data point. For example, a person can be described by his height and weight, which can affect blood pressure. So a set of data containing people, characterized by height and weight to predict blood pressure, is multivariate. Any data that can be plotted in a 3-dimensional space (2 independent variables and 1 dependent variable) is multivariate by definition.

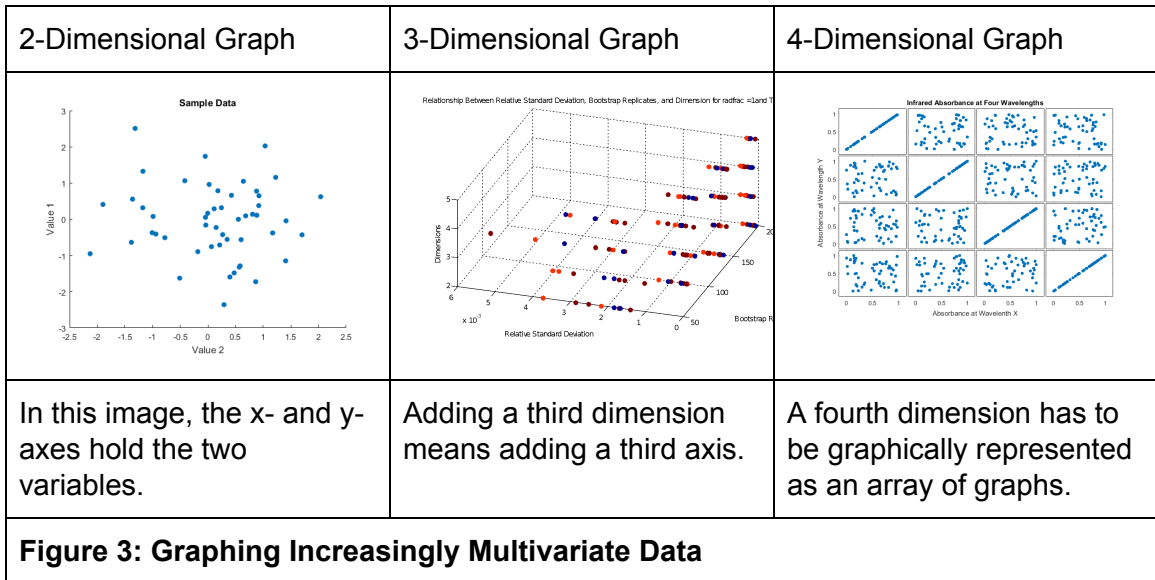
Furthermore, complex samples that seem on the surface to be a single continuous measurement can also be treated as multivariate data sets. Consider, for example, an infrared spectrum that represents the absorbance of light across a range of wavelengths:



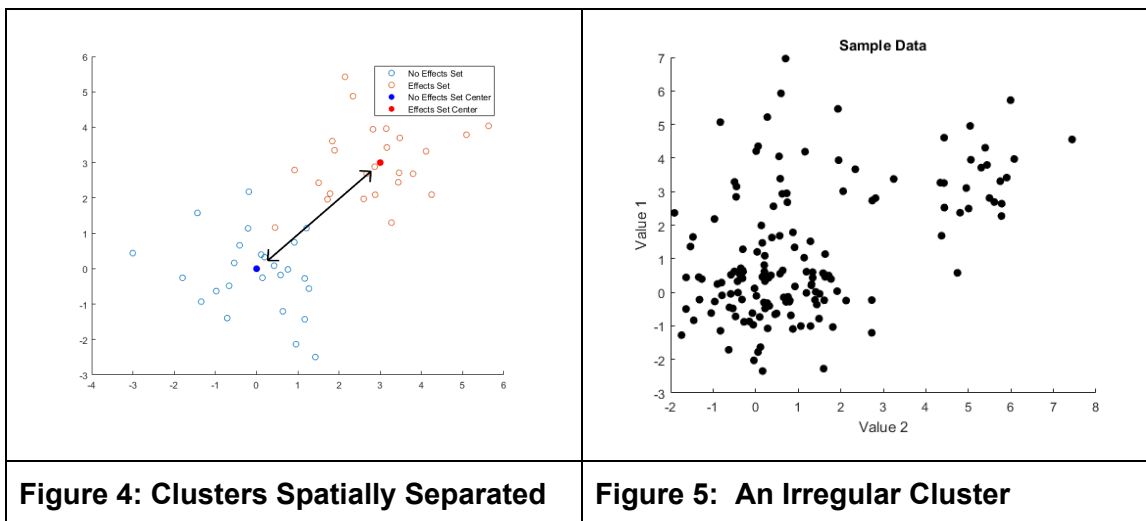
While these data may not seem to be multivariate at first glance, in reality, the absorbance at each individual wavelength of light used in the test may be considered a variable. This entire spectrum could be graphed as a single point in a large multivariate space. One would start by assigning each wavelength to an axis. For simplicity, let us first assume that we are only interested in λ_1 and λ_2 (labelled on the graph). We can assign absorbance at λ_1 to the x-axis and absorbance at λ_2 to the y-axis. Then, we can use the values derived from the original spectrum to graph a point reflecting these data.



We can use this method to graph any complex data, utilizing any number of axes. Pictured below are a standard x-y chart (a 2-dimensional graph) and a 3-dimensional graph. A graph can have any finite number of axes, though they become difficult to draw on paper and even more difficult to imagine. Though this is certainly difficult for a human, machines are incredibly capable at handling this kind of data. For this reason, it is important to be able to simplify complex data into a series of coordinate points that a computer can manipulate, and from the resulting clusters, determine statistical trends.



Finally, we can use a single graph to keep track of large numbers of different multivariate samples. Imagine that we had run an infrared spectral analysis on the same chemical a number of times, and each time we found slightly different results. We could graph all of the individual spectra on the same axis, as follows. When we look at the data, it becomes apparent that similar samples cluster together when represented as data points. These groups of points are referred to as clusters of data. These clusters have unique properties; the points may be very closely clustered along one axis, and more spread out along another. The significance of this will be discussed shortly.



Pharmaceutical Applications of Multivariate Data Analysis

Pharmaceutical applications of multivariate data analysis include the determination of the toxicity of novel pharmaceuticals. Toxicity determinations can be performed by experimentally determining the subject's current blood levels of a pharmaceutical and the dosage that subject was administered, and noting whether the pharmaceutical then produced an adverse effect in the subject. QBEST and discriminant cluster analysis can be applied to determine the safe dosage of the novel pharmaceutical agent. QBEST can be applied to the problem of calculating No-Observed Adverse Effect Levels (NOAELs) from published animal studies and clinical trials in a similar manner. The dosage and some numerical outcome variable (a minimum of one outcome variable) can be analyzed. Chapter 4 of this work addresses the generation of QNOAELs from the literature.

Additionally, QBEST can be used in the future to manage the increasingly vast amounts of health data generated in clinical trials and post marketing studies. The Mahalanobis equation is incapable of calculating a Mahalanobis distance in multidimensional SDs for all the data generated by many clinical trials. This is because the Mahalanobis requires

more data observations (e.g., subjects) than dimensions or independent variables (e.g., blood chemistry values, pathology measurements, etc.) in order to be calculated. Additionally, clinical studies observe a great number of other characteristics ranging from blood pressure and heart rate to weight and food intake. A single clinical study may record 50 or 100 data values for each patient. Unfortunately, because clinical trials are expensive, a phase 1 or phase 2 trial can easily have more data variables than subjects. Furthermore, in meta-analyses, it is unlikely that more than 50 or 100 clinical trials would be reported in the literature for a given pharmaceutical, again in large part because of cost. This means that all 50 or 100 variables measured could not be used as data in a Mahalanobis distance calculated by the Mahalanobis equation. QBEST, however, can be used even when the number of variables measured far exceeds the number of samples (subjects or clinical trials) measured. Data supporting this (situations where number of dimensions is far higher than number of training points) can be found in Chapter 5. For this reason, QBEST offers the best solution to solving clinical trial data problems and using all of the data generated in one single statistical test.

Even personal health monitoring systems could be improved by the implementation of QBEST. Numerous personal fitness systems currently monitor consumers' vitals almost continuously. Smartwatches, FitBits™, phones, and similar devices store information about heart rate, activity level, and even sleep quality. These devices generally establish a baseline for the person's vitals and note if an abnormal event happens. QBEST could easily be implemented in these device algorithms to detect the significance of deviation from normal, and to identify the nature of the aberration. For example, a test heart rate and pulse pattern could be compared to the patient's baseline training set as well as to a stored library of heart attack patterns and arrhythmias. Using QBEST, even a simple smartwatch could serve as an early detection system for heart conditions. Of course, if QBEST could be implemented in consumer devices, it could be put to even better use in medical devices such as continuous glucose monitoring insulin pumps.

Why a Statistical Method is Needed Beyond Simple Standard Deviation:

Most individuals involved in the sciences are familiar with the standard deviation. A standard deviation is a measure of variation around a central value. Distance in standard deviations often serves as a sort of limit by which we measure the reproducibility of data. (A limit of 3 SDs or 6 SDs is commonly used to denote class membership, for example.) I could state that I am 5'5" tall. Does that make me typical for a human being? Or am I gigantic, or perhaps diminutive? Answering that question is simple enough; One could estimate the average height of a human being with a good sampling of human beings, and then calculate the standard deviation of height. 68% of the height data should fall within one standard deviation of average if the heights follow the normal (Gaussian) distribution. By comparing my height to the average human height plus or minus one standard deviation, it becomes apparent that I fall within the +/- 34% of most people. So 5'5" tall is not an abnormal height.

But standard deviation is a mathematical parameter that is defined only in regard to a single variable. While it is possible to answer simple questions with standard deviation, such as “Is my height typical?”, more complicated questions cannot be answered using standard deviation alone. For instance, take the question, “Is my height and weight typical for a human?” It becomes readily apparent that adding a second variable (weight), thus making the problem multivariate, complicates the question considerably. To know whether my height and weight are typical, I need to know four things:

- 1) What is the average height and weight of all humans?
- 2) How much does height vary from person to person? (What is the standard deviation of height?)
- 3) How much does weight vary from person to person? (What is the standard deviation of weight?)

-and-

4) How do height and weight vary with regard to each other?

A covariance matrix is required to answer that last question. The covariance matrix is part of the Mahalanobis distance Equation. Specifically, the Mahalanobis equation represents the first statistical method capable of determining the difference between a test data point and a known cluster of multivariate data.

What is the Mahalanobis distance Equation?

The Mahalanobis metric has been described as “a rubber ruler.”^{2,3} The rubber ruler analogy can be explained as follows, by graphing a data sample of coordinate points.^{2,3}

1. The Mahalanobis equation starts with the center of a data sample. Along any line drawn in coordinate space that passes through the center of the data distribution, the Mahalanobis equation can be used to determine the scalar size of a standard deviation unit. In this way, the scalar standard deviation units are the markings along the ruler. The ruler stretches to different lengths in different directions to maintain 1 SD as a central 68% confidence interval for normally distributed data. In the figure below, the distribution is wider than it is high, so the ruler must stretch in the horizontal direction compared to the vertical direction.

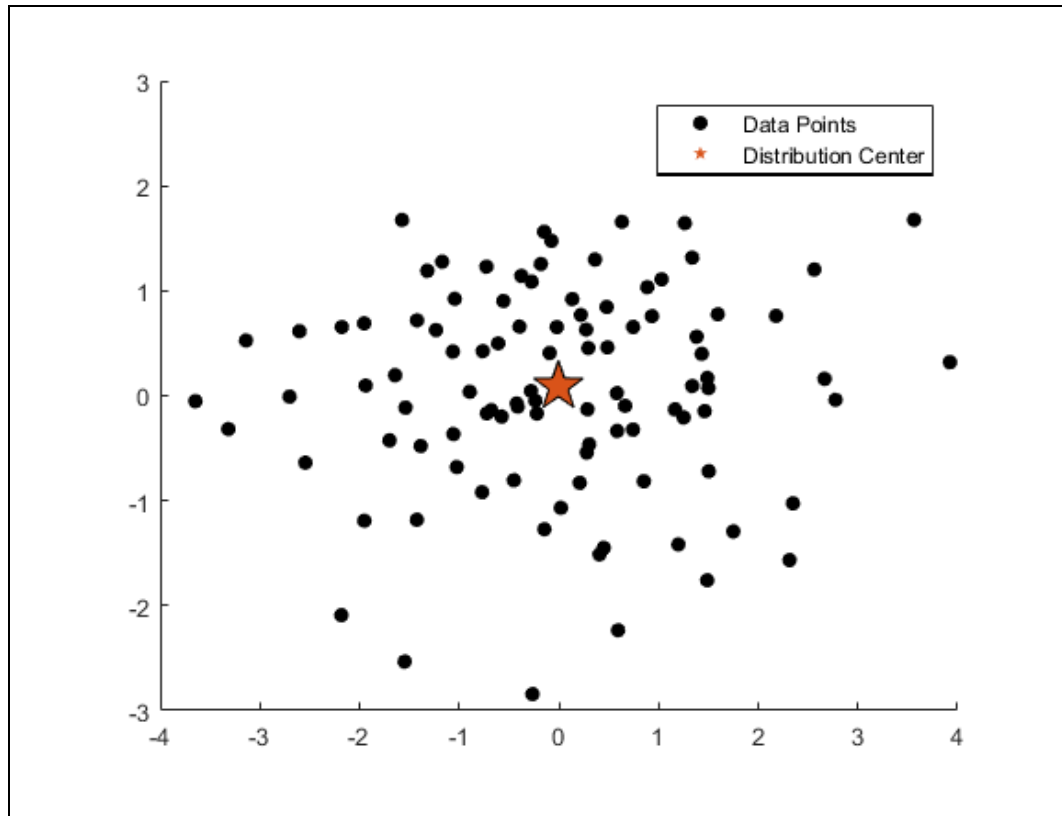


Figure 6.1: A Sample Distribution with Known Center

2. Rubber rulers can be thought of as being laid from end to end across the distribution to make the initial measurement to a new sample point (a point not in the original training set). Imagine laying a blank ruler one SD long across the data in a line through the center. Then, like a tailor flipping a ruler end-over-end to measure a bolt of cloth, we can extend that measurement infinitely out into hyperspace in the direction of each new sample point.

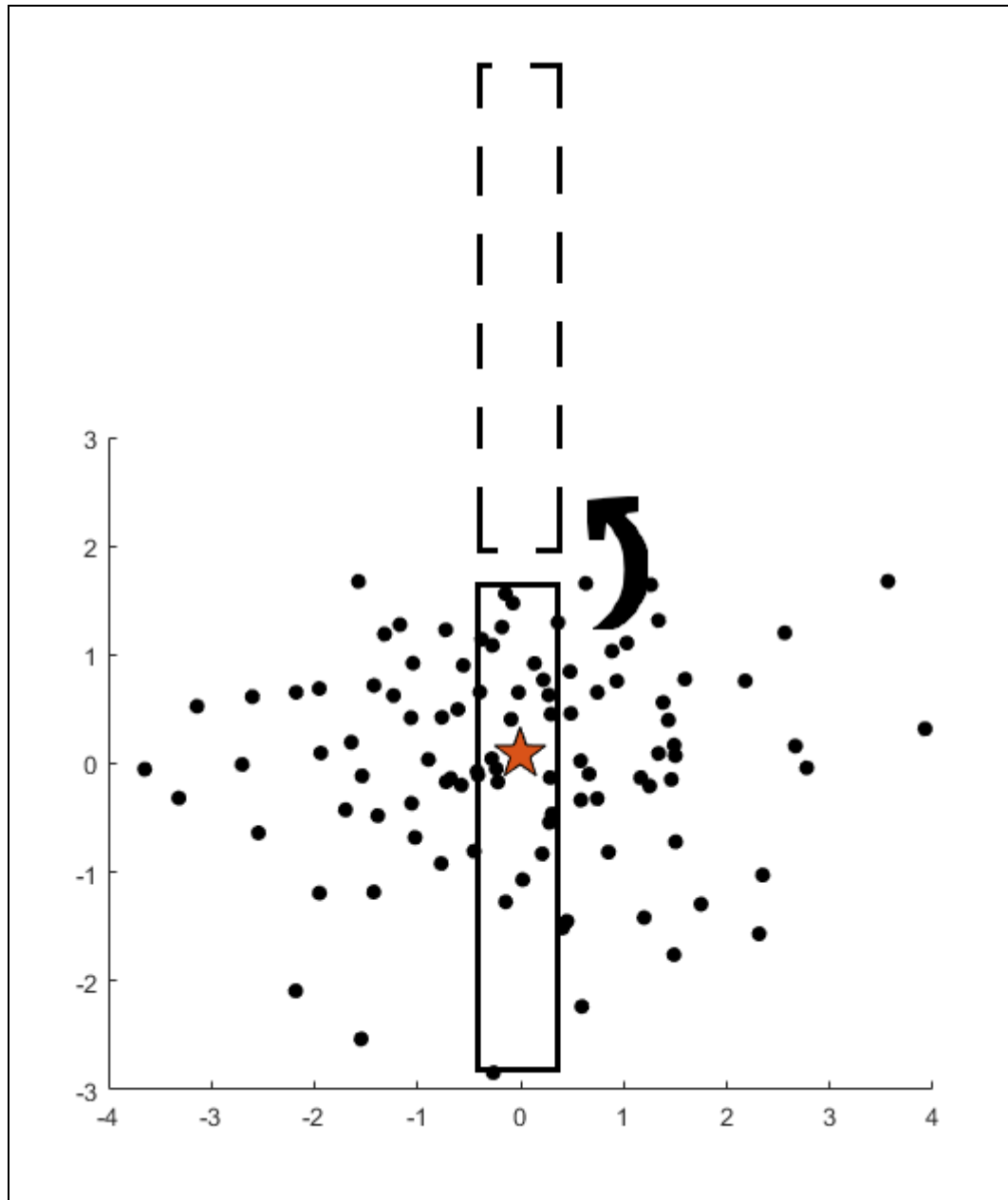


Figure 6.2: Centering the Rubber Ruler; Flipping the Ruler Out into Hyperspace

3. The markings on a “rubber ruler” indicate the number of standard deviation units a new sample point is from center. This is the “ruler” part of the rubber ruler analogy (with more than 1 SD marked on it).

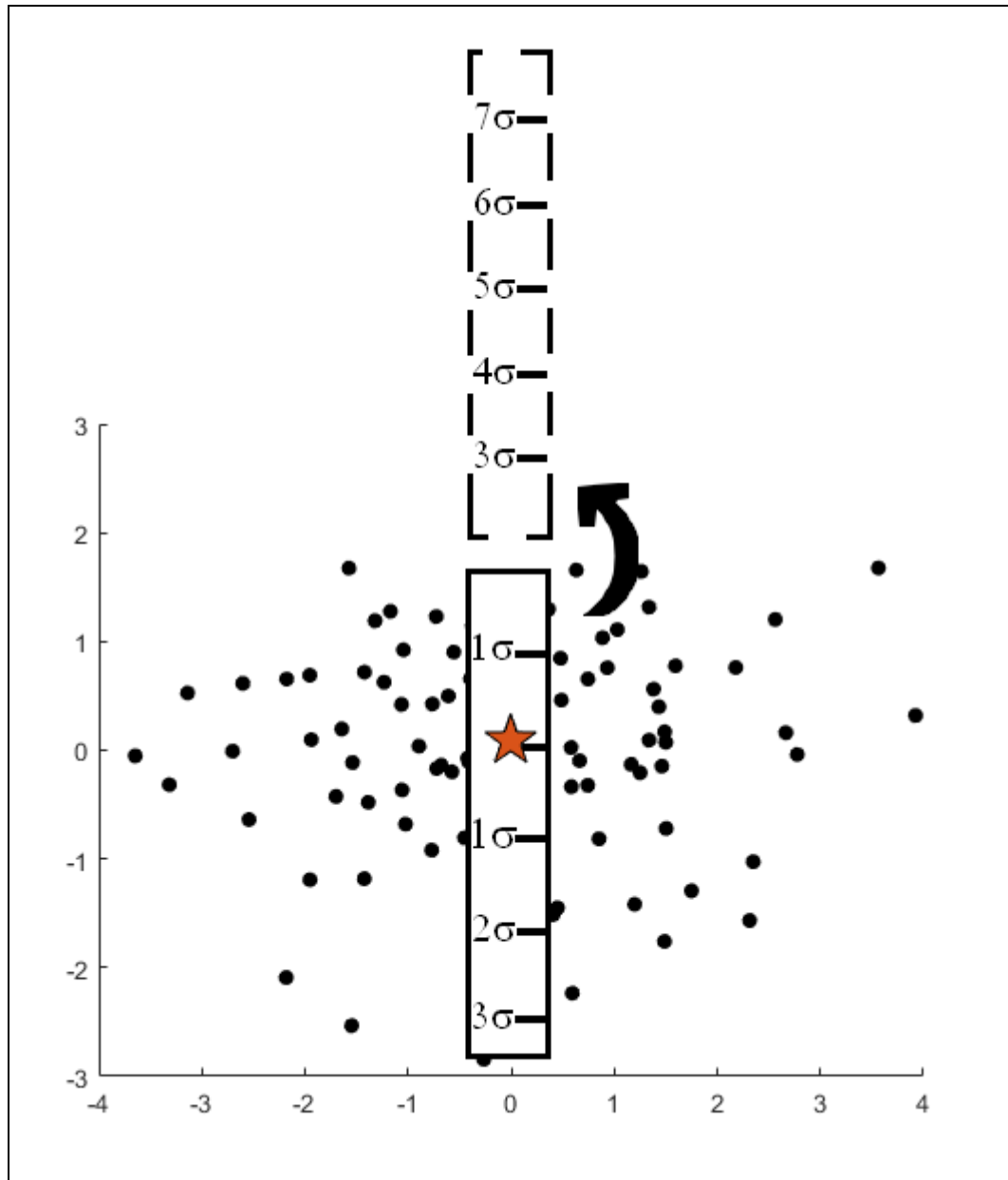
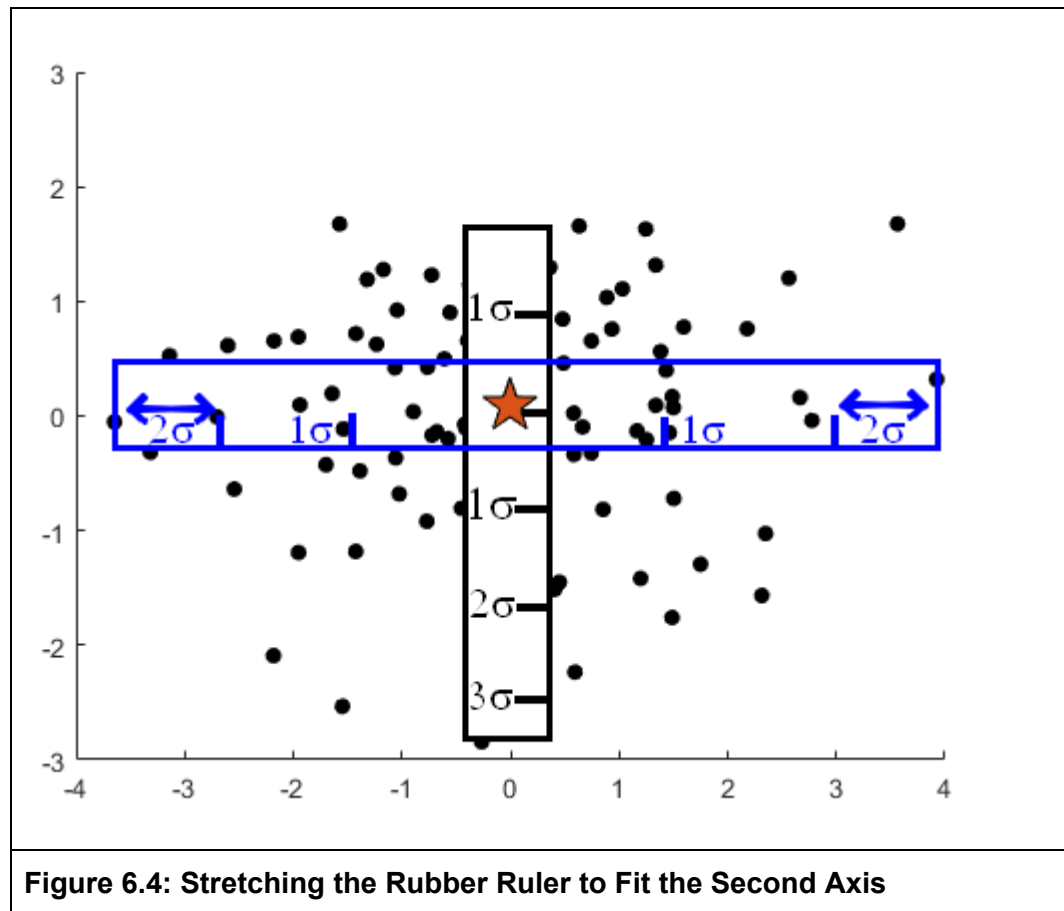


Figure 6.3: Determining Standard Deviation Unit Size Along the First Axis

4. The “rubber” part of the analogy is illustrated when you measure in a new direction. The same Mahalanobis equation can be used then to determine the standard deviation across any other line passing through the center of the distribution. Imagine picking up the ruler that you just made and rotating it 90 degrees to find the standard deviation units across the orthogonal direction of the data cluster. You center the ruler on the center of the cluster, but what do you find? The ruler is too short to cover the span of the data in this direction! So, you pinch both ends of the ruler and stretch it out until it extends to both ends of the

distribution. Now you know the standard deviation units of points along this axis as well.



5. In statistical terms, what this “rubber ruler” means is that SD is not equivalent along all directions; and to determine the confidence with which a single datapoint fits a distribution, your statistical test must be able to take into account the differing variation along any direction in a cluster, a major axis, a minor one, or a line falling anywhere in between.

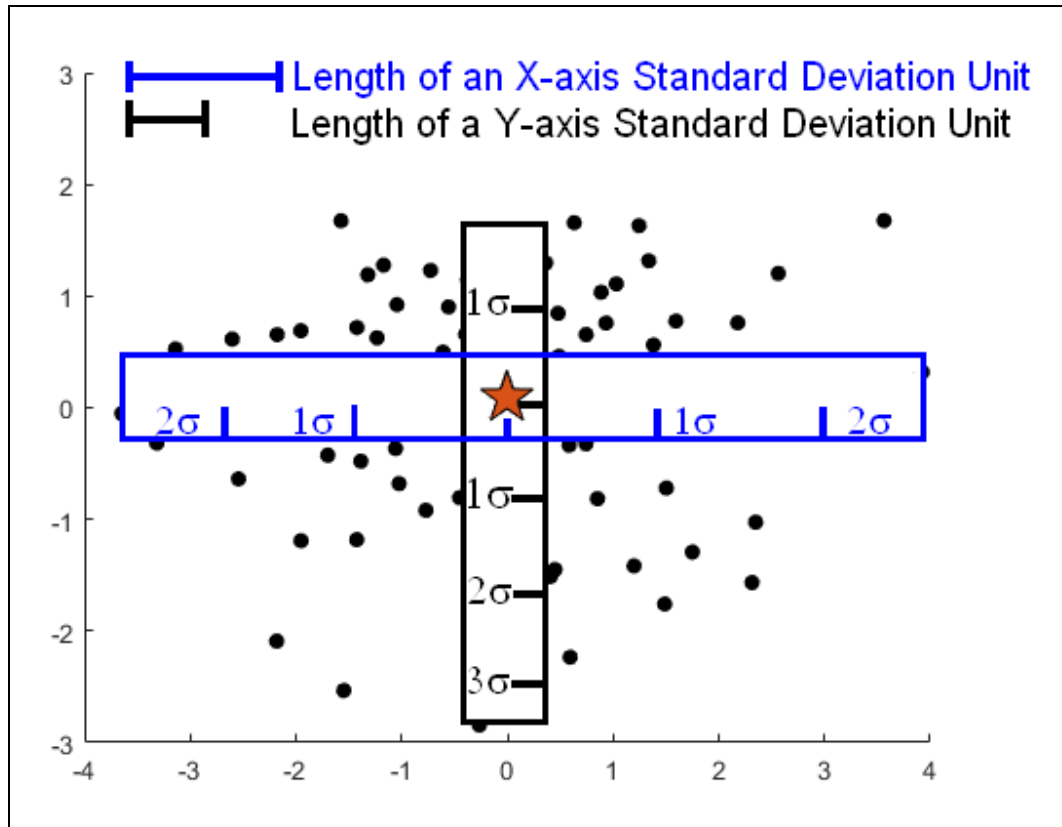


Figure 6.5: Determining Standard Deviation Unit Size Along the Second Axis

- By laying the “rubber ruler” over our cluster, we can see that one unit of spatial distance can be very big along one axis, while it is not so in another. The Mahalanobis equation allows us to quantify this change.

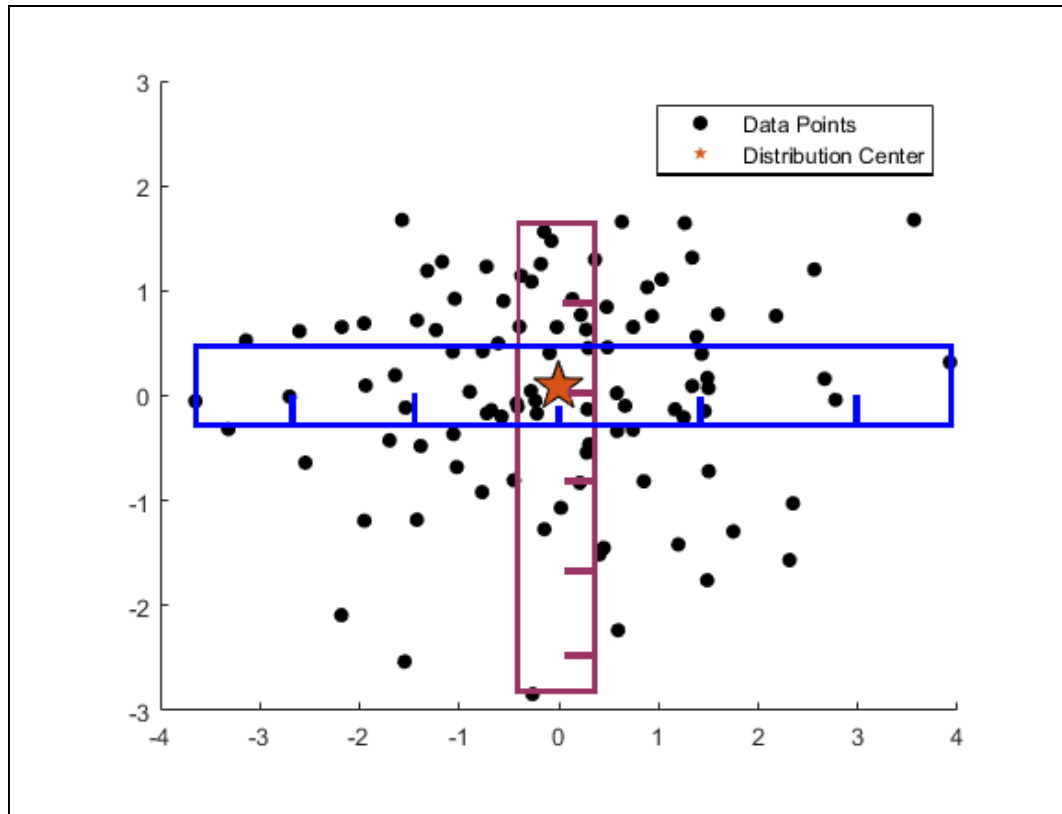


Figure 6.6: The Standard Deviation Unit Size is Reflective of the Spread of the Data Along Each Axis

Why is the Mahalanobis distance Insufficient?

The Mahalanobis equation is Inaccurate When Handling Asymmetrical and Irregular Data Clusters

Because the Mahalanobis equation is a rubber ruler whose length is determined by the span of a data cluster on a line passing through the center, the Mahalanobis equation is inherently prone to error when asymmetrical or irregular clusters of data are involved. Consider the following theoretical data clusters:

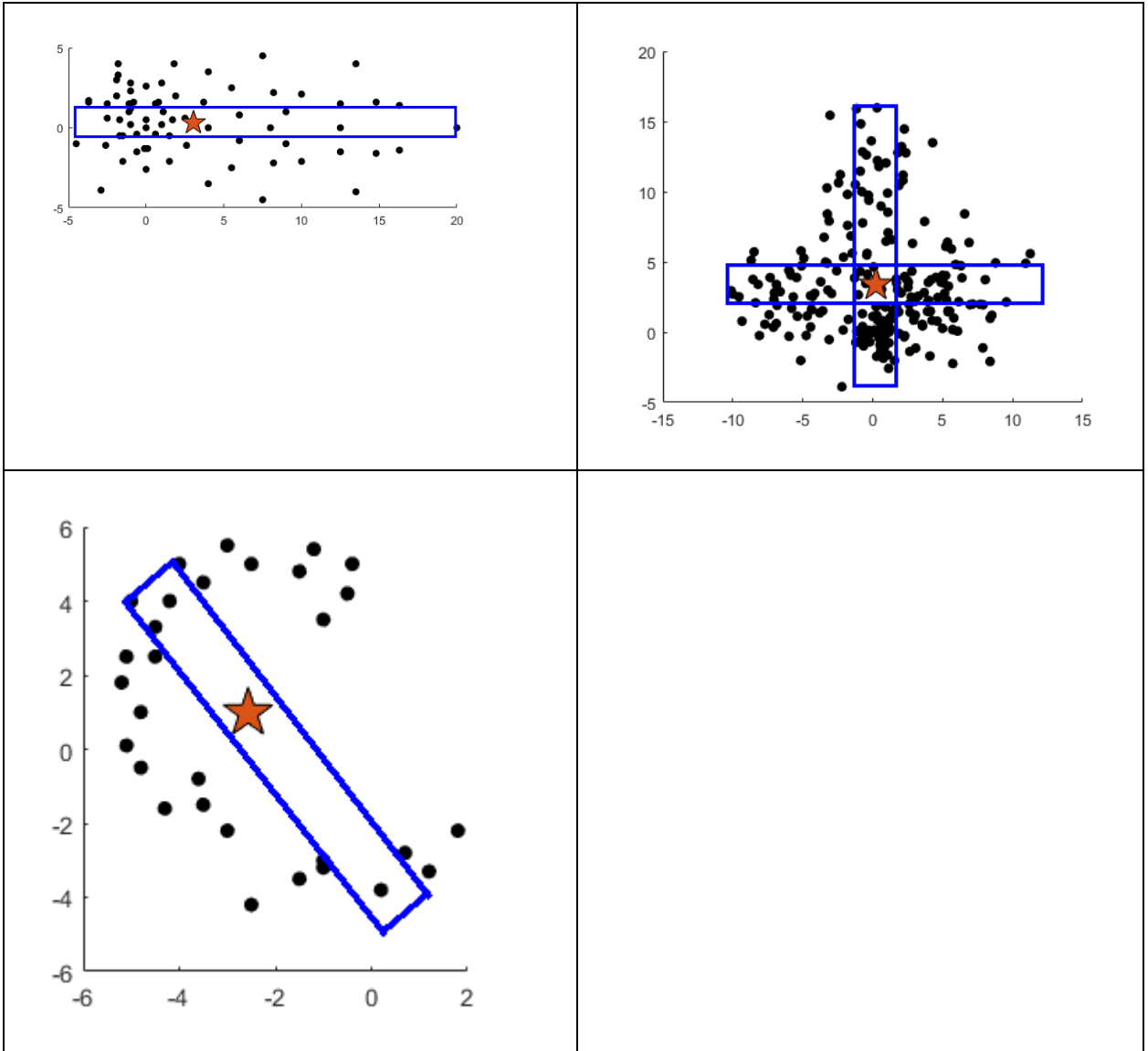


Figure 7: Three Asymmetrically-Distributed Data Clusters with Mahalanobis Rubber Ruler Drawn

Do the proposed rubber rulers really fit? Visually, it is apparent that it is ridiculous to measure the left and right-hand sides of these distributions with the same units of standard deviation. All of the pictured distributions have unequal dispersion along the left-hand or right-hand directional vector. This is where the QBEST is better, as will be explained later, because QBEST independently determines the right-hand and left-hand standard deviation units, that is the standard deviations along a line in opposite directions.

The Mahalanobis equation is Mathematically Incapable of Calculating Distances For Datasets with More Variables Than Observations

The Mahalanobis equation involves several matrix math operations that cannot be performed when there are fewer rows than columns of the matrices. In terms of pharmaceutical science, the Mahalanobis equation cannot be computed when the number of observations is less than the number of variables that describe those observations. Moreover, even when the number of samples exceeds the number of data observations, as the number of variables approaches the number of samples observed (the number of rows approaches the number of columns), the Mahalanobis equation loses accuracy and precision. However, QBEST is able to calculate a distance in SDs under those conditions with accuracy and precision.

What This Means in a Pharmaceutical Context

This capability of QBEST is particularly important in applications of multivariate data processing such as the meta-analysis of clinical trials. In such instances, a very few clinical studies may have been performed (clinical studies are expensive), but each study collected perhaps 100 endpoints. Even animal studies are very expensive to perform, both in terms of money and time resources, but human studies are even more so. As a result, studies tend to be fairly limited in quantity, and the researchers who have performed them use their subjects to the maximum extent possible. Animal studies and human subjects trials tend to have a lot of clinical endpoints (variables) monitored and reported at the end of the study.

Clinical monitoring is a multivariate by nature. All of the many biological factors noted about an organism are variables that could give some insight into the organisms well-being or condition. Until now, it has not been possible, mathematically speaking, to process all of the data provided by handful of clinical studies using a single comprehensive statistical test. The QBEST is the first statistical method following the form of the Mahalanobis equation that is able to comprehensively test all data as a single multivariate data points comparable to other multivariate data point clusters, and is also able mathematically to process a matrix with a greater number of columns than rows. In more practical terms, what this means is that QBEST is capable of analyzing, for instance, six clinical trials that each assayed the same 60 clinical end points. That feat is mathematically impossible using the Mahalanobis equation.

The Mahalanobis Requires Too Much Computing Power

The Mahalanobis equation is an order of d^3 equation, which means that for every d increase in the number of dimensions or variables describing the data, the equation takes d^3 more computing power to process. QBEST is an order of d equation, which means that for every d increase in the number of dimensions or variables describing the data, the equation takes d more computing power to process (a d^2 advantage in running

time) . This allows QBEST to quickly calculate the solution to Big Data problems with run times that would be prohibitively long for Mahalanobis equation calculation.

What is the Quantile Bootstrap Error Statistical Test in Detail?

The Quantile Bootstrap Error Statistical Test (QBEST) is a novel statistical method for the determination of the degree of difference between a test point and a training set of data points.

QBEST functions first by taking bootstrap replicates from a sample of data (called the training set). Then QBEST determines the center of the training set. QBEST relies on the Euclidean metric to determine the spatial distance between the two points along a straight line. (These straight lines are the hyperline between the center of the training set and each new sample test point, and the hyperline connecting the center of the training set to each bootstrap replicate point. The word “hyperline” is used because this line ordinarily exists in a hyperspace of many dimensions.) While it would be ideal to be able to consider just bootstrap replicate points that fall on this hyperline in making measurements of standard deviation in that direction, a line is infinitely thin, so the likelihood that any other given bootstrap replicate point will fall directly upon the hyperline is incredibly low.

If the hyperline used to measure the distance between the training set center and a test point is infinitely thin and no bootstrap replicates in the distribution fall on it, how do we measure the distribution of bootstrap replicate points along that hyperline in the direction of the new sample point? Instead of using an infinitely thin line, we instead use a hypercylinder in hyperspace. The radius of the hyper cylinder is increased until at least 50 bootstrap replicate points are captured within it. Replicate points in the direction of the test point will be engulfed by the growing hypercylinder. The QBEST equation requires that the user specify a “radfrac” or fraction of the training point spectra that should be included in this hypercylinder. Basically, the radfrac determines how big the radius of the hypercylinder should be - fat or thin depending on what proportion of the points should be included. The computer sets the radius based off of the radfrac input by the user.

Why is QBEST Superior to the Mahalanobis equation?

The Mahalanobis equation is limited in two very simple but very critical ways. First, the Mahalanobis equation is incapable of calculating a Mahalanobis distance when the number of variables comprising the data exceed the number of samples.³ That is to say, the Mahalanobis equation requires that the number of observations must far exceed the number of variables of the multivariate data In order to produce results with accuracy

and precision. Secondly, the Mahalanobis equation is very costly calculation to run in terms of computation.³ This means that for a given calculation, the Mahalanobis equation needs significantly more time and computational power than the QBEST does (more on that later).³

Observations vs. Variables

What This Means for Pharmaceutical Studies

Clinical trials and animal studies are incredibly expensive. This naturally limits the number of such studies that are conducted and published in the literature. Imagine that a dozen or so clinical trials are performed for a given therapeutic agent with roughly the same endpoints being measured. In these studies, blood concentrations of the drug are measured at many time points. Continuous blood pressure and heart rate monitoring are performed. Additionally, metabolic panels, lipid panels, and a number of drug-specific tests are done and the outcomes reported. Remember that continuous data can be broken down into a series of variables, so that where time is important, each time point at which blood pressure (for example) was measured becomes an additional variable. It is reasonable to expect that perhaps one hundred variables are measured in the course of a single clinical trial.

In this example, it is mathematically impossible to calculate a Mahalanobis distance using the Mahalanobis equation. There are simply too few clinical trials in relation to the number of variables measured for each trial. QBEST, however, can handle the given problem with statistical accuracy and precision.

Computational Cost

The Mahalanobis equation is order of d^3 in computational complexity, which is to say that for every d increase in variables (or dimensions), the computation cost increases by d^3 .³ In practical application, this causes the computer running the Mahalanobis equation to very quickly become “bogged down” as more variables are collected.

QBEST is an order of d equation, which means that for every d increase in the number of dimensions of the data, QBEST requires only d more memory and computational power to complete the calculation. The same problem that would cause a supercomputer running the Mahalanobis equation to freeze or take hundreds of years to run can easily be completed using the QBEST method on a cheap personal laptop computer.³ QBEST makes advanced computation possible for individuals, researchers, and small corporations using the most meager of resources; QBEST allows researchers using advanced supercomputers to attempt multivariate data problems not yet dreamed of.

What is the QBEST?

“A Rubber Ruler with a Nail in the Center”

If the Mahalanobis equation is a rubber ruler, then the QBEST equation is a rubber ruler with a nail in the center. The QBEST ruler can be stretched longer or shorter independently on opposite sides of a data cluster. In this way, it is capable of independently calculating different-sized standard deviation units for opposite halves of an asymmetric data cluster^{1,3-9}.

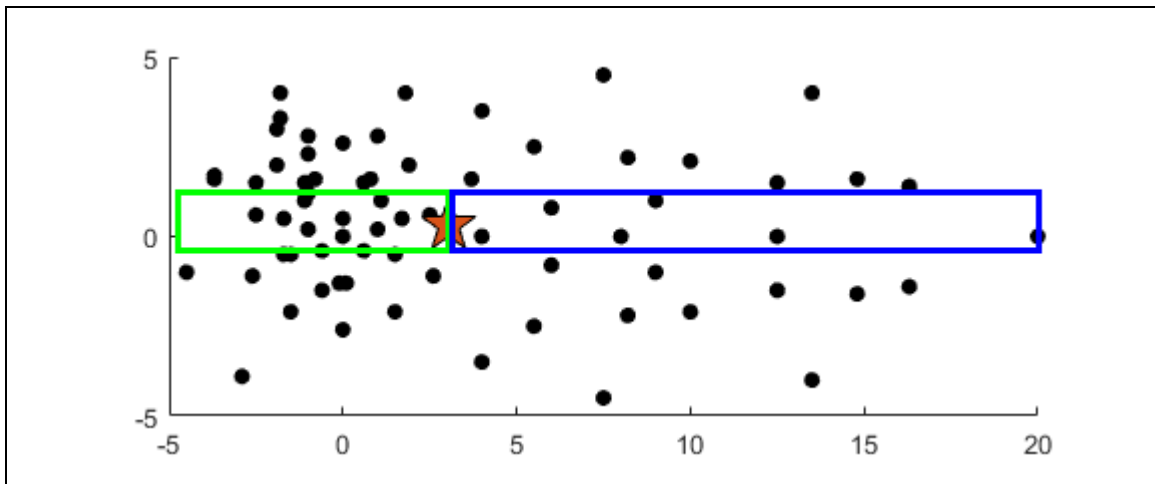


Figure 8: QBEST is a “Rubber Ruler with a Nail”

Shown in green is the left-hand QBEST rubber ruler with a small standard deviation. Shown in blue is the right-hand QBEST rubber ruler with a larger standard deviation. The standard deviation units for each side are calculated independently by QBEST, which is reflective of the skew of the data.

How Does QBEST Work?

QBEST is a multivariate nonparametric statistical method. First, the original training set is replicated using a Monte Carlo integration of the bootstrap distribution function. The replicates allow for the approximation of traditional parametric statistics like bias, variance, confidence intervals and prediction error.⁵⁻⁹

The QBEST calculates the number of standard deviation units between a test point and the center of the training set. It does this by calculating the Euclidean distance between the two points and scaling that distance with the Euclidean distance of one SD of the training set points. A hypercylinder is used for the selection of comparable training spectrum points.⁵⁻⁹

The equation for determining two-dimensional Euclidean distance is defined as the following:

$$\text{dist}((x,y),(a,b)) = \sqrt{(x-a)^2 + (y-b)^2}$$

The equation for determining multidimensional Euclidean distance is defined as follows:

$$D_{ij}^2 = \sum_{v=1}^d (x_{vi} - x_{vj})^2$$

What is Discriminant Cluster Analysis?

Simply put, discriminant cluster analysis is the process of determining whether a given point belongs to one cluster of data points or another.

How Can Discriminant Cluster Analysis be Applied to Drug Development?

Detecting Contamination

Suppose there are two clusters of data, one describing safe dosages of a pharmaceutical and one describing adulterated drug. Given a new unknown test article, analysis via QBEST would determine the likelihood that the test point belonged to one category (safe) or the other (adulterated). This method has been used to differentiate between the near infrared spectra of pharmaceutical capsules contaminated with cyanide and normal capsules.^{4,5}

No Observed Adverse Effect Limits

Alternatively, suppose that the two clusters of data represent drug studies done on a novel pharmaceutical agent. Each point represents one dosage given in a study along with any number of multivariate factors on different spatial dimensions, including pharmacokinetic information like Cmax, or a measure of physiological changes induced by the medication, such as a measure of the increase in heart rate after dosage. One cluster represents dosages that produced no adverse effects in the patient; the other represents dosages that caused some adverse effect. From the existing literature available on this novel therapeutic, we want to determine the maximum “safe” dose of the agent to aid us in powering our clinical studies. To do this, we would like to determine the No Observed Adverse Effect Limit (NOAEL). This metric represents the highest dose of an agent that can be administered to a patient that will cause no adverse effects. (Any higher dose is likely to cause some negative effect.)

Traditionally, NOAELs and No Observed Effect Levels (NOELs) have been experimentally determined in a single experiment by administering increasingly high

doses of an agent to patients and determining which of the doses was the highest given that produced no adverse effect. The scientific problems with this methodology are fairly obvious: it is impossible to determine a NOAEL or NOEL value for any dose other than an exact dosage that was administered, leaving much room for error because the levels are usually spaced an order of magnitude or more apart. NOAEL studies cannot be easily meta-analyzed or pooled, as methodology of the study protocol can greatly influence the outcome of the study. It is costly to repeat NOAEL studies to increase precision and requires the use of new human or animal subjects.

Discriminant cluster analysis using QBEST can be used for meta-analysis in place of a traditional NOAEL study or the Benchmark Dose Method . QBEST provides greater accuracy and reproducibility than a traditional NOAEL and yields more intuitively comprehensible results than a BMD (Benchmark Dose) analysis. Furthermore, performing a discriminant cluster meta-analysis re-utilizes data that has already been generated and published in the literature and prevents the waste of resources and research animals in redundant studies. Finally, QBEST can provide an estimate of an effect level that can be used to power additional preclinical and clinical trials. QNOAEL and QNOEL determination are described in detail in Chapter 2 and Chapter 4.

What is new about QBEST?

Nonparametric Method

QBEST represents a novel means to analyze data in several ways. First, QBEST is a nonparametric statistical method, which means that it is capable of treating data that does not follow a known statistical distribution (such as a normal distribution). Parametric methods require an assumption that the data follows a certain distribution. If these assumptions are not met, the results of the analysis are inaccurate and it is inappropriate to apply the method in question. QBEST is unique in that it makes no assumptions about the distribution of the data. As such, QBEST can also be applied to parametric data.

Unique Handling of Skewed Data

QBEST is novel in that it treats opposite sides of a data cluster separately when determining standard deviation units from center. QBEST does not assume that the left- and right-hand sides of a cluster are equally distributed. This allows QBEST to accurately determine the significance of the distance of a test point from the center of a cluster even when the cluster is abnormally-shaped. The most widely used method, the Mahalanobis method, does not account for skewed clusters. The QBEST distance is considerably more accurate than the Mahalanobis distance for irregular and skewed data clusters.

Discriminant Cluster Analysis Applied Non-Traditionally

Discriminant cluster analysis, the method of determining whether a test point belongs to one or the other possibly overlapping classified clusters of data points, has never been applied to the determination of toxicology studies¹¹. This method is primarily used in analytical chemistry for the classification of substances and the determination of impurities. The method has been used in other applications, but is not used as widely as perhaps it should be. Some current applications include analysis in the tourism industry and data analysis in banking^{10, 11}. Discriminant cluster analysis is an incredibly powerful statistical technique with many possible applications from business to science to artificial intelligence and machine learning. The application of discriminant cluster analysis to toxicology data is a novel aspect of this work that lays the ground for future medical applications of discriminant cluster analysis.

How do you use QBEST?

The Variables

The Training Set

The training set is the set of data points of known outcomes used to test a new (unknown) "test point". Training sets are common in calibrating neural networks, for example. In a pharmaceutical context, the training sets could be groups of points representing preclinical or clinical studies of medication that caused no adverse effects and groups of points representing studies of medication that caused adverse effects. A test point could be an individual study.

The training points input into the QBEST are critical to obtaining valid results. The old maxim, "garbage in, garbage out" holds for all training sets, not just QBEST, but the Mahalanobis equation and neural networks as well. As with all data analysis, first and foremost it is important that quality data representative of the population are used. Precautions such as Cochrane Review of data can ensure that appropriate data are chosen in the first place.

After quality data have been selected, it is the task of the researcher to inspect the data. The size of the training set and the distribution of training points determine the value of other parameters that should be set when using the QBEST equation. Guidelines for setting these parameters can be found in Chapter 4.

Radfrac

The radfrac is the fraction of bootstrap replicate points used to generate the hypercylinder. Points within the hypercylinder form the basis of statistical comparison to the test point.

When is it appropriate to use a large/small radfrac?

Radfrac is important because it determines how bootstrap replicate points are determined to be relevant to the test point. When the number of bootstrap replicate points is very large, one can decrease the radfrac because the large number of points provides more in the hypercylinder, and narrowing the radius of the hypercylinder increases accuracy in this case for clusters that contain a concavity. Furthermore, when the cluster is highly irregular, it is important to decrease the radfrac in order to avoid including points that do not fit a concave surface for that part of the cluster.

Number of Bootstrap Replicates

The number of bootstrap replicate points made for a new training set is important because having more points increases precision, albeit at a small cost of increased memory and computation time.

Number of Variables or Dimensions

The number of variables or dimensions of the multivariate data affects the accuracy of the algorithm. The algorithm is more accurate when largely multivariate data also has a large training set size.

When Is QBEST Insufficient?

QBEST is meant for the analysis of multivariate data. It can be used effectively for numerical response variables. QBEST can also theoretically be used to compare ordinal or binary response variables (given that the variable is assigned a numerical score) alone or in combination with numerical response variables. However, at present, no testing has been done on ordinal or binary response data.

As with any meta-analysis method, the accuracy of QBEST is most negatively impacted when the quality of the training set of data is poor. Increasing the quality of the training

set and the quantity of the training set observations increases the accuracy and reproducibility of QBEST. When a very few training set observations are supplied, the reproducibility of QBEST's output is decreased. These trends are discussed in detail in Chapter 5.

As with any other statistical method, the researcher should always visually inspect all combinations of two variables of the data graphed against one another to ensure that there are no problems (like nonlinearities) that would complicate discriminant cluster analysis. QBEST takes into account the typical variation in points. So, for example, for ultrasonic resonance spectral data that follow a Lissajous function, QBEST may generate a less accurate probability measurement.

Many irregular clusters can be evaluated with accuracy if the radfrac is set to be appropriately high or low after the data are inspected. The danger of setting radfrac without visually inspecting the data on all axes is that too high of a radfrac could include points that are not indicative of the trend along a given vector, while too low of a radfrac would exclude points that ought to be included. This principle is especially important for data clusters that include concavities, which could otherwise be handled poorly by the algorithm. (See the Conclusion for an explanation of data with concavities.) The selection of an appropriate radfrac for the shape of an irregular cluster of data very strongly affects the accuracy of the algorithm and great care should be taken in assigning this value.

CHAPTER 2: A NOVEL STATISTICAL APPROACH TO NOAEL: QBEST APPLIED TO DOSING OF ELLAGIC ACID AND THE QNOAEL VS. BMD FOR POINT OF DEPARTURE

Purpose: This article describes the research strategy and goals for this thesis. This chapter was published as “The QNOAEL vs. BMD for Point of Departure” on BioRxiv. It is the preprint of a later paper, published as “A Novel Statistical Approach to NOAEL: QBEST Applied to Dosing of Ellagic Acid,” on Webmedcentral.com.

Publication and Copyright Information:

This section was published to Biorxiv.com as follows:

Cynthia Dickerson, Robert A. Lodder. (2018) “The QNOAEL vs. BMD for Point of Departure.” 24 May 2018, bioRxiv doi: <https://doi.org/10.1101/329763>

“The copyright holder for this preprint (which was not peer-reviewed) is the author/funder. It is made available under a CC-BY 4.0 International license.”

The QNOAEL vs. BMD for Point of Departure

Abstract

Quantile bootstrap (QB) methods can be applied to the problem of estimating the No Observed Adverse Effect Level (NOAEL) of a New Molecular Entity (NME) to anticipate a safe starting dose for beginning clinical trials. An estimate of the NOAEL from the extended QB method (called the QNOAEL) can be calculated using multiple disparate studies in the literature and/or from laboratory experiments. The QNOAEL is similar in some ways to the Benchmark Dose (BMD) and is superior to the BMD in others. The Benchmark Dose method is currently widely used in toxicological research. Results are used in a simulation based on nonparametric cluster analysis methods to calculate confidence levels on the difference between the Effect and the No Effect studies. The QNOAEL simulation generates an intuitive curve that is comparable to the dose-response curve.

The QNOAEL of ellagic acid (EA) will be calculated for clinical trials of its use as a component therapeutic agent (in BSN476) for treating Chikungunya infections. This will be the first application of QB to the problem of NOAEL estimation for a drug. The specific aims of the proposed study are to evaluate the accuracy and precision of the QB Simulation and QNOAEL compared to the Benchmark Dose Method, and to calculate the QNOAEL of EA for BSN476 Drug Development.

Specific Aims

Nonparametric statistics are statistics that are not based on parameterized families of probability distributions, like the normal distribution. They are important because data frequently follow a distribution other than a known one, like the normal distribution. The NOAEL is an important part of the non-clinical risk assessment for new drugs like BSN476, a drug for treating Chikungunya. The NOAEL is a professional opinion based on the design of the study, indication of the drug, expected pharmacology, and spectrum of off-target effects. It is the highest dose at which there was not an observed toxic or adverse effect¹¹. There are important theoretical limitations to the traditional NOAEL calculation, which led to the newer Benchmark Dose method, which also has a number of problems. In brief, the traditional NOAEL is determined by administering a few different doses of drug to a group of subjects, observing those subjects for physiological change, and assigning the dosages to the categories of “having an adverse effect” and “not having an adverse effect”. The highest dosage resulting in no adverse effect is determined to be the NOAEL.

The NOAEL method is problematic because (1) dose levels are often an order of magnitude apart, and it is highly unlikely that the exact NOAEL dosage will be

administered in any particular study; (2) determination of what constitutes an “effect” can be difficult when negative effects are of a highly subjective nature (for example, when mood is affected)⁴. A nonparametric simulation using extended QB (Quantile Bootstrap) methods can solve the problems associated with the use of the traditional NOAEL or the Benchmark Dose (BMD) and enable accurate toxicological estimates to be made.

Evaluate the accuracy and precision of the QB Simulation and QNOAEL compared to the Benchmark Dose Method

Utilizing synthetic data with known characteristics, the BMD and QNOAEL will be calculated. The QNOAEL will then be compared to the BMD to determine which is closest to be known answer for the synthetic data.

Calculate the QNOAEL of Ellagic Acid for BSN476 Drug Development

Chikungunya is a rapidly spreading mosquito-borne disease that now infects over 3 million people worldwide¹². BSN476, a drug for treating chikungunya infections, contains in part EA. QB will estimate a safe level of EA for the first-in-human study in order to develop a treatment for Chikungunya. An EA toxicity meta-analysis using food consumption will be completed as part of the Investigational New Drug (IND) application to the FDA. Studies will be selected from the literature and analyzed according to the Cochrane protocols, and the QNOAEL of EA will be calculated along with the NOAEL and BMD. These results will serve as the basis for the first-in-human study of EA.

Strategy

Significance

Method Significance

The NOAEL depends strongly on the dose selection, dose spacing, and sample size of a single study from which the critical effect has been identified. The primary goal of BMD modeling is to define a point of departure that is largely independent of study design. But while the BMD effectively enables multiple studies to be pooled to increase accuracy, it does not handle studies with conflicting results gracefully, as will be seen below³.

	NOAEL	Benchmark Dose	QNOAEL
Advantages	<ul style="list-style-type: none"> • Very simple, early method of deriving a POD that has been utilized for years • Defined as the highest dose which generates no effect in the studied population • Can be used when data will not work with BMD mode 	<ul style="list-style-type: none"> • Calculation is not limited to specific experimental doses, but can calculate a BMD within the tested range • Multiple studies may be pooled to increase accuracy • Allows for weighting of results based on study quality • Accounts for dose-response curve and pharmacokinetic modeling • Benchmark Dose is defined as the dose which generates a specific benchmark effect 	<p>The advantages of BMD, plus:</p> <ul style="list-style-type: none"> • Calculates statistical measures of certainty to gracefully handle conflicting input study results • Generates an intuitive correlation vs. dose curve that shows the statistical likelihood of effect at a given dose • Able to account for publication bias in the literature • Bootstrapping algorithm is robust and nonparametric (and thus able to solve problems in which the underlying distribution is unknown) • Can extrapolate to lower doses than dosing range studied/reported in the literature. • Computed quickly(an order of n1 algorithm)
Limitations	<ul style="list-style-type: none"> • Limited to doses studied/reported in the literature • Lowest Observed Average Effect Level (LOAEL) 	<ul style="list-style-type: none"> • More complicated and time-consuming process than simple NOAEL • Limited strongly 	<ul style="list-style-type: none"> • More complicated and time-consuming process than simple NOAEL • Limited

Table 1 Continued			
	<p>cannot be used to generate NOAEL</p> <ul style="list-style-type: none"> • Not designed for pooling of multiple studies • Does not take into account the statistical likelihood of effect across a large population • Does not represent the population as a whole; is limited to the population studied 	<p>by the quality of information and the inclusion criteria chosen by the researcher</p> <ul style="list-style-type: none"> • Unable to account well for conflicting study results • Generates a non-intuitive mean response x dose line which is ambiguous for doses which sometimes cause an effect, or which cause an effect in a certain percentage of the population • Results do not well indicate the response of a large population to a drug • Parametric test. • “Effect” vs. “No Effect” classification can be difficult when studies examined different endpoints; ex: effect on heart rate vs. mood alteration 	<p>somewhat by the quality of information and the study inclusion criteria chosen by the researcher</p>

The BMDL (the statistical lower confidence limit on the benchmark dose, or BMD) is used as the point of departure (POD) for most non-cancer and cancer risk estimates derived by the U.S. EPA. The initial step in the risk assessment process is hazard identification, which is defined as the identification of effects on health noted as the result of exposure to a specified chemical. Hazard identification is followed by determination of the critical effect on which to construct NOAELs (No Adverse Effect Levels) or BMDs and BMDLs.

Both the NOAEL and BMD approach require some common considerations of the general quality of a particular study. A few of these considerations include:

- a. Sample Size. Were the sample sizes used large enough to properly detect treatment effects?
- b. Exposure. Were the exposure durations adequate? Were relevant routes of exposure employed in the study?
- c. Endpoints. Did the study measure endpoints of interest?
- d. Quality. Did the study employ standard quality control procedures like good laboratory practice (GLP)?

In addition to these common data quality considerations that affect both the NOAEL and BMD estimates, there are added BMD-specific points to consider in the identification of datasets that are appropriate for BMD modeling. For example, when sample size decreases, which results in decreased power to detect treatment effects, the NOAEL procedure produces POD estimates while the BMD approach produces lower (extra precautionary) PODs. To maintain consistency and reproducibility, most scientists employ a six-step process for BMD analysis. The six steps involved in the BMD analysis are (1) choice of a BMR, (2) selecting a set of models, (3) assessing model fit, (4) model selection when BMDLs are divergent, (5) model selection when BMDLs are not divergent, and (6) data reporting.

The new QB nonparametric meta-analysis of multiple studies so far appears to be superior to BMD modeling. Unlike BMD, the QNOAEL estimate is not limited by the format of the data presented. The QNOAEL is no more time-consuming to calculate than the BMD, and provides a simpler decision-making process. For example, the graphs below show the BMD and QB simulations for THC in hemp seed. Note that the same clinical studies were used for both analyses. QB not only more clearly demonstrates the trend of data, but also produces a correlation curve which is intuitively noncontradictory. (It is to be expected that different studies may reach contradictory results on the effects of a given dosage, as study methods and populations vary.)

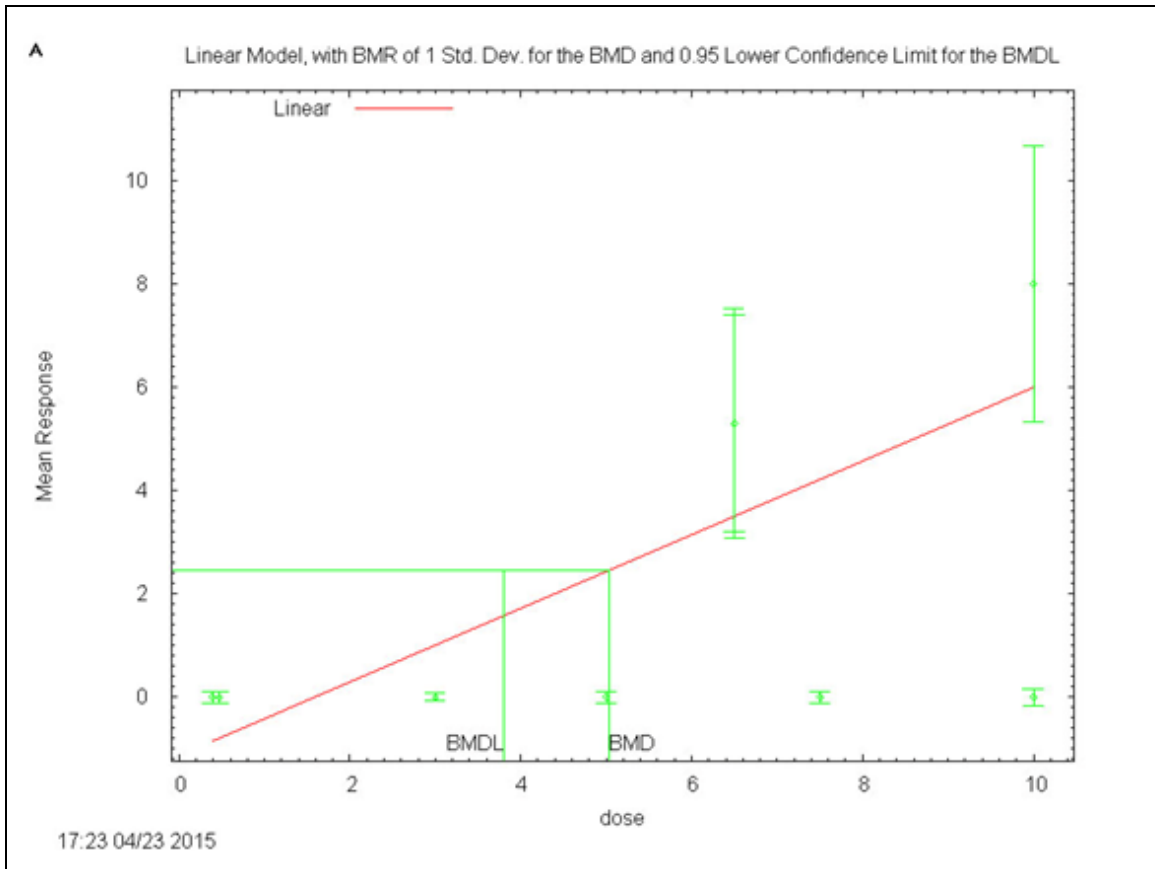


Figure 9: Linear BMD Model for THC in Hemp Seed

The BMD method attempts to fit a model to studies at different THC doses (green circles) that sometimes show an effect, and sometimes do not. This fitting process can seem to make little sense considering the data. Note that two different studies at 10 mg THC show opposite effects. (Dosages falling on the 0 line of “Mean Response” are no-effect dosages in this figure.)

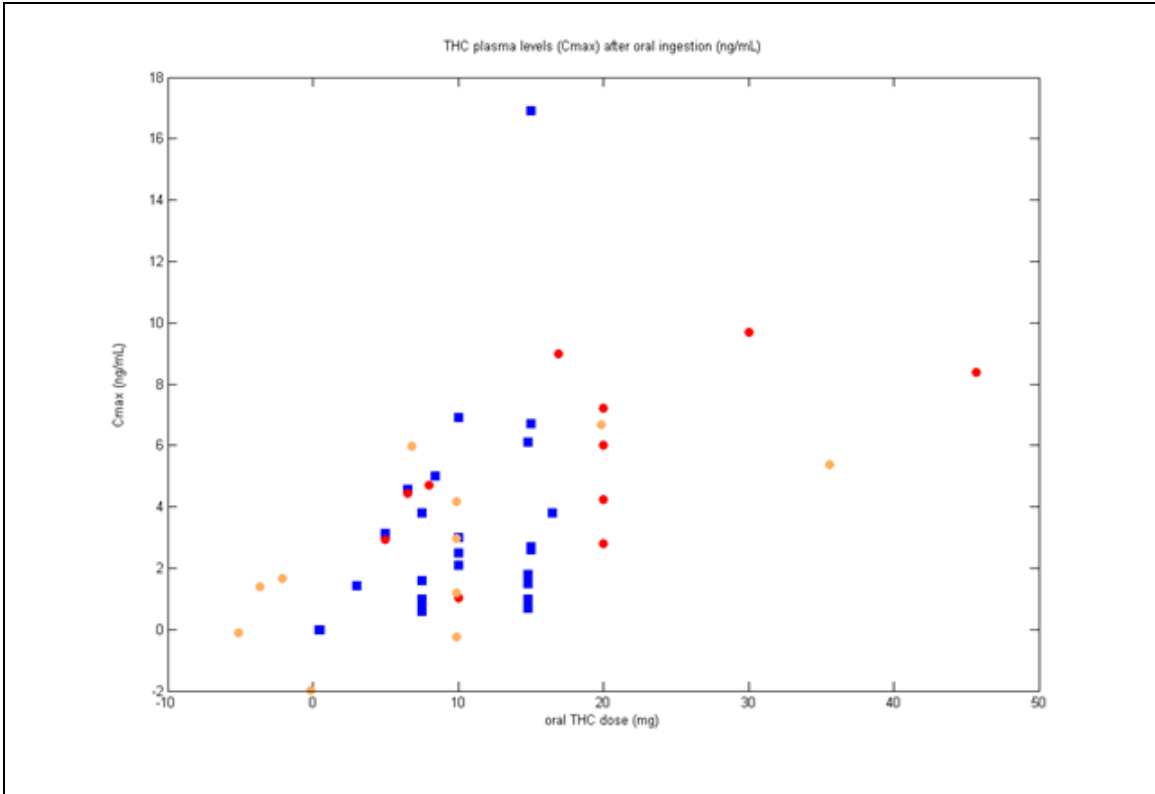


Figure 10: THC Doses Showing Effect or No Effect Graphed in Multivariate Space

On the other hand, in multivariate space the doses show no discontinuities. Red circles are test set study results (showed effect on heart rate HR and blood pressure BP). Blue squares are training set study results (no effect on HR or BP, sometimes an effect on Cmax). Orange circles are test set after recentering on the training set.

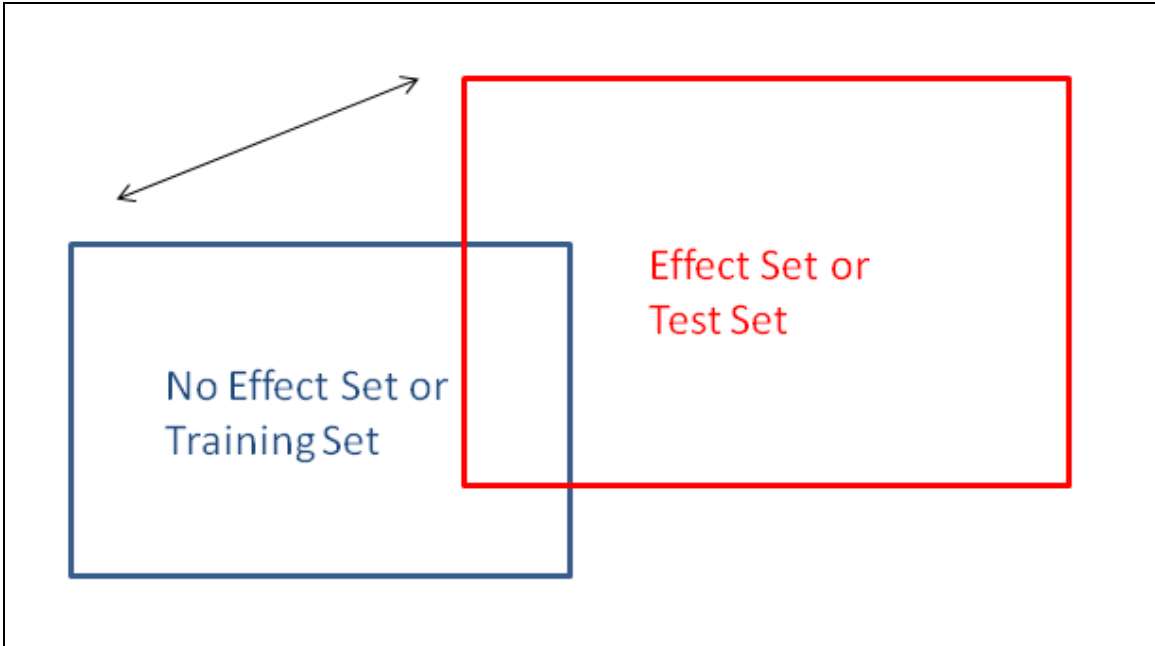


Figure 11: Determining 98% Confidence Levels

In the new nonparametric method, 98% confidence limits are set on the No Effect Set (or Training Set) by correlating integrals of bootstrap replicates of the training set with themselves. The simulation then projects the Effect Set (or Test Set) into the same space as the No Effect Set, and then translates the Effect Set toward/away from the center of the No Effect Set to determine when the correlations between the No Effect and Effect replicate integrals become significantly different.

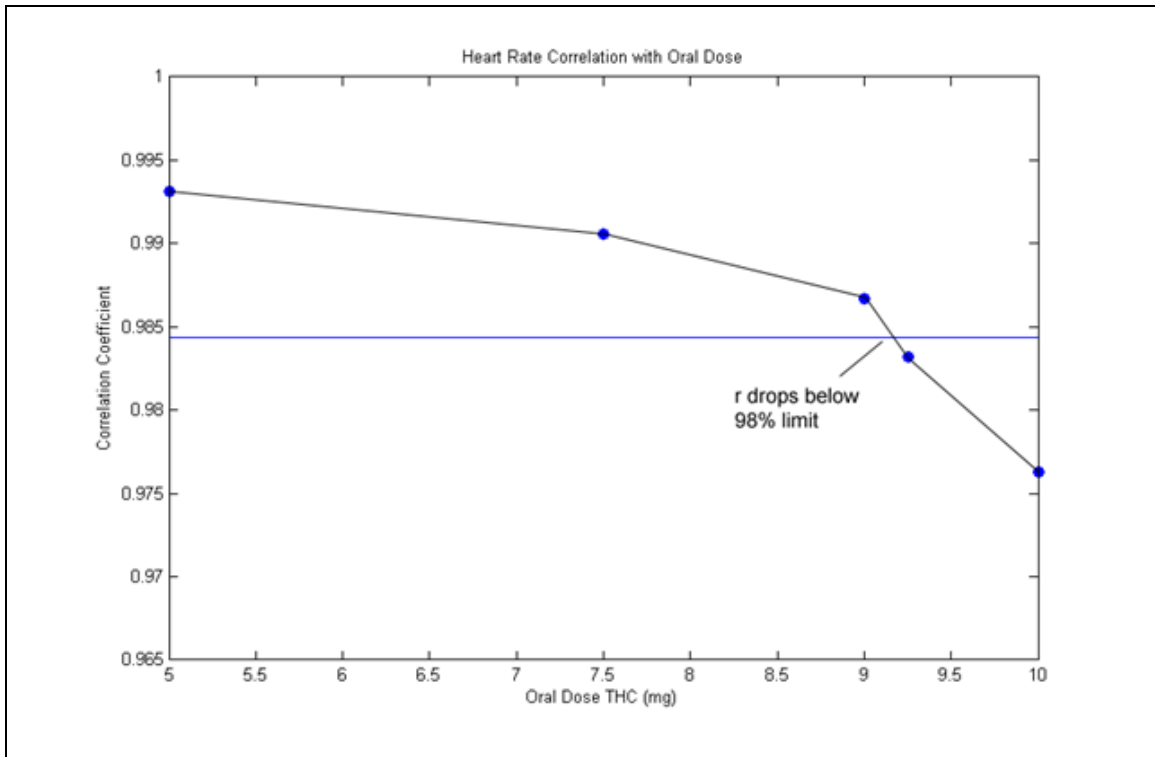


Figure 12: Determining the QNOEL of THC

The gradual approach of the line to the 98% confidence limit reveals that the scale of the test set is larger than the scale of the training set. The horizontal blue line represents the 98% confidence limit on the training set (the no effect on HR/BP set). The 98% confidence limit on the training set was 9.1 mg orally (which is the NOEL for heart rate) with a $C_{max} = 2.8$ ng/ml.

Application Significance

Chikungunya is a rapidly spreading mosquito-borne disease that now infects over 3 million people worldwide¹². The disease originated in Africa around 1700 A.D., and until recent years, reported infections were limited to the African continent and Southeast Asia¹. The disease was first identified in 1952 during an outbreak so serious that infections were clinically indistinguishable from dengue fever¹⁹. Throughout the 1960s and 1970's, outbreaks were reported in Southeast Asia¹⁹. After decades without another Southeast Asian outbreak, a 1999 outbreak in Indonesia led to a massive outbreak reported in India in 2006, the strain responsible for this resurgence bearing 99% similarity to the strain responsible for a 1989 outbreak in Uganda^{3,19}. In December, 2013, the disease made its debut in the Americas, with its first local transmission occurring on the island of St. Martin; local transmission in French Guiana on the South American continent occurred later that month^{15,17,18}. After only two years, local transmission had been documented in 19 Caribbean countries, including Puerto Rico, as well as in nearly

every country on the South American continent¹⁵⁻¹⁸. The WHO has currently issued a level-1 watch for travelers visiting South America and the Caribbean and expects Chikungunya to spread^{15,18}.

BSN476 contains in part EA. EA is a polyphenolic compound with antiproliferative and antiviral properties⁶. EA at 10 uM produces 99.6% inhibition of Chikungunya virus in vitro⁶. EA is found in a number of plant extracts, usually in the form of hydrolyzable ellagitannins which are complex esters of EA with glucose^{5,14}. Ellagitannins are broken down in the intestine to eventually release EA^{5,14}. To develop BSN476 as a treatment for Chikungunya, the PK of the drug must be studied in a first-in-human (FIH) trial, and a safe range of exposure must be determined for that trial.

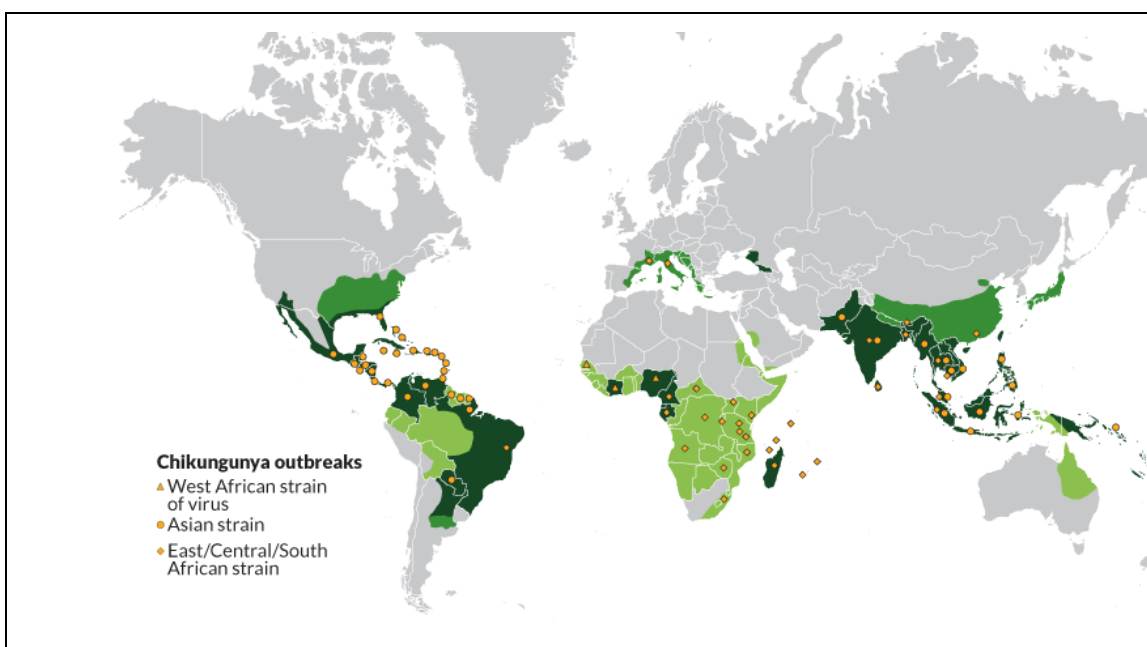


Figure 13: Map of Chikungunya Outbreaks

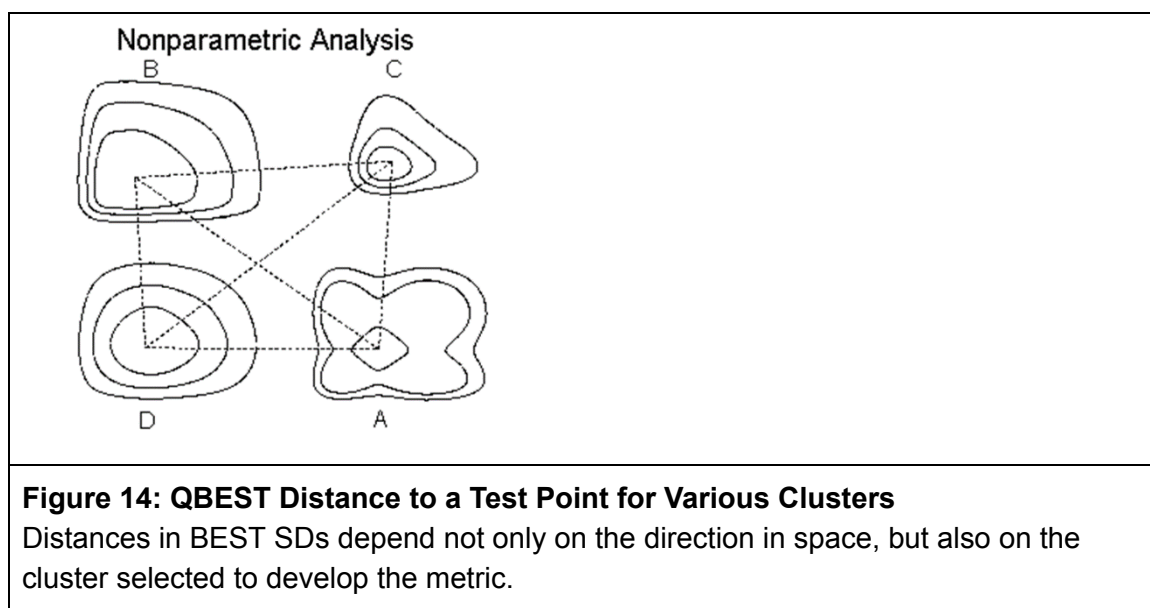
Yellow dots mark the locations of chikungunya outbreaks, while the green areas mark the ranges of the mosquitoes able to propagate the outbreaks. This map was shows data only from June, 2015 and earlier.¹²

QB is applied within the Cochrane framework for meta-analysis (the Cochrane framework provides a “Garbage-In-Garbage-Out” standard for data inputs - generally clinical studies) to determine doses for the first in human study². The QNOAEL of EA will be estimated from previously published food consumption studies. Bootstrap replication and manipulation of data clusters will reveal the QNOAEL of EA with 98% confidence. This project will use the QNOAEL to estimate a safe range of exposure to EA for the FIH study on the way to developing a treatment for Chikungunya. QB is a robust O(n)

algorithm that is designed for massively parallel computers, and is a very powerful meta-analysis tool (most algorithms use matrix factorization and are $O(n^3)$ in execution time). The QB algorithm will be translated from MATLAB into Python to make it more accessible to the scientific community. Very large sample datasets will be used to test the memory usage and reproducibility of the results of the algorithm, as well as minimum parameters for its usage. These data have permitted estimation of the maximum analysis capability of cloud computing services and the National Science Foundation XSEDE supercomputer (Comet).

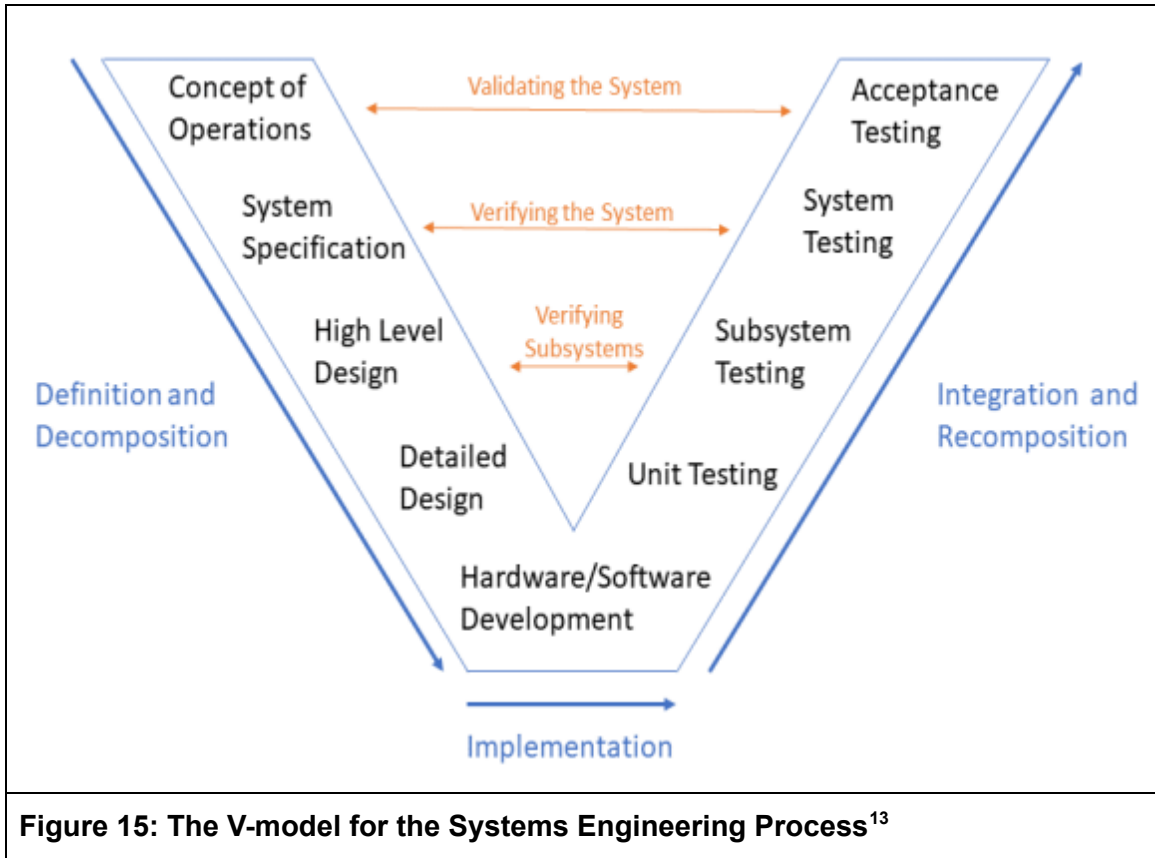
Innovation

This proposed study utilizes a Quantile Bootstrap statistical method designed for Big Data problems, SOB, inside a new simulation to estimate the QNOAEL for a drug with 98% confidence from a set of small studies⁷⁻¹⁰. SOB is a form of cluster analysis, which is a common analytical technique for determining chemical identity and purity⁷⁻¹⁰. So far, the QNOAEL appears to be superior to the NOAEL and the BMD.



QB is applied within the Cochrane framework for meta-analysis. QB works by analyzing clusters of studies that found an effect, and clusters of studies that found no effect (the studies can use different dose levels). By analyzing the quantiles of each cluster and adjusting for cluster skew, QB can measure the distance between clusters in probability space, until it finds the dose that yields no adverse effects for the entire human population at a specified level of statistical significance (see Figure 21, 98% level set by default).

Approach



This is a software development project undertaken as part of a larger drug development project. Good Engineering Practice, Standard Operating Procedures (SOPs) and working practice guidelines have been implemented for project design as well as execution (see Figure 22). A robust change control system must be implemented in this project.

The Design SOPs and Configuration Management system will be applied to the system designed to resolve each specific aim in this project. Each specific aim will begin with a Needs Analysis to determine what the new system needs to be able to do. A requirements analysis will also be conducted to determine what is required to fill those needs. A System Requirements Review (SRR) will demonstrate understanding of the requirements documents (scope, specifications, schedule, validation plans, and budget). SRR will determine the initial design direction and describe preliminary data and progress, and how these will converge to an optimum and complete system configuration for the specific aim. The memory needed to run QB on large datasets and evaluate the performance of the algorithm must be quantified. (Very large sample

datasets to test the memory usage and reproducibility of the results of the QB algorithm are being created, as well as to set ranges of parameters for its usage. Parameters include the number of bootstrap replications desired (b), the number of variables in the multivariate analysis (d), and the size of the dataset used (n). Preliminary data indicate that the typical laptop computer can process over one million bootstrap replication samples, independent variables, or sample data points when the other two parameters are minimized, or over one-thousand bootstrap replications, variables, and data points [all maximized at approximately 1000 inputs]).

As the system evolves through the development process, topic experts will be invited to later design reviews (especially CDR, TRR, and MRR). System Design Review (SDR) acts as a control gate that reviews and approves the top-level system design solution and rationale¹³. It is the decision point to proceed with system specification flow down to individual physical and process configuration items¹³. System limitations will be refined at SDR.

Using the run-time and memory usage data determined in Specific Aim 1, the performance capabilities of the algorithm on an NSF supercomputer (XSEDE Comet) will be calculated. QB is capable of tackling immensely large datasets, and the computing capabilities of the algorithm will be stretched on a massively parallel machine to demonstrate proof-of-concept. An SDR report will be added to the Design History file for FDA.

A Preliminary Design Review (PDR) will be performed on each configuration item or group of configuration items to: (1) Evaluate the progress, technical acceptability, and risk resolution, (2) Measure its harmony with performance and engineering specialty requirements of the Configuration Item development specification, (3) Evaluate the extent of definition and evaluate the technical risk connected with the selected methods/processes, and (4) Demonstrate the existence and compatibility of the physical and functional interfaces among the configuration item and other items of equipment, facilities, computer software, and personnel¹³. Topic experts will also be invited to the review. A Blue team and a Red team are used for design and validation, respectively. A PDR report will be added to the Design History file for FDA in the annual reporting system.

Critical Design Review (CDR) is the last design review conducted before an action is taken that is irreversible. (1) CDR is a review to establish that detail design of the configuration item under review meets cost, schedule, and performance requirements. (2) CDR will establish detail design compatibility among the configuration item and other items of equipment, facilities, computer software and personnel. (3) CDR will gauge configuration item risk areas (on a technical performance, cost, and schedule basis). Topic experts will again be invited to the review. A Blue team and a Red team are used

for design and validation, respectively. A CDR report will be added to the Design History file for FDA in the annual reporting system.

Deployment Readiness Review (DRR) is held to confirm readiness for deployment. This review is conducted to ensure that all deficiencies are corrected before actual use. The complete system is challenged every feasible way (conceptually, physically, cyber-, etc.). The DRR demands the review and analysis of all subsystem/unit level testing preceding the formal acceptance tests. Topic experts will be invited to the review. A Blue team and a Red team are used for design and validation, respectively. A DRR report will be added to the Design History file for FDA.

The QB algorithm will be translated into Python to make it more accessible to the scientific community. Once QB is available in Python it will run on Amazon Web Services, Microsoft Azure, and Google Compute Engine as well as the NSF XSEDE Comet supercomputer currently being used. The REPLICICA algorithm has already been translated from Matlab into Python to make it more accessible to the scientific community. However, the algorithm on which QB relies has yet to be translated and the entirety of the program remains to be validated.

Credit

The preliminary data and proposal described were supported in part by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant UL1TR001998. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. This project was also supported by NSF ACI-1053575 allocation number BIO170011.

Reference List

See “References” appendix.. References have been renumbered compared to those published in the original paper.

Appendix

Programs “REPLICICA” and “SOB” published in the original paper are included in the Appendix.

CHAPTER 3: ESTABLISHING EDI FOR A CLINICAL TRIAL OF A TREATMENT FOR CHIKUNGUNYA

Purpose: Establishing the EDI of ellagic acid will assist in the determination of a safe starting dose of ellagic acid for clinical trials in BSN476.

Publication and Copyright Information:

Dickerson C., Ensor M., Lodder R.A. (2018) Establishing EDI for a Clinical Trial of a Treatment for Chikungunya. In: Shi Y. et al. (eds) Computational Science – ICCS 2018. ICCS 2018. Lecture Notes in Computer Science, vol 10861. Springer, Cham
© Springer International Publishing AG, part of Springer Nature 2018

Reproduced with permission. License for reproduction can be found in the Appendix.

Establishing EDI for a Clinical Trial of a Treatment for Chikungunya

Abstract

Ellagic acid (EA) is a polyphenolic compound with antiviral activity against chikungunya, a rapidly spreading new tropical disease transmitted to humans by mosquitoes and now affecting millions worldwide. The most common symptoms of chikungunya virus infection are fever and joint pain. Other manifestations of infection can include encephalitis and an arthritic joint swelling with pain that may persist for months or years after the initial infection. The disease has recently spread to the U.S.A., with locally-transmitted cases of chikungunya virus reported in Florida. There is no approved vaccine to prevent or medicine to treat chikungunya virus infections. In this study, the Estimated Daily Intake (EDI) of EA from the food supply established using the National Health and Nutrition Examination Survey (NHANES) is used to set a maximum dose of an EA formulation for a high priority clinical trial.

Keywords: Tropical disease NHANES Drug development

1 Introduction

1.1 Compound

Ellagic acid (EA) is a polyphenolic compound with health benefits including antioxidant, anti-inflammatory, anti-proliferative, athero-protective, anti-hepatotoxic and anti-viral properties^{1,2}. EA is found in many plant extracts, fruits and nuts, usually in the form of hydrolyzable ellagitannins that are complex esters of EA with glucose. Natural sources high in ellagitannins include a variety of plant extracts including green tea, nuts such as walnuts, pecans and almonds, and fruits, particularly berries, such as blackberries, raspberries and strawberries, as well as grapes and pomegranates.

1.2 Chikungunya

Chikungunya virus is transmitted to humans by mosquitoes. Typical symptoms of chikungunya virus infection are fever and joint pain. Other manifestations may include headache, encephalitis, muscle pain, rash, and an arthritis-like joint swelling with pain that may persist for months or years after the initial infection. The word 'chikungunya' is thought to be derived from its description in the Makonde language, meaning "that which bends up" the deformed posture of people with the severe joint pain and arthritic symptoms associated with this disease³. There is no vaccine to prevent or medicine to treat chikungunya virus infections.

Millions of people worldwide suffer from chikungunya infections. The disease spreads quickly once it is established in an area. Outbreaks of chikungunya have occurred in

countries in Africa, Asia, Europe, and the Indian and Pacific Oceans. Before 2006, chikungunya virus disease was only rarely pinpointed in U.S. travelers. In 2006–2013, studies found a mean of 28 people per year in the United States with positive tests for recent chikungunya infection. All of these people were travelers visiting or returning to the United States from affected areas in Asia, Africa, or the Indian Ocean.

In late 2013, the first local transmission of chikungunya virus in the Americas was identified on the island of St. Martin, and since then all of the other Caribbean countries and territories. (Local transmission means that mosquitoes in the area have been infected with the virus and are spreading it to people.)

Beginning in 2014, chikungunya virus disease cases were reported among U.S. travelers returning from affected areas in the Americas and local transmission was identified in Florida, Puerto Rico, and the U.S. Virgin Islands. In 2014, there were 11 locally-transmitted cases of chikungunya virus in the U.S. All were reported in Florida. There were 2,781 travel-associated cases reported in the U.S. The first locally acquired cases of chikungunya were reported in Florida on July 17, 2014. These cases represent the first time that mosquitoes in the continental United States are thought to have spread the virus to non-travelers. Unfortunately, this new disease seems certain to spread quickly. Data Driven Computational Science (DDCS) offers ways to accelerate drug development in response to the spread of this disease.

EA has been shown to be an inhibitor of chikungunya virus replication in high throughput screening of small molecules for chikungunya⁴. In screening a natural products library of 502 compounds from Enzo Life Sciences, EA at 10 μ M produced 99.6% inhibition of chikungunya in an in vitro assay.

1.3 Metabolism

Ellagitannins are broken down in the intestine to eventually release EA. The bioavailability of ellagitannins and EA have been shown to be low in both humans and in animal models, likely because the compounds are hydrophobic and they are metabolized by gut microorganisms⁵⁻⁸. The amount of ellagitannins and EA reaching the systemic circulation and peripheral tissues after ingestion is small to none⁷. It is established that ellagitannins are not absorbed while there is high variability in EA and EA metabolites found in human plasma after ingestion of standardized amounts of ellagitannins and EA⁹⁻¹¹. These studies indicate that small amounts of EA are absorbed and detectable in plasma with a C_{max} of approximately 100 nM (using standardized doses) and a T_{max} of 1 h^{9,10}. EA is metabolized to glucuronides and methyl-glucuronide derivatives in the plasma. The most common metabolite found in urine and plasma is EA dimethyl ether glucuronide¹².

It appears that the majority of ingested ellagitannins and EA are metabolized by the gut microbiota into a variety of urolithins. Urolithins are dibenzopyran-6-one derivatives that are produced from EA through the loss of one of the two lactones present in EA and then by successive removal of hydroxyl groups. Urolithin D is produced first, followed sequentially by urolithin C, urolithin A, and urolithin B. Urolithins appear in the circulatory system almost exclusively as glucuronide, sulfate and methylated forms as a result of phase II metabolism after absorption in the colon and passage through the liver¹³. While the amount of EA in the circulation is in the nanomolar range, urolithins and their glucuronide and sulfate conjugates circulate at concentrations in the range of 0.2–20 nM¹⁴. In light of the much larger concentrations of urolithins in the circulation compared to EA, it must be considered that the reported in vivo health effects of ellagitannin and EA may be largely due to the gut-produced urolithins. Growing evidence, mostly in vitro, supports the idea that urolithins have many of the same effects as EA in vitro. Various studies have shown evidence of anti-inflammatory¹⁵⁻¹⁷, anticarcinogenic¹⁸⁻²¹, anti-glycative²², possibly antioxidant^{6, 23}, and antimicrobial²⁴ effects of urolithins.

There is variation in how people metabolize EA into the various urolithins²⁵⁻²⁷. This is not surprising in light of the known differences between individuals in intestinal microbiotic composition. Tomás-Barberán²⁶ evaluated the urinary urolithin profiles of healthy volunteers after consuming walnuts and pomegranate extracts. They found that, consistent with previous findings, that urolithin A was the main metabolite produced in humans. However, they noted that the subjects could be divided into three groups based on their urinary profiles of urolithins. One group excreted only urolithin A metabolites while a second group excreted urolithin A and isourolithin A in addition to urolithin B. The third group had undetectable levels of urolithins in their urine. These results suggest that people will benefit differently from eating ellagitannin rich foods.

1.4 Use of EDI

Knowledge of the Estimated Daily Intake (EDI) can permit pharmacokinetic and formulation studies to be conducted without prior expensive and time-consuming toxicology studies, especially when the molecule is naturally present in the food supply (see Figure 16). A subject's dietary level of the compound would normally vary around the EDI. A subject is brought in to the drug evaluation unit, and after the usual ICH E6 procedures and informed consent, is "washed out" of any of the compound might be present from previous food consumption. Typically, washout is accomplished by maintaining the subject on a diet containing none of the compound to be investigated for a period of five or more half-lives. The subject then receives a dose of the compound and blood samples are collected for pharmacokinetic or other analysis. The concentration of the dose is calculated to keep the subject's exposure below the EDI. For this reason, it is important to establish the EDI before the clinical trial is designed and executed. After sufficient samples have been collected, the subject is released and the trial is complete for that subject. The subject then returns to a normal diet and levels

increase again to levels similar to those before the study.

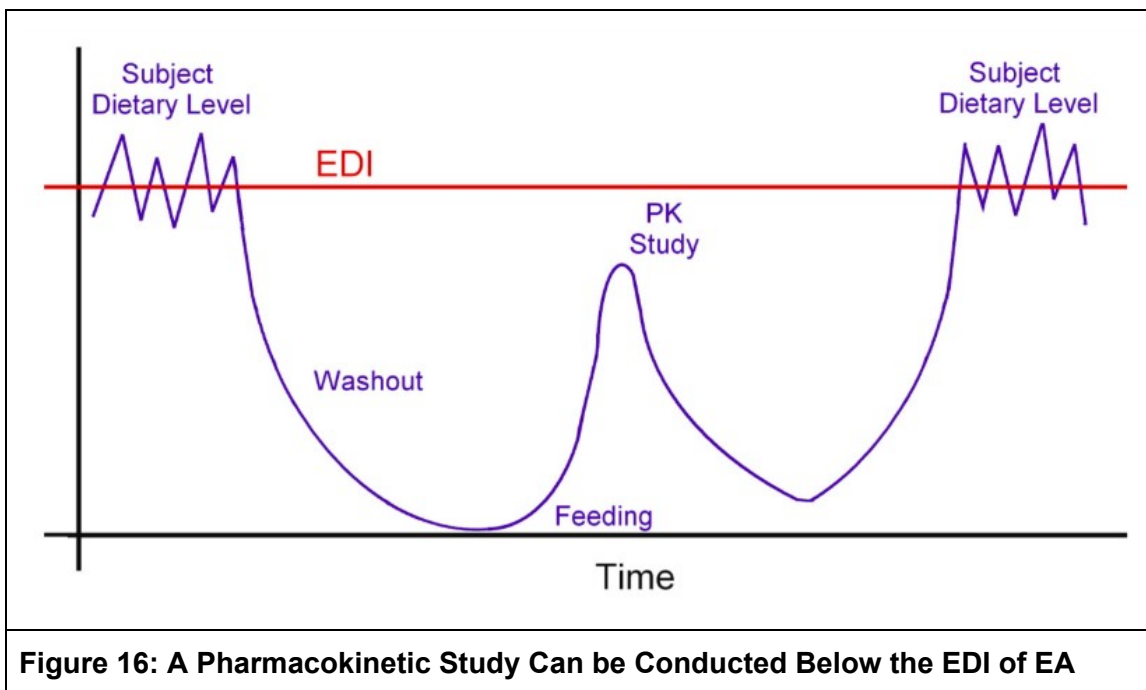


Figure 16: A Pharmacokinetic Study Can be Conducted Below the EDI of EA

2 Assessment of EA Use

An assessment of the consumption of EA (EA) by the U.S. population resulting from the approved uses of EA was conducted. Estimates for the intake of EA were based on the approved food uses and maximum use level in conjunction with food consumption data included in the National Center for Health Statistics' (NCHS) 2009–2010, 2011–2012, and 2013–2014 National Health and Nutrition Examination Surveys (NHANES) [27–29]. Calculations for the mean and 90th percentile intakes were performed for representative approved food uses of EA combined. The intakes were reported for these seven population groups:

1. infants, age 0 to 1 year
2. toddlers, age 1 to 2 years
3. children, ages 2 to 5 years
4. children, ages 6 to 12 years
5. teenagers, ages 13 to 19 years
6. adults, ages 20 years and up
7. total population (all age groups combined, excluding ages 0–2 years).

3 Food Consumption Survey Data

3.1 Survey Description

The most recent National Health and Nutrition Examination Surveys (NHANES) for the years 2013–2014 are available for public use. NHANES are conducted as a continuous, annual survey, and are released in 2-year cycles. In each cycle, approximately 10,000 people across the U.S. complete the health examination component of the survey. Any combination of consecutive years of data collection is a nationally representative sample of the U.S. population. It is well established that the length of a dietary survey affects the estimated consumption of individual users and that short-term surveys, such as the typical 1-day dietary survey, overestimate consumption over longer time periods³¹. Because two 24-h dietary recalls administered on 2 non-consecutive days (Day 1 and Day 2) are available from the NHANES 2003–2004 and 2013–2014 surveys, these data were used to generate estimates for the current intake analysis.

The NHANES provide the most appropriate data for evaluating food-use and food-consumption patterns in the United States, containing 2 years of data on individuals selected via stratified multistage probability sample of civilian non-institutionalized population of the U.S. NHANES survey data were collected from individuals and households via 24-h dietary recalls administered on 2 non-consecutive days (Day 1 and Day 2) throughout all 4 seasons of the year. Day 1 data were collected in-person in the Mobile Examination Center (MEC), and Day 2 data were collected by telephone in the following 3 to 10 days, on different days of the week, to achieve the desired degree of statistical independence. The data were collected by first selecting Primary Sampling Units (PSUs), which were counties throughout the U.S. Small counties were combined to attain a minimum population size. These PSUs were segmented and households were chosen within each segment. One or more participants within a household were interviewed. Fifteen PSUs are visited each year. For example, in the 2009–2010 NHANES, there were 13,272 persons selected; of these 10,253 were considered respondents to the MEC examination and data collection. 9754 of the MEC respondents provided complete dietary intakes for Day 1 and of those providing the Day 1 data, 8,405 provided complete dietary intakes for Day 2. The release data does not necessarily include all the questions asked in a section. Data items may have been removed due to confidentiality, quality, or other considerations. For this reason, it is possible that a dataset does not completely match all the questions asked in a questionnaire section. Each data file has been edited to include only those sample persons eligible for that particular section or component, so the numbers vary.

In addition to collecting information on the types and quantities of foods being consumed, the NHANES surveys collected socioeconomic, physiological, and

demographic information from individual participants in the survey, such as sex, age, height and weight, and other variables useful in characterizing consumption. The inclusion of this information allows for further assessment of food intake based on consumption by specific population groups of interest within the total population.

Sample weights were incorporated with NHANES surveys to compensate for the potential under-representation of intakes from specific population groups as a result of sample variability due to survey design, differential non-response rates, or other factors, such as deficiencies in the sampling frame^{29, 30}.

3.2 Methods

Consumption data from individual dietary records, detailing food items ingested by each survey participant, were collated by computer in Matlab and used to generate estimates for the intake of EA by the U.S. population. Estimates for the daily intake of EA represent projected 2-day averages for each individual from Day 1 and Day 2 of NHANES data; these average amounts comprised the distribution from which mean and percentile intake estimates were produced. Mean and percentile estimates were generated incorporating sample weights in order to provide representative intakes for the entire U.S. population. “All-user” intake refers to the estimated intake of EA by those individuals consuming food products containing EA. Individuals were considered users if they consumed 1 or more food products containing EA on either Day 1 or Day 2 of the survey.

3.3 Food Data

Food codes representative of each approved use were chosen from the Food and Nutrition Database for Dietary Studies (FNDDS) for the corresponding biennial NHANES survey. In FNDDS, the primary (usually generic) description of a given food is assigned a unique 8-digit food code^{29, 30}.

3.4 Food Survey Results

The estimated “all-user” total intakes of EA from all approved food uses of EA in the U.S. by population group is summarized in Figures 17, 18, 19 and 20.

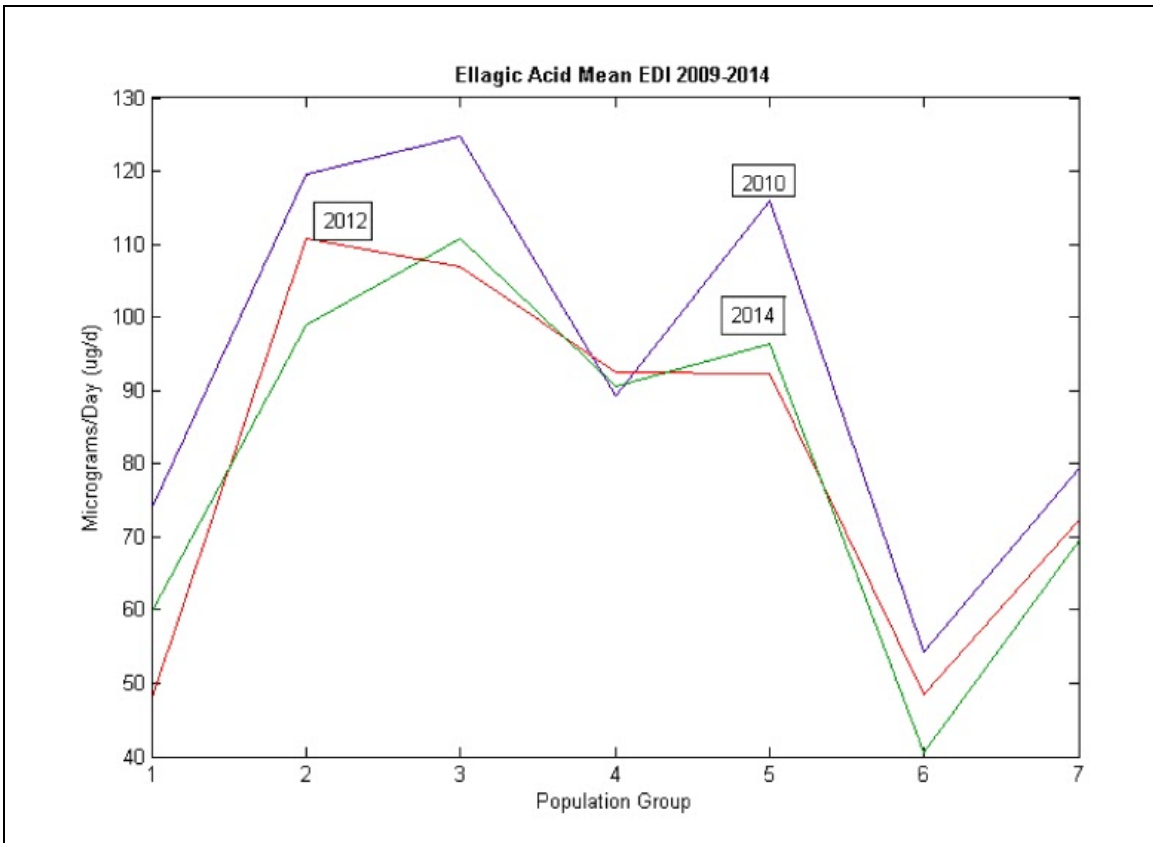


Figure 17: Ellagic Acid Mean EDI

Children consume more EA on average than adults. Baby foods are often made from ingredients high in EA. The blue line shows data from the 2009–2010 NHANES, the red line data from the 2011–2012 NHANES, and the green line data from the 2013–2014 NHANES.

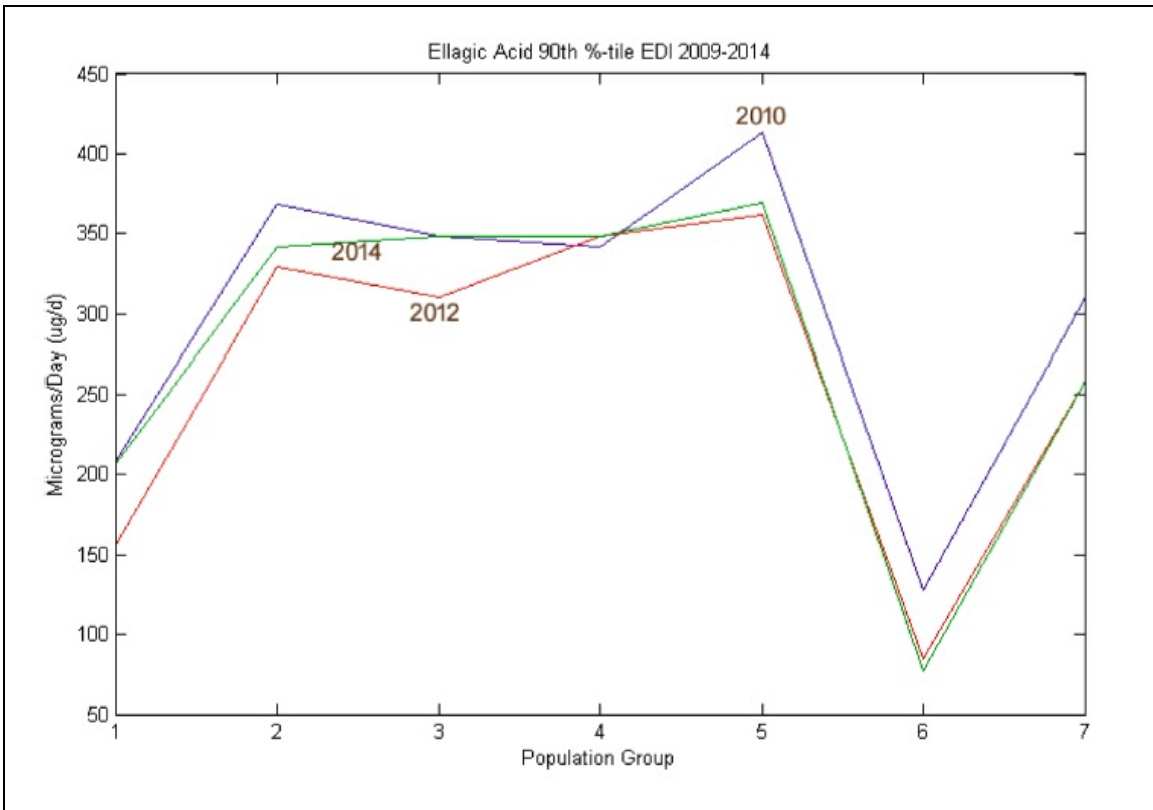


Figure 18: Ellagic Acid 90th %-tile EDI

Teenagers contribute the highest peak in the 90th percentile consumers of EA. The blue line shows data from the 2009–2010 NHANES, the red line data from the 2011–2012 NHANES, and the green line data from the 2013–2014 NHANES.

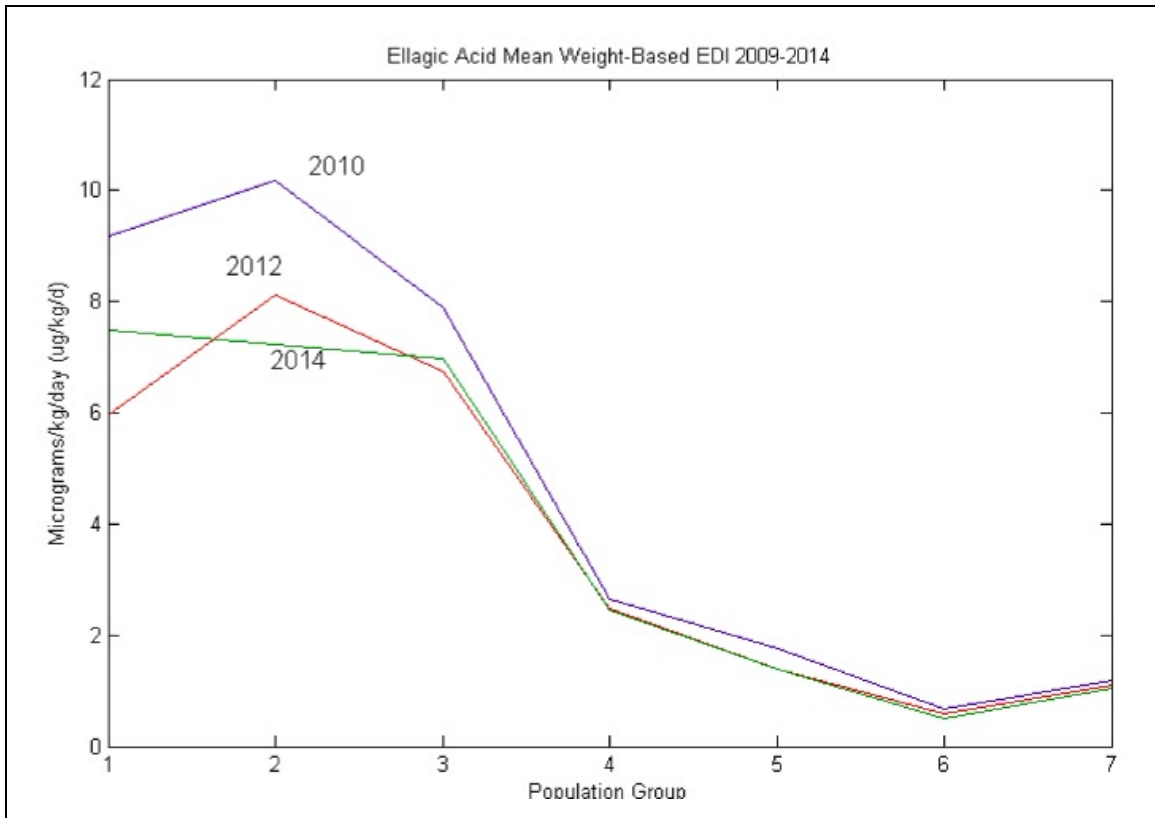


Figure 19: Ellagic Acid Mean Weight-Based EDI When EA exposure is calculated on a per kilogram of body weight basis, toddlers aged 1 to 2 years are exposed to the most EA on average. The blue line shows data from the 2009–2010 NHANES, the red line data from the 2011–2012 NHANES, and the green line data from the 2013–2014 NHANES.

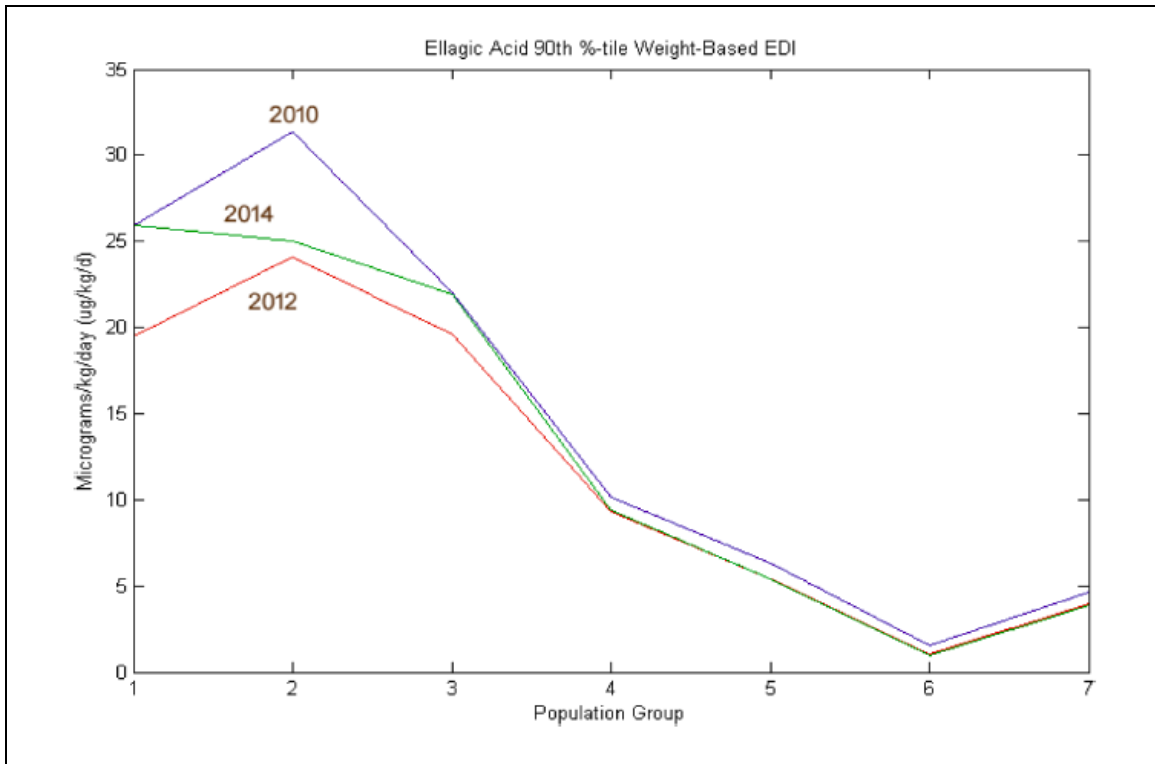


Figure 20: Ellagic Acid 90th %-tile Weight-Based EDI

When EA exposure is calculated on a per kilogram of body weight basis for the 90th percentile consumers, toddlers aged 1 to 2 years are again exposed to the most EA. The blue line shows data from the 2009–2010 NHANES, the red line data from the 2011–2012 NHANES, and the green line data from the 2013–2014 NHANES.

The estimated “all-user” total intakes of EA from all approved food uses of EA in the U.S. by population group are graphed using NHANES data in Figs. 17, 18, 19 and 20 for 2009–2010, 2011–2012, and 2013–2014. The figures show that over 6 years, the consumption of EA has been fairly constant and that children and teenagers are the major consumers.

4 Conclusions

In summary, 28.3% of the total U.S. population of 2+ years was identified as consumers of EA from the approved food uses in the 2013–2014 survey. The mean intakes of EA by all EA consumers age 2+ (“all-user”) from all approved food uses were estimated to be 69.58 µg/person/day or 1.05 µg/kg body weight/day. The heavy consumer (90th percentile all-user) intakes of EA from all approved food-uses were estimated to be 258.33 µg/person/day or 3.89 µg/kg body weight/day. The EDI (red line in Figure 16) is set at 70 µg/person/day from the 2013-2014 NHANES for consumers ages 2 and up. The next experiment will be an actual trial of EA in human subjects at the EDI with a dose of 3.89 µg/kg body weight/day (see Figure 16), as determined by this DDCCS study.

Support

The project described was supported in part by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant UL1TR001998. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. This project was also supported by NSF ACI-1053575 allocation number BIO170011.

References

See References Section for the original references of this paper.

CHAPTER 4: CALCULATING THE ACCEPTABLE DAILY INTAKE OR NOAEL WITH QBEST (QNOAEL)

Purpose: The acceptable daily intake (QNOAEL) of ellagic acid is calculated using an analysis of animal studies. This is proof-of-concept that QBEST can be used to generate NOAELs and NOELs in a novel fashion.

Calculating the Acceptable Daily Intake or NOEL with QBEST (QNOEL)

Introduction:

Current Best Practice

Before a novel therapeutic agent can progress from animal studies to first-in-human studies, the FDA requires evidence that the investigational new drug will be safe for human use. Determining the No Observed Adverse Effect Limit (NOAEL), or the highest dose of a drug that can be administered to an animal without producing any undesirable effects, is important in this process because it places an upper bound on the dosage that should be administered to a human. In general, the maximum dose administered to humans is limited even further as a precaution in case the drug has unexpectedly greater toxicity in humans than in the animal model.

Traditionally, NOAELs have been determined by administering increasingly large doses of a drug compound to animal subjects until adverse effects become apparent. The highest dose administered during the course of the study that did not produce any adverse effect is designated the NOAEL. There are problems with this method, namely that the NOAEL must be determined to be a dose that was administered in the study, and often only 3 or 4 doses, spaced an order of magnitude or more apart, are used because of cost. For this reason, the NOAEL determined through traditional means often does not capture the true population NOAEL well. Traditional NOAEL determinations can be confounded in other ways, as well, and it is difficult to meta-analyze and pool NOAEL studies. This can lead to NOAEL determinations that are influenced strongly by sampling.

The Benchmark Dose Method is another way of estimating toxicity. The benchmark dose is defined as the statistically calculated lower 95 % confidence limit on the dose that produces a defined response (called the benchmark response or BMR, usually 5 % or 10 %) of an adverse effect compared to background, often defined as 0 % or 5 %. The benchmark dose describes a quantity that is in some ways very different from a NOAEL; the benchmark dose is an estimate of the quantity of therapeutic agent that will cause an adverse reaction, whereas NOAEL indicates the highest dose that is completely tolerable. The benchmark dose (BMD) can be calculated from a series of separate studies, unlike to NOAEL. However, the BMD can be confounded by conflicting studies. In any given population, it is expected that there will be some variation in response. When calculated with conflicting studies shown later in this dissertation, some of which found an effect to occur at a given dose while some found no effect at the same given dose, the Benchmark Dose Method generates a mean dose-response line through studies with opposite conclusions. Though the Benchmark Dose Method is in many

ways superior to the traditional determination of NOAEL, as it is able to meta-analyze multiple toxicity studies and can generate a BMD that was not a dose directly administered in a study, it is still not an ideal method.

Discriminant Cluster Analysis and QNOAEL

Discriminant cluster analysis is a method commonly utilized in chemometrics and analytical chemistry for determining whether a given sample belongs in the same class as a set of training samples. The QBEST equation is an application of this method to estimation of NOAELs.

At present, the Mahalanobis distance is the more commonly used metric for statistical multivariate discriminant cluster analysis. Though the premise of the Mahalanobis equation is basically the same as that of QBEST, the Mahalanobis is limited by a crucial requirement: In order to calculate a Mahalanobis distance, the algorithm requires that significantly more data points be included in the analysis than variables that describe those data points. In terms of meta-analysis of drug development data, this means that in order to use the Mahalanobis equation, one would need to have significantly more clinical trials or animal studies than clinical parameters measured in those studies. Clinical trials and animal studies are very expensive, and generate vast amounts of data. This means that while there may be a dozen or so relevant clinical studies for a given drug in the literature (which form rows in a matrix), perhaps upwards of a hundred separate measurements will have been collected for each patient (columns in the matrix). Moreover, the change in many biological parameters may be recorded over time. It is mathematically impossible to calculate a Mahalanobis distance using all of the parameters of the studies given so very few studies - the number of rows of the matrix must exceed the number of columns. This problem is a result of the matrix inversion / factorization method by which Mahalanobis distances are calculated.

QBEST is ideal for these circumstances, and can be applied to generate a QNOAEL which is in many ways superior to the traditional NOAEL or BMD. The method of calculation differs from the Mahalanobis equation, and in consequence the number of studies can be very low in comparison to the number of variables and the equation still maintains mathematical accuracy and precision. QBEST is able to maximally utilize the vast amounts of data generated by clinical studies. Meta-analyzing with QBEST can reduce the need to repeat costly clinical studies, as often occurs with traditional NOAEL determination. Like the Benchmark Dose Method, QBEST is capable of calculating a safety limit other than a dose that was administered. Unlike the Benchmark Dose Method, however, QBEST generates a QNOAEL that represents the dose which at a given confidence level will produce no effect in the population. QBEST handles conflicting studies in a more intuitively acceptable way than the Benchmark Dose Model, which will be seen below. Unlike the Benchmark Dose Method, the QBEST is a

nonparametric statistical test and can be more accurately applied when parametric assumptions (such as assumptions of normality) are not met.

Ellagic Acid

Ellagic acid (EA) is a naturally-occurring polyphenol with strong antiviral, anti-inflammatory, and anti-cancer properties. EA has been shown to significantly reduce growth rate and multiplicity of tumors in animals with various cancers. It also affects the clotting of blood and the lipid profile of animals with diabetes or models of heart disease. EA has been reported numerous times to have an attenuating effect on physiological changes induced by various chronic disease states. EA is also the principal component of an investigational new drug, BSN476, which will be moved into clinical trials as a treatment for Chikungunya. Toxicology tests must be conducted in at least one rodent and one nonrodent species prior to human exposure. (Often rats and beagles are used. Rats are usually the first species, and the rat results help select the range of doses for the beagle studies. The rodent and nonrodent results, in turn, help to select the range of doses for first-in-human studies.) Currently, no NOAEL or NOEL has been published in the literature for human consumption of EA, nor has a Generally Recognized as Safe (GRAS) memorandum from the US FDA been made publicly available. Tasaki et al estimated the no-observed-effect level (NOEL) to be 3011 mg/kg b.w./day for male F344 rats and the NOAEL and NOEL in females were to be 3254 mg/kg b.w./day and <778 mg/kg b.w./day, respectively²⁰. This determination was made on the clinical endpoint of body weight gain suppression²⁰. Tasaki et al. also measured alkaline phosphatase activity, but concluded that alterations in this and other enzymatic activity were sporadic and not dose-dependent, so did not consider these or blood content levels of various ions to be grounds for determining the NOAEL or NOEL²⁰.

Methods:

Literature Search:

In order to access a sufficiently large sampling of the literature on ellagic acid, several search methods were used. The results of these searches were documented for reproducibility. A total of 156 published peer-reviewed journal articles were indicated as useful by search results. Only one of the indicated articles could not be obtained, namely Yang 2008. Of the 156 published articles, 82 described animal studies in which ellagic acid was administered. Details of the literature search are as follows.

Literature Search Methodology

A private corporation provided a small database of studies on ellagic acid, many of which were animal studies. These were all included in the literature review. Toxnet, University of Kentucky Library "InfoKat" database, and Google Scholar were utilized. Finally,

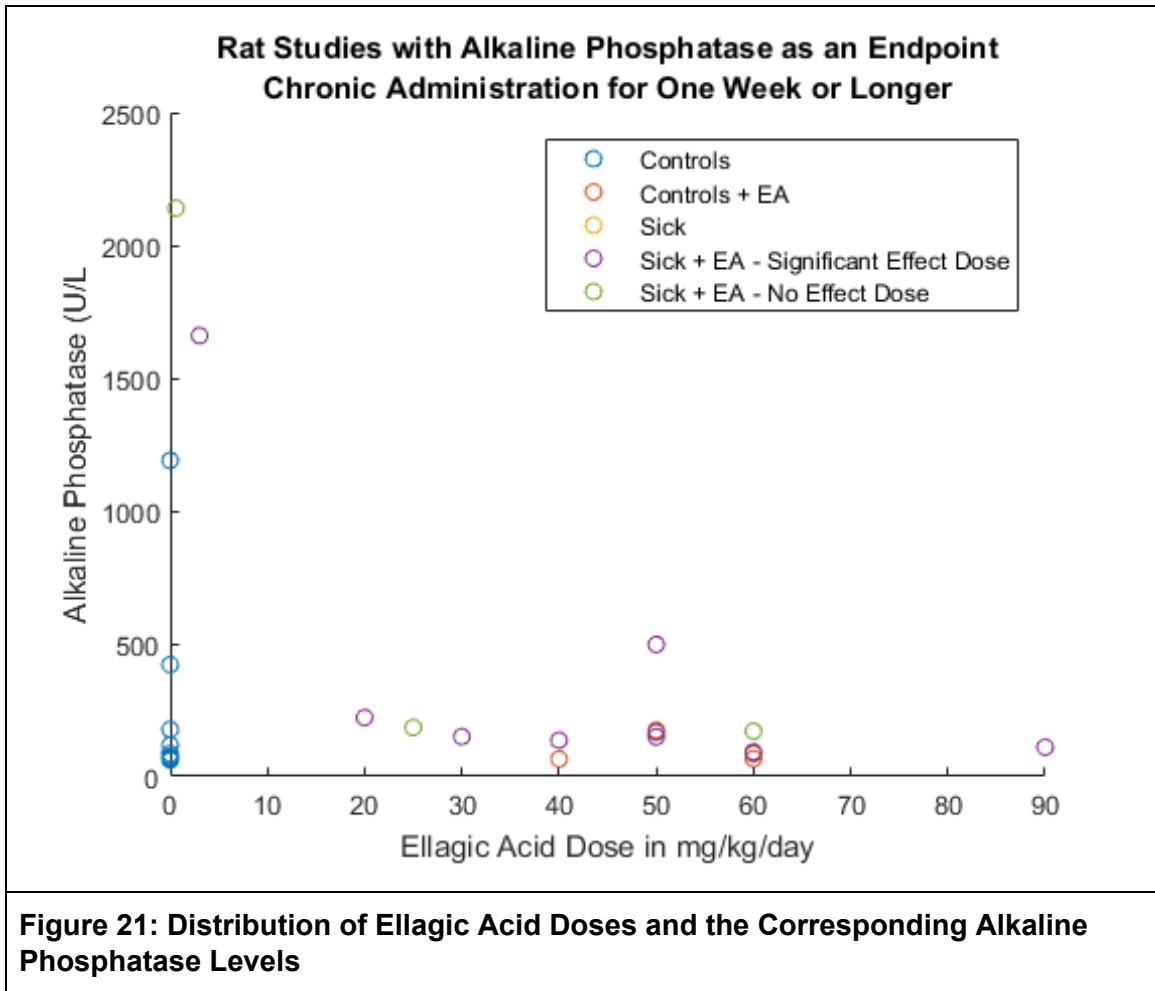
relevant sources cited by all materials were included. To ensure data integrity, a Cochrane protocol was put in place before the selection of studies for review.

Cochrane Protocol:

In order to maintain research integrity, the following Cochrane Protocol for data selection was utilized:

1. Only animal studies with an n of 6 or greater were considered.
2. The studies were analyzed for sameness of endpoints. The most common endpoints shared were oxidative damage and inflammatory markers (such as levels of interleukins).
3. All studies falling into these categories of endpoints were compared, and the endpoint with the most studies having been conducted using the same administrative route in the same animal species for the same time scale (chronic or acute administration) was selected.
4. In the event that there was a tie in the endpoint with the highest number of representative studies, priority would have been given first to an endpoint for which all studies shared an additional endpoint and then to a numeric response variable over a categorical variable.
5. All studies were validated to have significant difference between the control and disease-state control group.

The study criteria determined by this method were as follows: chronic oral-administration studies of ellagic acid performed in rats for a time period of no less than one week that reported alkaline phosphatase activity as an endpoint. Eight studies were isolated that met these criteria. [Gumus (2011); Murugan (2009); Panchal (2013); Ahad (2014); Singh (2009); Thresiamma (1996); Devipriya (2007); Umesalma (2010)] Tasaki (2008) was not considered. Distribution of the doses of ellagic acid and the activity of alkaline phosphatase are shown in Figure 21. The level of confidence for this designation was $p < 0.05$ for all included studies.



Converting Enzymatic Activity Units:

One of the studies indicated by the study criteria, Thresiamma (1996), reported alkaline phosphatase activity in terms of Kingston-Armstrong units per 100 mL serum. The activity was converted into international units per liter utilizing conversion data from “International Enzyme Units and Isoenzyme Nomenclature”. The 2-point format of a line was used to extrapolate the conversion of KAUs first to Bodansky units, and then to international units. It should be noted that the extrapolated quantity was outside the literature range for estimation, which likely contributes to the fact that the alkaline phosphatase activity of all groups was reported to be noticeably higher for this study than the others.

Application of QBEST to Determine QNOAEL

First, a bootstrap replication method was applied to the data. Bootstrapping utilizes random sampling of data from the training set of n rows, with replacement from the training set, to form a new set of samples of n rows (called a bootstrap sample).

Repeated resampling of the training set, forming repeated bootstrap samples, yields the bootstrap sample distribution. The bootstrapping process utilized is multidimensional.

This bootstrapping method is applied because it is nonparametric. Empirical cumulative distribution functions (ECDFs) are functions that approximate the true cumulative distribution function of the data. When the original observational data are bootstrapped nonparametrically using a Monte Carlo approximation to the bootstrap distribution, an estimate of the underlying true cumulative distribution function is obtained.

Doses that were determined to have no significant effect on the activity of alkaline phosphatase were assigned the label “No Effect Doses”. Doses that were determined to have significant effect on the activity of alkaline phosphatase were assigned the label “Effect Doses”. The ECDF of the “No Effect Doses” dataset was integrated and compared with the ECDF of the set of the combined “Effect Doses” and “No Effect Doses”, quantile-by-quantile. The test statistic of this procedure is the correlation coefficient between the two ECDFs, based off of the quantile-quantile plot and the location and scales of quantiles. The hypothesis tested is that the two distributions are the same - that the correlation between the two ECDFs is close to one - with a confidence limit of 98% (the default setting, which can be changed by the user if desired).

A simulation run was conducted in which the test set and the training set were moved far apart in hyperspace (correlation coefficient <1) and then closer together in hyperspace to generate a curve of correlation coefficients at various distances separating the training set and test set. When the simulation finally centered the “Effect Doses” (training) set on the “No Effect Doses” (test) set, the correlation coefficient was close to 1 and there was no significant difference between the two sets. The simulation employed a bisecting divide and conquer algorithm to move the training set and test set together. As long as the two sets remained statistically different, each successive iteration of the simulation run would move the test set one-half the remaining distance toward the training set. The process was repeated until the difference between the sets became statistically insignificant, as indicated by a correlation greater than the 98% level between the two ECDFs formed by testing the training set against itself. Stated another way, the algorithm progressively translated both sets or clusters of data away from each other until it became statistically apparent that the clusters were different. The 98% confidence limits (of the x- and y-axes) on the “Effects” were determined to be the QNOAEL for ellagic acid.

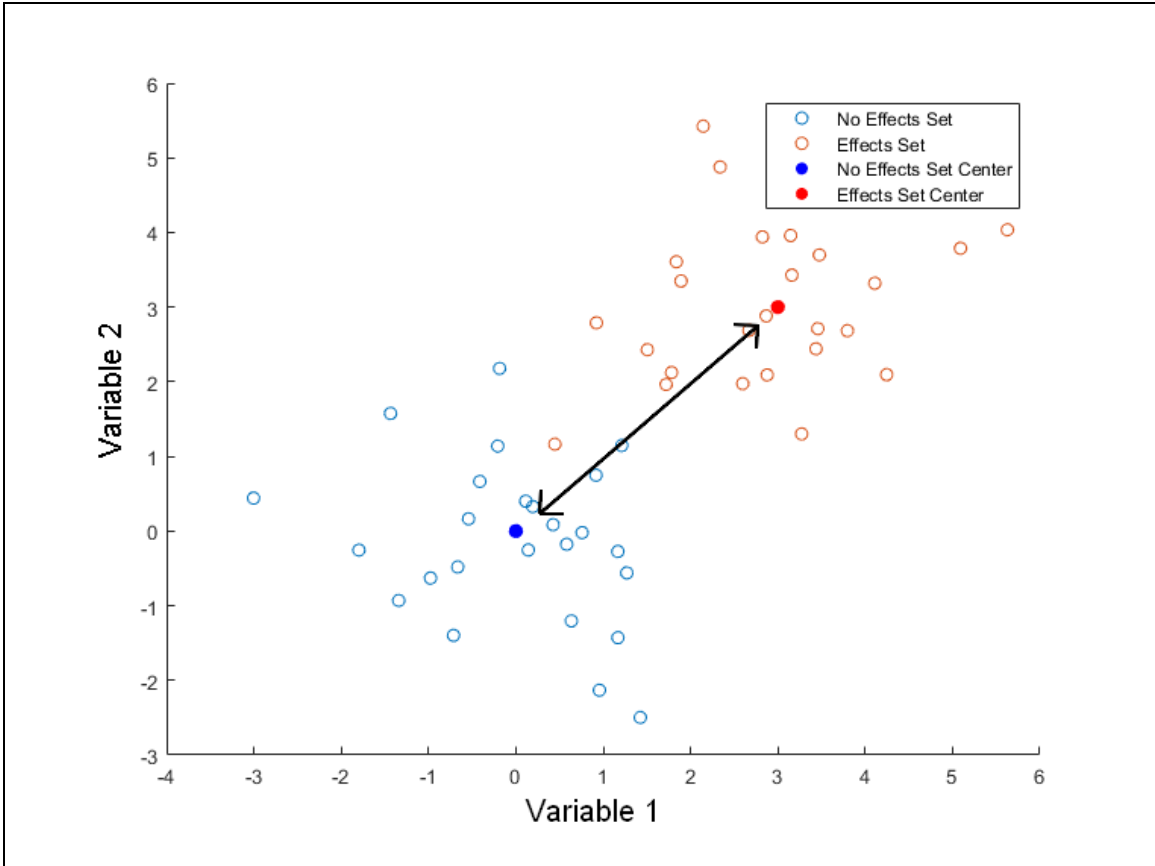


Figure 22: Translating Clusters of Data to Attain the 98% Confidence Level of Difference

The above depicts two theoretical clusters of data being translated toward each other in a stepwise fashion along both axes until the 98% confidence level of difference between the two ECDFs is reached.

Results:

Out of the 82 animal studies reviewed for this experiment, only the following doses were reported to have caused harm to the animal rather than attenuation of a symptom or imbalance induced by a disease state.

Study	Dose	Effect
Damas 1987 "Studies..."	30 mg/kg Intravenous Injection	Swelling of the spleen and noticeable histological changes in lymph nodes.
Paivarinta 2006	1,565 mg/kg/day Oral Feeding for a Prolonged Period	Increases size of adenomas in the duodenum. p=0.029
Damas 1987 "Thrombopenic..."	30 mg/kg Intravenous Injection	Hematological aberrations and swelling of the lymphatic tissues.
Hassoun 1997	6 mg/kg/day on days 10,11,12 of pregnancy and 3 mg/kg on day 13	Ellagic acid control significantly increased fetal death rate despite generally protective effects against teratogenesis when co-administered with TCDD
Bohn 1993	20 microliters in DMSO (0.2 microCi and 100 micrograms) injected into a random uterine horn	Apparent increase in fetal mortality, though the sample size was only three rats.

The most common reported statistically significant effects in the literature were the attenuations of various markers of oxidative stress, malignancy, and metabolic imbalance. In all of the eight studies selected, ellagic acid controls showed no significant change in alkaline phosphatase activity. Only in rats with chronic conditions that lowered the activity of alkaline phosphatase were significant effects of EA administration observed, and levels of alkaline phosphatase were attenuated. In light of this finding, the No Observed Effect Level (QNOEL) of ellagic acid was calculated rather than the NOAEL.

The 98% confidence limit on the "Effects Set" was 32.8129 mg/kg/day ellagic acid. This indicates that the QNOEL of ellagic acid is 32.8129 mg/kg/day in rats.

Discussion:

This study demonstrates the feasibility of using the QBEST equation to generate QNOAELs and QNOELs from animal studies or clinical trials published in the literature. The QBEST equation was utilized to perform discriminant cluster analysis between “Effects Doses” and “No Effects Doses” of ellagic acid orally administered to rats. “Effect” versus “No Effect” classification was done on the basis of the author’s determination that the clinical endpoint was or was not significantly affected with $p < 0.05$. The result of this discriminant cluster analysis was a 98% confidence limit QNOEL of ellagic acid, or a dose that would cause no observed effects 98% of the time if the same studies were to be repeated. This is the first published study demonstrating the feasibility of generating a QNOEL value using the QBEST equation. It is a novel method for estimating the acceptable daily intake (ADI) of a food additive that takes advantage of the published literature, reducing the need to perform costly animal studies (or human studies where human data is available from the literature).

The QNOEL of ellagic acid was determined to be 32.8129 mg/kg/day in rats on the basis of significant increase in alkaline phosphatase activity. This is generally in keeping with the literature in that amelioration of symptoms or attenuation of illness-induced symptoms seems generally to be noted at oral doses up to 30 mg/kg.

However, it should be noted that this determination was largely limited by the literature available. Of the studies surveyed, few had overlapping numeric endpoints and as a result, only eight studies were eligible for analysis. A number of doses were repeated in those studies, as well, reducing the scope of the analysis further. Though the QBEST Equation is capable of calculating a QNOEL based off of this data, increasing training set size has been shown to improve the accuracy of the output. (See Chapter 5 for the effects of training set size on accuracy and precision of QBEST.)

Furthermore, the designation of adverse effect versus effect doses is complicated in the case of ellagic acid. Attenuation of alkaline phosphatase activity which has been decreased by an induced illness could arguably be called an effect, but not an adverse effect of ellagic acid administration. Consistently, doses of up to 0.5g/kg are reported to have caused attenuation of a symptom when administered to rats with induced chronic conditions but no effect on the same endpoint when tested in healthy rats. In fact, Mandal 1990 reported no adverse effects when doses ranging from 0.4 - 8g/kg/day of ellagic acid were orally-administered to rats for a period of four weeks.

Paller et al. has administered ellagic acid orally to human cancer patients as a component of muscadine grape extract tablets. At the highest dose permitted by the study protocol, 8 tablets/day containing at total of 9.6 mg of ellagic acid per day, patients still experienced no adverse effects except for “flatulence, soft stools, and eructation”

when the sample size was increased to 14. Assuming that the average patient in the study weighed 62 kg, the highest daily dose of ellagic acid that patients received was 0.15 mg/kg/day. This daily dose is approximately 220-fold lower than the calculated QNOEL of 32.8129 mg/kg/day in rats.

This discrepancy is important because the literature on ellagic acid strongly suggests a dose-dependent attenuation of cancer symptoms, inflammatory responses, and inhibition of metastasis and tumor growth. Largely, though, these effects are not seen when less than 30 mg/kg/day is administered orally in rodents. This is likely due in part to EA's poor solubility in water and relatively poor absorption in its natural state.

The results of the present study suggest that ellagic acid has a wider therapeutic window than is currently recognized. In future human clinical trials, it may be appropriate to increase the starting dose of ellagic acid. Here, the QNOEL of ellagic acid was determined to be 32.81 mg/kg/day in rats but current human clinical trials administer a maximum daily dose of roughly 0.15 mg/kg/day. The calculated QNOEL indicates that the maximum daily human dose is insufficient to achieve attenuating effects of ellagic acid, as measured by alkaline phosphatase activity.

A study of American food consumption data from the NHANES database found that the mean estimated intake of ellagic acid from the food supply is 1.05 $\mu\text{g}/\text{kg}$ body weight/day, while the 90th percentile daily intake is 3.89 $\mu\text{g}/\text{kg}$ body weight/day¹⁵⁴. The 90th percentile daily intake of ellagic acid from the food supply is indicative of a "heavy user" dose for human administration. In order to elucidate the NOEL of ellagic acid in humans, an administration study should be performed between this value and QNOEL determined in rats (with the maximum dose being reduced by a factor of 6.2¹⁵⁵, as is suggested when moving from rat toxicity studies into human studies). A clarification for whether adverse effects are induced by pure ellagic acid at the dose of 0.15 mg/kg/day, the ellagic acid dose-equivalent for muscadine grape tablets administered in human clinical trials by Paller et al., should be provided by the proposed study and will indicate whether ellagic acid alone produces observed effects

Appendix: Search Methodology & Cochrane Protocol

Literature Search Methodology

Corporate Literature Provision:

A private corporation provided a small database of studies on ellagic acid, many of which were animal studies. These were all included in the literature review.

Web-Based Search Terms:

1. "Ellagic Acid" search on Toxnet yielded few relevant results, though all were included in the source database.
2. "Ellagic Acid NOAEL" was searched on the University of Kentucky Library "InfoKat" database. All relevant papers returned in the 69 results were included.
3. "Ellagic Acid NOEL" returned no relevant results on the first page of UK library search, so this search was concluded.
4. "ellagic acid noael" Searched on 5/22/2018 returned on Google Scholar 726 results. Many are duplicates of papers from Company Files or UK Library search. Relevant non duplicate items in the first 40 results were obtained.
5. Johanningsmeier 2011 was not included due to the large number of pomegranate review papers already included.

Sources Cited by Previous Entries:

Relevant sources cited by the papers found by the above methods (company files, UKY search, and Google Scholar search) were determined. Because the corporate literature search provided a fairly comprehensive sampling of papers published up to 2008, citations were taken from all papers published in 2008 or later.

1. In reviewing previously obtained sources, some citations stood out in the bodies of the text as being particularly relevant. If a source was noted in the body of text as being a human or animal study involving ellagic acid, the cited source was generally obtained immediately.
2. The titles of all sources in the citation sections were reviewed. Any titles including the words "ellagic acid" were obtained, unless the study titles directly indicated that the source was an *in vitro* study or cell assay. Studies whose titles indicated that they covered only the administration of pomegranate products or plant extracts were also omitted. Articles of questionable relevance were reviewed briefly before being included in the in-depth literature review.
3. Resources were not duplicated if they were already included.

4. Sources were not taken from Faria 2011 because a copy was not obtained before the citation search. No sources were used from this article as it could not be gotten in full and there are more recent pomegranate review articles, the sources of which were reviewed.
5. The search for cited sources was begun with the most recent references.
6. A very few papers used citation methods that omit the title of the cited article. The text was searched for instances of relevant citations, and then relevance was validated by obtaining the journal article.

Cochrane Protocol:

In order to maintain research integrity, the following Cochrane Protocol for data selection was utilized:

6. Only animal studies were considered with an n of 6 or greater.
7. The studies were analyzed for sameness of endpoints. The most common endpoints shared were oxidative damage and inflammatory markers (such as levels of interleukins).
8. All studies falling into these categories of endpoints were compared, and the endpoint with the most studies having been conducted using the same administrative route in the same animal species for the same time scale (chronic or acute administration) was selected.
9. In the event that there was a tie in the endpoint with the highest number of representative studies, priority would have been given first to an endpoint for which all studies shared an additional endpoint and then to a numeric response variable over a categorical variable.
10. All studies were validated to have significant difference between the control and disease-state control group.

The study criteria determined by this method were as follows: chronic oral-administration studies of ellagic acid performed in rats for a time period of no less than one week that reported alkaline phosphatase activity as an endpoint.

CHAPTER 5: VALIDATING THE ACCURACY OF QBEST ON SYNTHETIC DATA: ROUND THE COMPASS ROSE

Purpose: This chapter evaluates the accuracy of the QBEST algorithm on synthetic data and evaluates what factors most affect the accuracy of the algorithm.

Validating the Accuracy of QBEST on Synthetic Data: Round the Compass Rose

Introduction:

QBEST is a discriminant cluster analysis method that can be used to identify whether a new point ought to be classified as a member of one or another multivariate data clusters⁴. It is an improvement over the Mahalanobis equation, which has been described as “a rubber ruler”. Specifically, the Mahalanobis equation calculates the number of standard deviation units of a point from the center of a cluster of normally distributed data points along a given axis, or off that axis. QBEST, on the other hand, is a “rubber ruler with a nail in the center”, meaning that it calculates the standard deviation of a point from a cluster of data points along a given vector, extending from the center of the cluster towards the test point. This strength allows the QBEST to be applied to multivariate clusters of data that are irregular in shape (e.g., not normally distributed) with greater accuracy than the Mahalanobis.

QBEST has already been applied to real-world problems of sample differentiation. QBEST has been applied to the problem of differentiating clusters of data from infrared (IR) spectroscopy of medicine capsule samples^{1, 5}. QBEST was demonstrated to be able to identify capsules contaminated with metals or cyanide⁵. In another experiment utilizing QBEST, pure Anacin capsules were distinguished from capsules contaminated with common industrial contaminants, such as aluminum powder and dust⁶. It has also been used to differentiate between benzoic acid and false samples via IR analysis⁷.

The algorithm’s accuracy and performance when handling spherically distributed $N(0,1)$ data has been evaluated^{3,4}, but an evaluation of the algorithm’s accuracy and bias when handling elliptically distributed $N(0,1)$ data has never been calculated. The purposes of this experiment were (1) to validate QBEST’s accurate handling of elliptical data; (2) to evaluate what factors most influence the accuracy of the algorithm; (3) to determine the maximum accuracy with which the average laptop computer could calculate BEST distances for elliptical data. To this end, synthetic data clusters’ known statistical parameters were compared to the theoretical QBEST distance calculated by the algorithm. A standard ellipse served as the basis of this model. Future research should evaluate QBEST’s treatment of skewed clusters.

Materials:

System Specifications:

As reported by Matlab's "memory" command, the laptop used for the memory usage test had the following capabilities:

Maximum possible array:	5089 MB (5.337e+009 bytes) *
Memory available for all arrays:	5089 MB (5.337e+009 bytes) *
Memory used by MATLAB:	624 MB (6.545e+008 bytes)
Physical Memory (RAM):	7366 MB (7.724e+009 bytes)

All memory data was generated on a laptop PC with 7366 MB of RAM. The operating system used was Windows 10.

Other data were generated on a Lenovo 3437CTO X-64 based laptop PC with an Intel® Core™ i7-3520M CPU at 2.90 GHz, 2901 MHz, with 2 cores and 4 logical processors. The operating system used was Windows 7 Professional, Version 6.1.701 Service Pack 1 Build 7601.

Matlab R2017a was used for all calculations.

Methods:

Standard Circular Data:

In order to validate the program for the standard data generation, first a circular cluster of data were generated in Matlab. Utilizing the `mvnrnd()` function, data with a known standard deviation of 1 along both the x and y axes, centered on the point (0,0) were generated. Figure 23 is an example of a cluster of data thus generated. The mean, or center point, of the cluster is shown as a red point and the theoretical first standard deviation is shown as a red circle with radius=1. The second standard deviation is shown in blue. In this example, 45/100 fall within the red circle. 43/100 points fall between the red and blue circles or on the blue circle. 12/100 points fall further from center than the second standard deviation. Theoretically, for normally distributed data, 68% of the points should fall within one standard deviation of the mean, and 95% within two standard deviations, leaving less than 2% of data outside the blue circle. It is known that random number generators do not perfectly generate data fitting the input parameters, but an approximation is required for this research. Furthermore, it is likely that the command in question merely considers the standard deviation along the x axis when generating the x coordinates and the standard deviation along the y-axis when

generating the y-coordinates; that is to say that Matlab does not use a target covariance matrix when generating points using `mvnrnd()`.

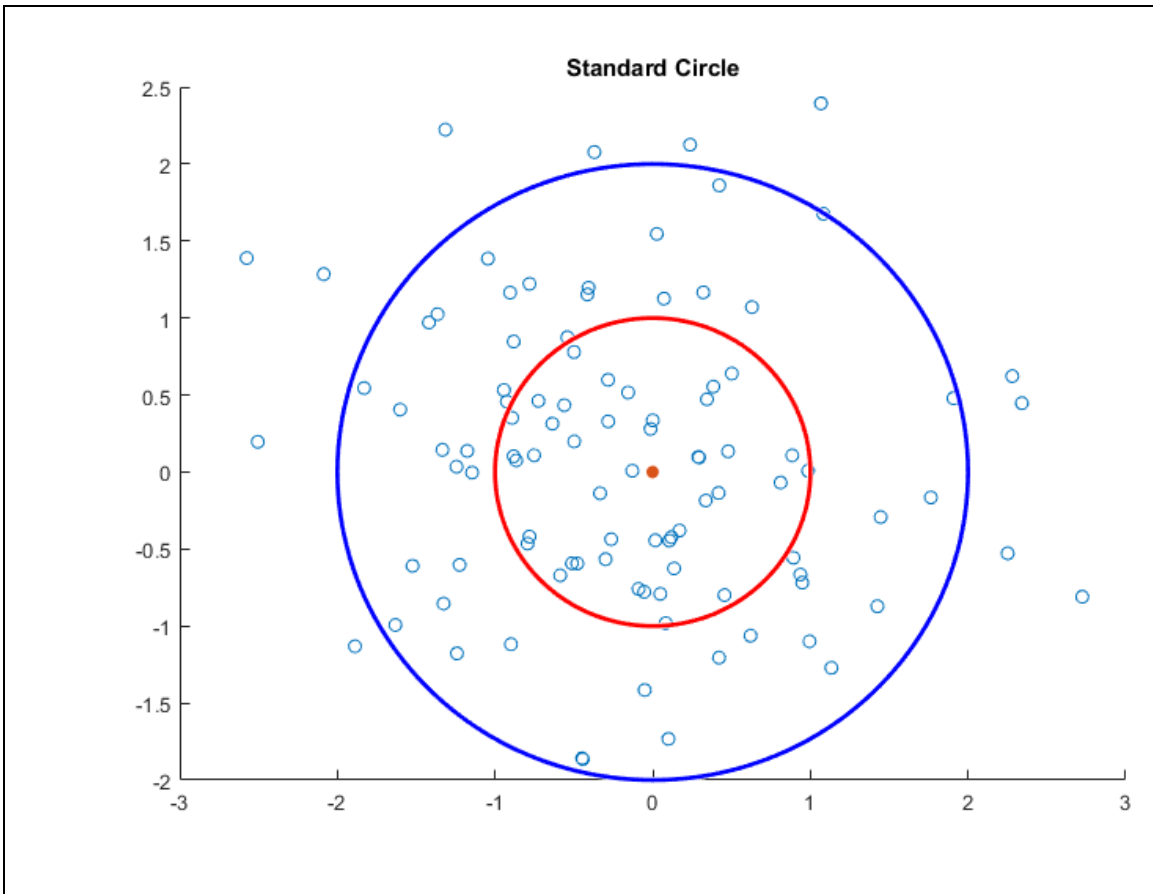


Figure 23: Standard Circle

A standard circular cluster of data with center (0,0) and standard deviation of 1 in all directions. The first standard deviation distance is shown in red and the second standard deviation distance in blue.

Standard Elliptical Data:

Then, using the `mvnrnd()` function, data were generated with a known standard deviation of 1 along the y axis and 2 along the x axis. Figure 32 is an example of 100 data points generated in this fashion with center shown in red and the standard ellipse with principal axis length equal to 2 and minor axis length equal to 1 shown in red. The eight compass points utilized in the experiment are shown in red.

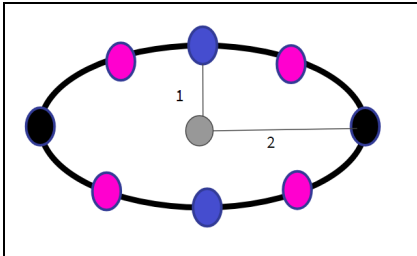


Figure 24: Stylized Standard Ellipse

A stylized standard ellipse with major principal axis points shown in black, minor principal axis points in blue, and diagonal points in pink.

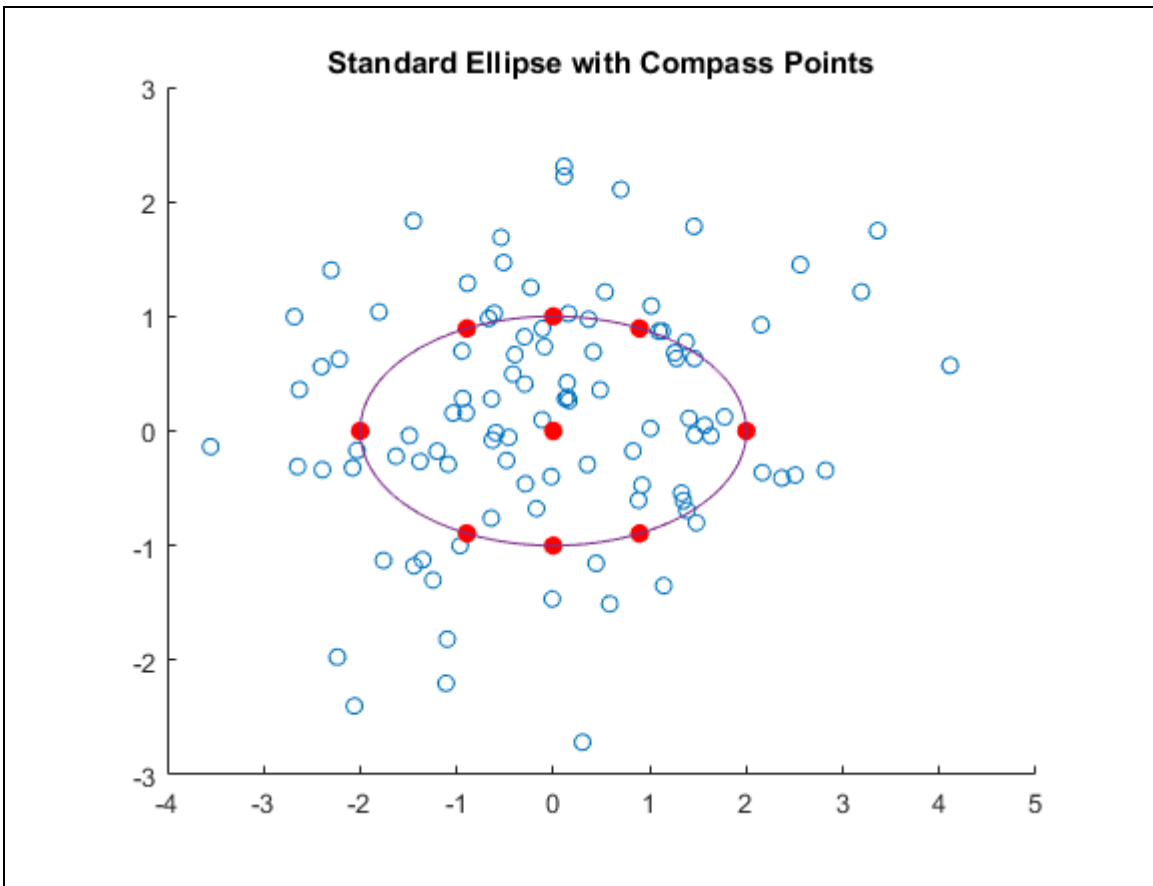


Figure 25: Standard Ellipse with Compass Points

An example of the elliptical cluster of data generated in this experiment with $\mu = (0,0)$ and $\sigma=(2,1)$. Compass points on the standard ellipse shown in red. The red line indicates 1 standard deviation distance from center.

The eight compass points shown are as follows, and comprise the farthest points along the major and minor axes as well as the points located at a 45-degree angle from the x and y axes:

Table 3: The Eight Compass Points of the Standard Ellipse		
Principal Axis Points	Non-Principal Axis Points	45 Degree Diagonal Points
(-2,0)	(0,1)	(-2/(sqrt(5)),(2/(sqrt(5))))
(2,0)	(0,-1)	(-2/(sqrt(5)),(-2/(sqrt(5))))
		(2/(sqrt(5)),(-2/(sqrt(5))))
		(2/(sqrt(5)),(2/(sqrt(5))))

Increasing Dimensions:

The same process was used to increase the dimensionality of the circular and elliptical clusters. For increased dimensions in the data, the center remained the same but was assigned additional zero-coordinates. For example, the three-dimensional center would be (0,0,0). A standard deviation of 1 was assigned to the `mvrnd()` cluster in all directions.

Nested Iterative Loops

The program `Round_The_Compass_Rose_increasing_training_set.m` was used to evaluate the algorithm's handling of the data as each variable increased. This program, available at our lab website ([link](#)) utilizes a series of nested iterative loops.

The outermost loop changes the fraction of data points used in the generation of the hypercylinder, "radfrac". A minimum, step, and maximum value are set for radfrac.

The second nested iterative loop increases the number of datapoints in the training set (denoted by the variable "u") by increasing the number of points generated by the `mvrnd()` function. Note that the actual number of training points generated is equivalent to $50*u$. The line of code relevant to this command is:

```
TNSPEC=mvrnd(mu,sigma,50*u);
```

The third nested iterative loop increases the dimensions or number of wavelengths of the data points (denoted by the variable "v"). "v" is indicative of the number of dimensions of the data points above two. That is, when $v=0$, the data is generated in a 2-D ellipse.

When $v=1$, the data is three-dimensional; when $v=2$, the data is four-dimensional; etc. Dimension is increased in the training set by centering the `mvrnd()` cluster on $(0,0,\dots,0)$, with the appropriate number of zeroes being added to the coordinate for the number of dimensions - that center point being assigned to the variable "mu". Then, "sigma" is added such that the standard deviation of the `mvrnd()` cluster is $(2,1,1,\dots,1)$, with ones being appended such that the number of dimensions is increased to the proper value. In this loop, as well, the compass points are defined and added with zeroes to make them the proper dimensions. The compass points are defined:
`compass_points=[0,1;(-2/(sqrt(5))),(2/(sqrt(5))); -2,0;(-2/(sqrt(5))),(2/(sqrt(5)));`
`0,-1;(2/(sqrt(5))),(2/(sqrt(5))); 2,0; (2/(sqrt(5))),(2/(sqrt(5)))];`

Compass points of increased dimensions would have all further dimensions equal to zero. $(0,1, 0,0,\dots,0)$, for example.

Furthermore, this loop defines the expected standard deviation from center of each compass point:
`expected_standard_deviation = [1;1.2649;2;1.2649;1;1.2649;2;1.2649];`

The fourth nested iterative loop defines "newspec", the test spectrum, as being equivalent to one of the eight compass points. This loop tests each compass point using the method of the fifth iterative loop. It also creates storage for the solutions generated in the fifth iterative loop. The fifth iterative loop is run a total of six times, with different bootstrap replicates being generated, and the average and standard deviation of the outputs is calculated and stored.

The fifth iterative loop increases the number of bootstrap replicates generated. Bootstrap replicates are equivalent to $50 \times "z"$. This loop actually calls the `qb.m` program. The code for this loop is:
`for i=1:z`

```

    B=50*i;
    sds = nan; bias = nan; sds skew = nan;
    try
    [BTRAIN,CNTER]=replica(TNSPEC,B);
    [sds,sds skew,qrr] = qb(TNSPEC,BTRAIN,newspec,CNTER,radfrac,sensitiv);
    bias=sds-expected_standard_deviation(j);
    catch ME
    end
    SDS_matrix((i,:))=[compass_points(j,:),B,sds,bias,sds skew];
end

```

The program used for multivariate polynomial regression of the final results was <https://www.mathworks.com/matlabcentral/fileexchange/34918-multivariate-polynomial-regression> .

Memory Usage Tests:

In order to determine the maximum parameter size for each required input of the algorithm, the input values of each variable training set size, number of bootstrap replications, and dimension were increased until the program ran for a continuous 20 minutes without completing. (This was determined to be indicative that the program would not complete its run, even given more than an hour.) Then, variables were increased together. For these tests, the same input was used for both or all variables being tested. All variables were increased until the run time exceeded 20 minutes, the program raised a memory error and terminated, or the cpu time of the last run approached 200 seconds (as this was also indicative of an imminent crash).

Results:

Memory Usage:

The following chart shows the maximum input sizes attained for each variable alone or in concert with another variable. The amount of data that a personal laptop computer can analyze far exceeds what is possible using the Mahalanobis equation. A personal laptop computer can run up to 6 million bootstrap replications with a small dataset, or can handle 1 million dimensional data, or 5 million training points provided that the other parameters are small with a run time of less than 20 minutes. This far exceeds the needs of most research scientists.

Table 4: Memory Usage of QBEST				
Bootstrap	Dimension	TNSPEC	Memory Used by Matlab (MB)	Clock Time of Second Operation (s)
6*10 ⁶	2	100	Not Determined	Not Determined
100	10 ⁶	150	Not Determined	Not Determined
100	2	5*10 ⁶	Not Determined	Not Determined
10 ⁴	2	10 ⁴	490	11.803887
100	10 ⁴	10 ⁴	1257	839.131357
10 ⁴	10 ⁴	100	1225	194.179892
10 ³	10 ³	10 ³	464	42.909076
3*10 ⁶	2	2*10 ²	500	87.75957
3*10 ⁶	30	100	1182	66.53017
100	5*10 ⁵	1.5*10 ²	1422	268.2744
150	5*10 ⁵	150	1608	405.3309
5*10 ³	350	5*10 ³	473	198.7201
10 ⁴	200	10 ⁴	476	489.2994
125	10 ⁴	10 ⁴	1257	1359.073
150	5*10 ³	5*10 ³	685	155.3649
10 ⁴	10 ⁴	150	1257	387.0736
5*10 ³	5*10 ³	200	668	105.0412
100	50	2.5*10 ⁶	1466	1073.131
10 ²	2	2.5*10 ⁶	579	56.26981

Relative Error or Bias

The algorithm's absolute error was calculated to be equal to one minus the average sds output of qb.m, over a course of six runs. Tested at each compass point, the theoretical true number of standard deviations of distance that each test point has from center is one. SDs represents the number of standard deviations of distance calculated by the algorithm. Thus, the absolute error of the answer is one minus the average value of sds.

Effect of Radfrac & Bootstrap Replication on Average Relative Error:

Relative error remains constant regardless of the the number of bootstrap replications performed or the fraction of points used in generating the hypercylinder (when considering elliptical data that has no concavities). As shown in Figure 26, the average percent error remains relatively unchanged as bootstrap replicates increase (across the x-axes of each graph) or as radfrac increases (increased from top figure to bottom figure).

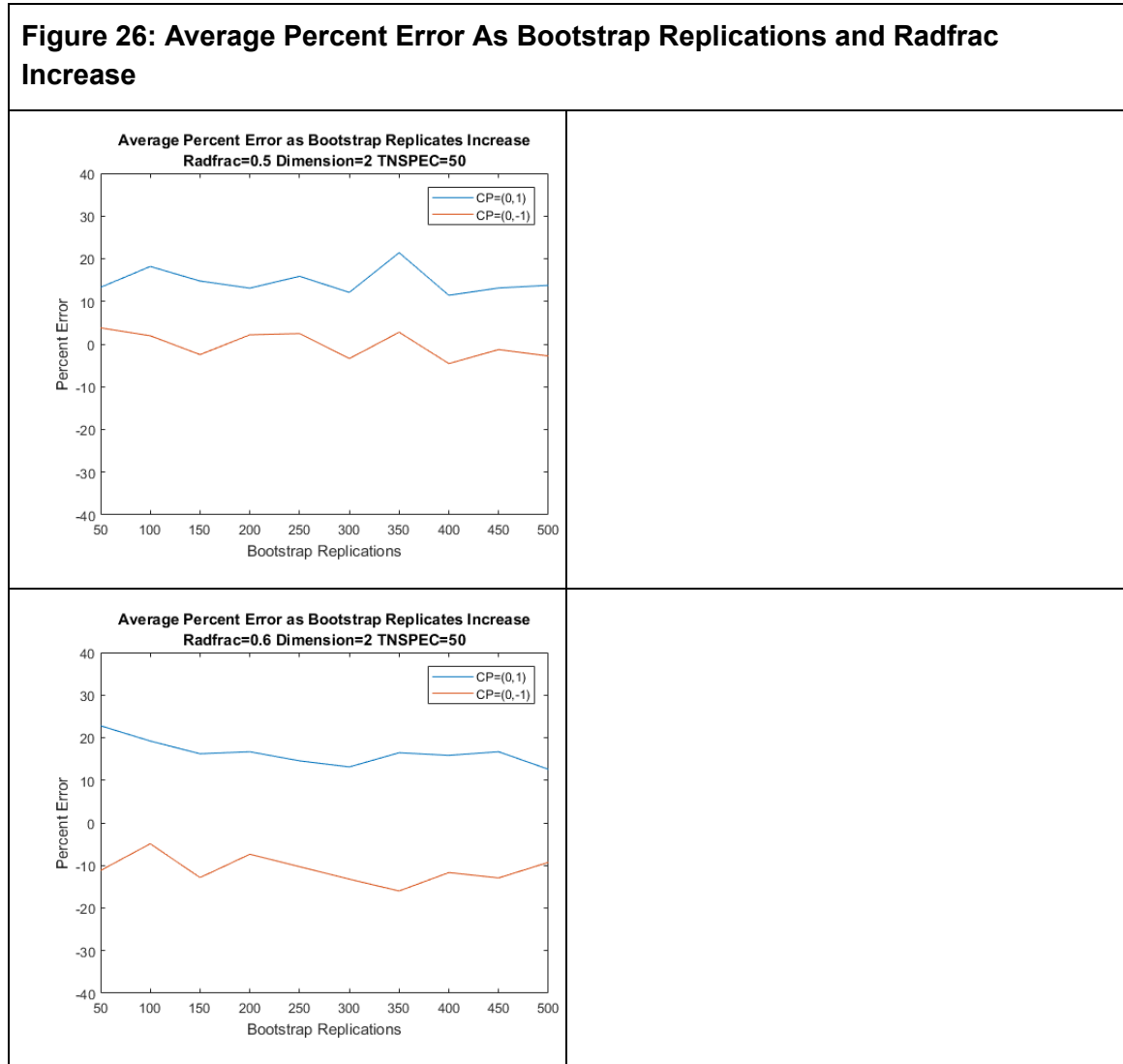


Figure 26 Continued

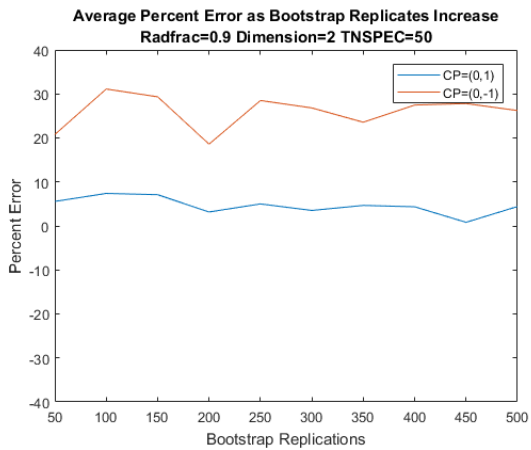
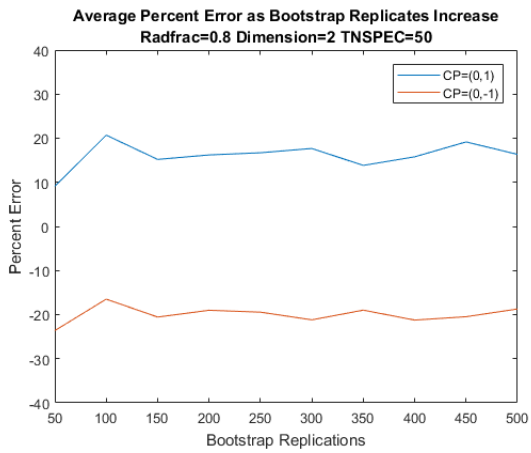
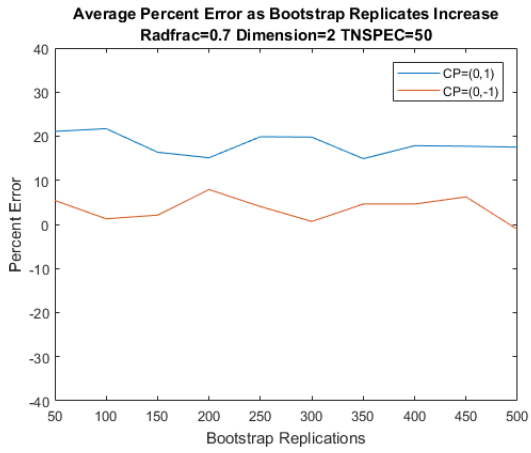
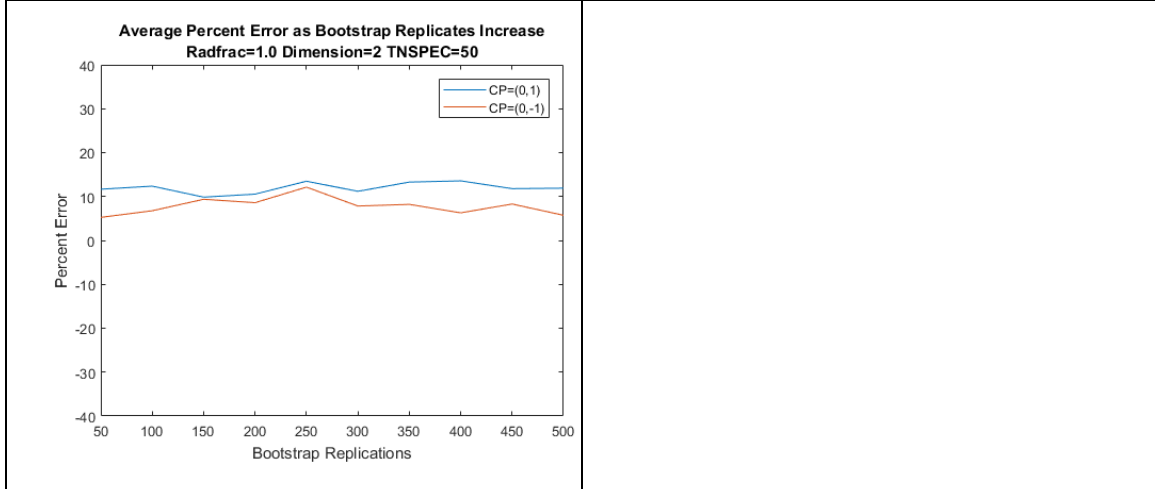


Figure 26 Continued



Effect of Training Spectrum Size and Dimension on Relative Error:

Neither training spectrum size nor dimension are significant predictors of relative error. Increasing the size of the training spectrum appears to decrease the percent error slightly, though it is difficult to say how much of this decrease is due to the inborn error of the training spectrum generation by the `mvrnd()` command and how much of the decrease is due to increased competency when the algorithm bootstrap replicates from a larger training set of data. Increasing the number of dimensions of the datapoints in question while holding other factors constant appears to slightly increase absolute error. These effects are largely negligible, though. This data is shown in Figure 27.

Figure 27: Effect of Training Spectrum Size and Dimension on Relative Error

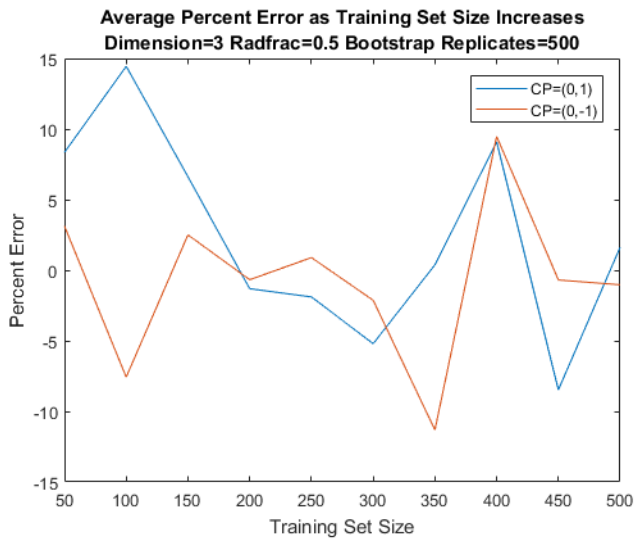
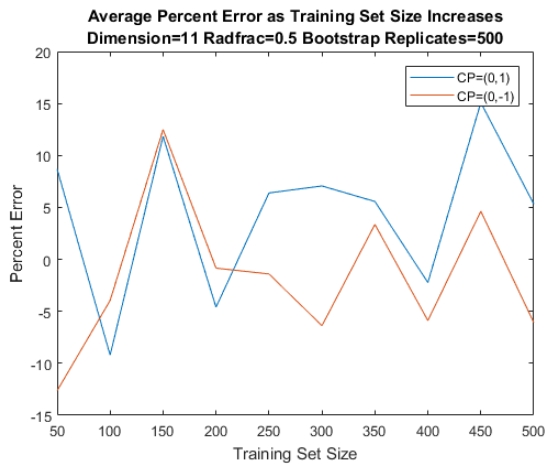
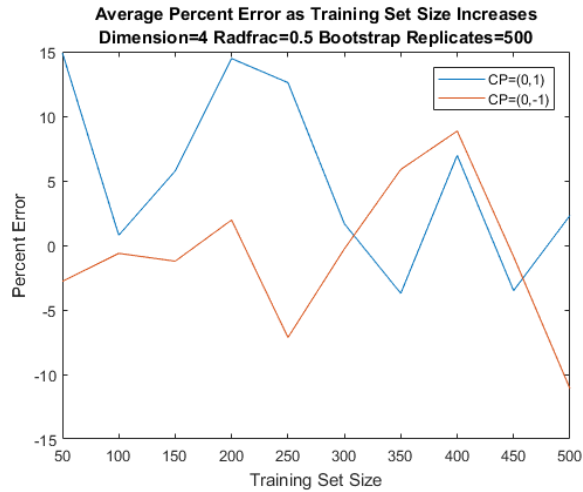
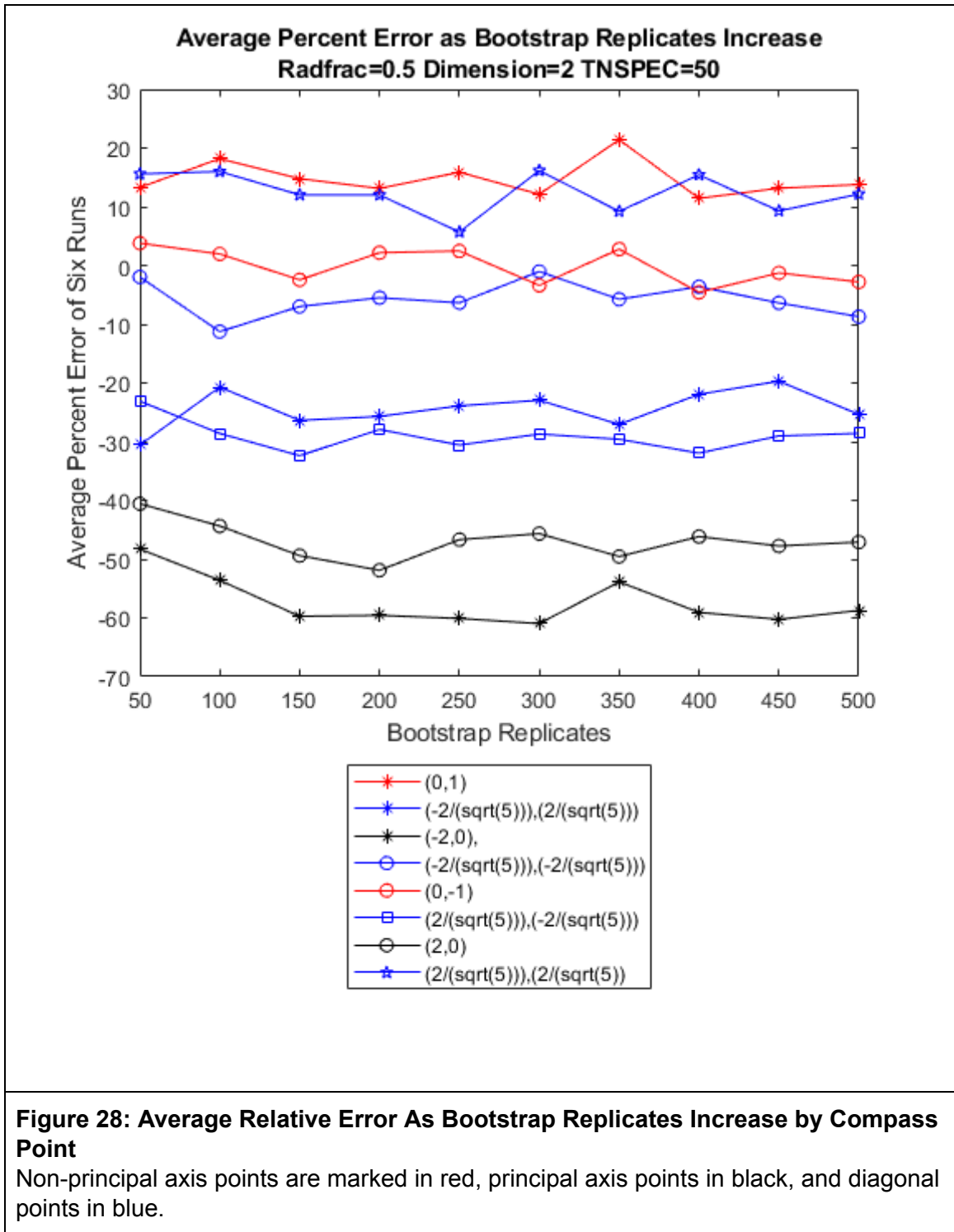


Figure 27 Continued



Effect of Compass Point on Relative Error:

Compass point is not a significant predictor of relative error. However, like compass points tend to cluster together in terms of relative error due to eccentricities in the training set. The training set largely determines the relative error of each compass point.



Relative Standard Deviation:

The algorithm's relative standard deviation (RSD) was calculated by dividing the standard deviation of the six runs by the average distances in SDs of the six runs. Relative standard deviation is decreased by increasing the number of bootstrap replications run on the data. Increasing the size of the training spectrum or increasing the number of dimensions of the data appears to have no significant effect on the relative standard deviation of the output of the algorithm. This indicates that the algorithm handles with equivalent precision increasingly multivariate datasets, as well as datasets for which more or less training points were input. The effect of *radfrac*, or fraction of points used to generate the hypercylinder on relative standard deviation is low in proportion to the effect of bootstrap replication for data that is elliptical, or lacking in concavities. Non-linear functions, such as those seen in acoustic data, can create data clusters with concavities (e.g., clusters that appear shaped like the letter "c"). In these instances, the influence of *radfrac* should theoretically be quite great. Further research should be done on data clusters including concavities to determine the nature of this effect.

Figure 29: Relative Standard Deviation as Bootstrap Replicates and Radfrac Increase

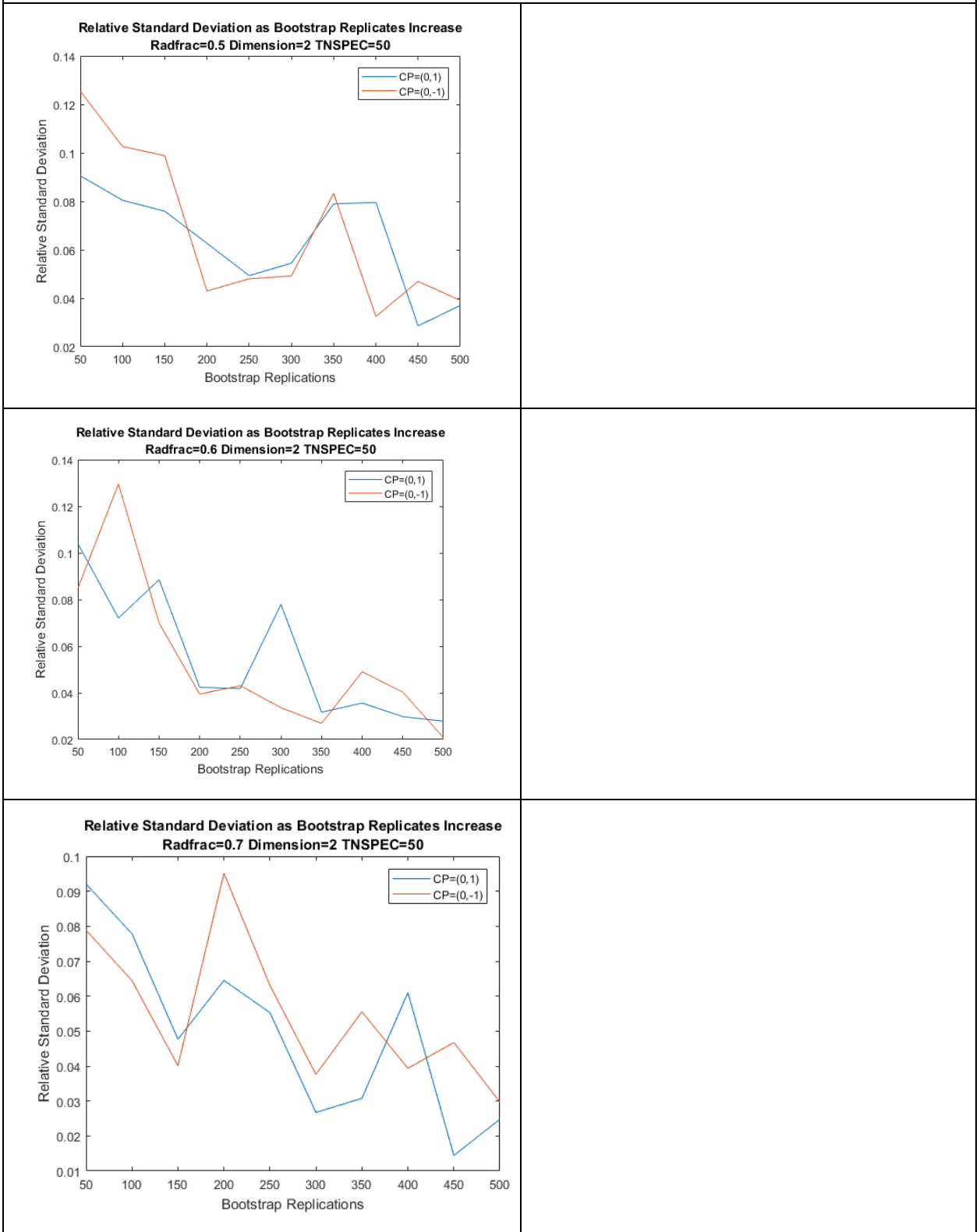
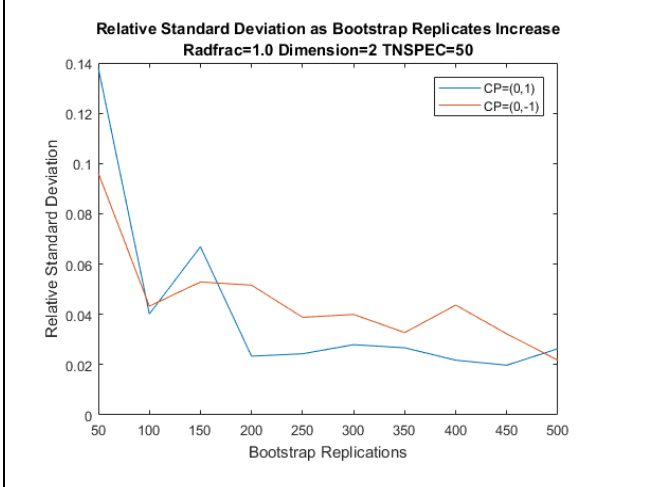
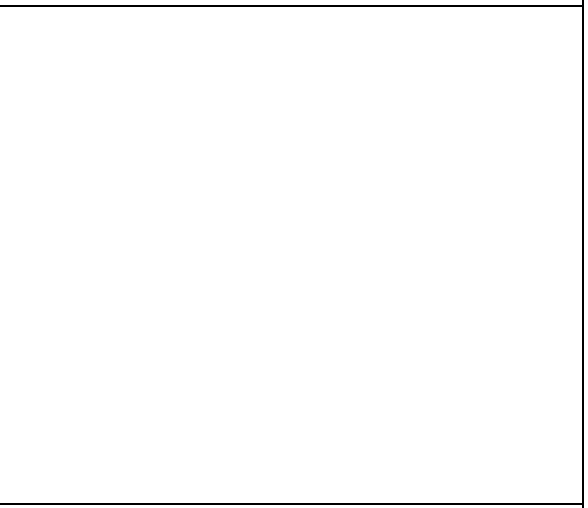
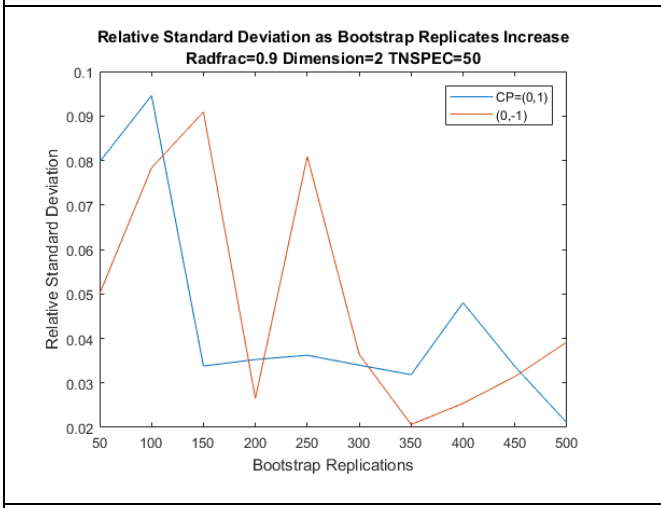
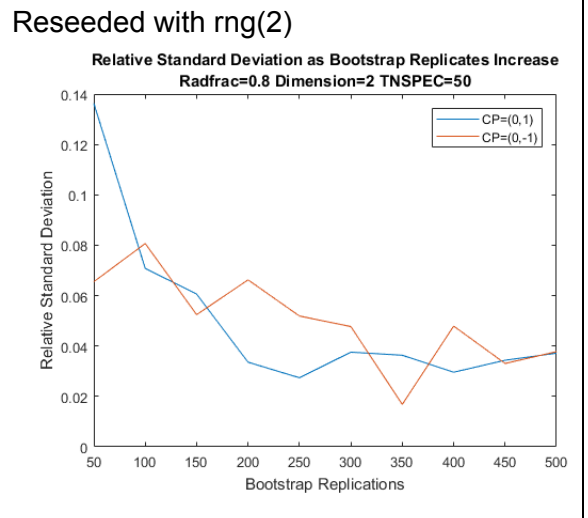
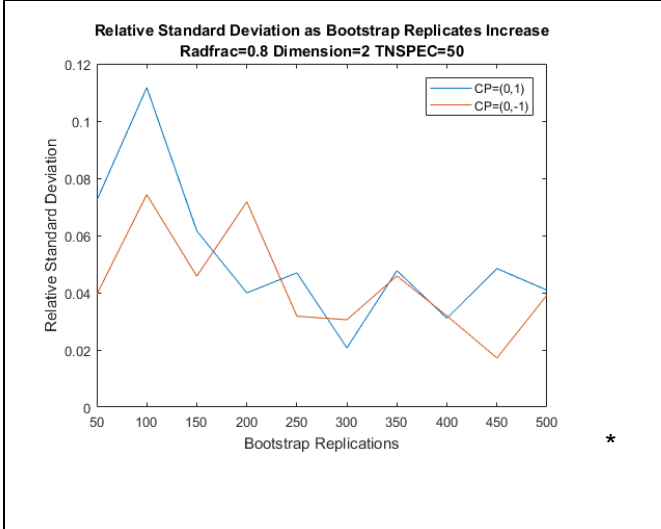


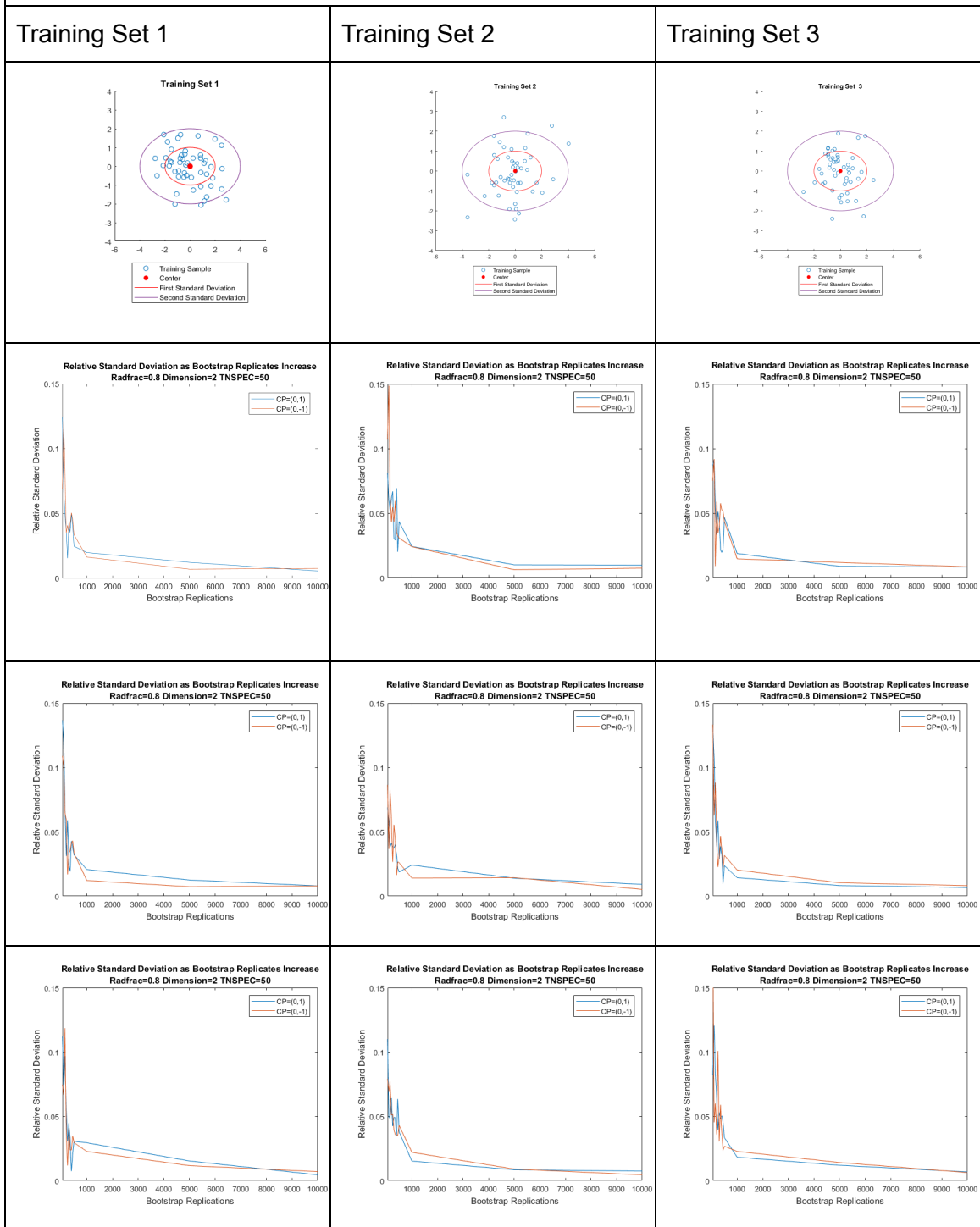
Figure 29 Continued



RSD is Dependent on Training Set

The relative standard deviation of QBEST's output is most significantly affected by the training set chosen. In Figure 30, below, three different elliptical clusters of data have been randomly generated using the same parameters. The RSD vs. bootstrap replicate graphs below them are three graphs generated from the same training spectrum, but bootstrap replicated with a different random number generator seed. Graphs generated from the same training spectrum are nearly identical to one another, but significant variation exists between graphs generated from a different training spectrum.

Figure 30: Relative Standard Deviation as Bootstrap Replicates Increase: Effect of Training Sample and Random Bootstrap Replicates

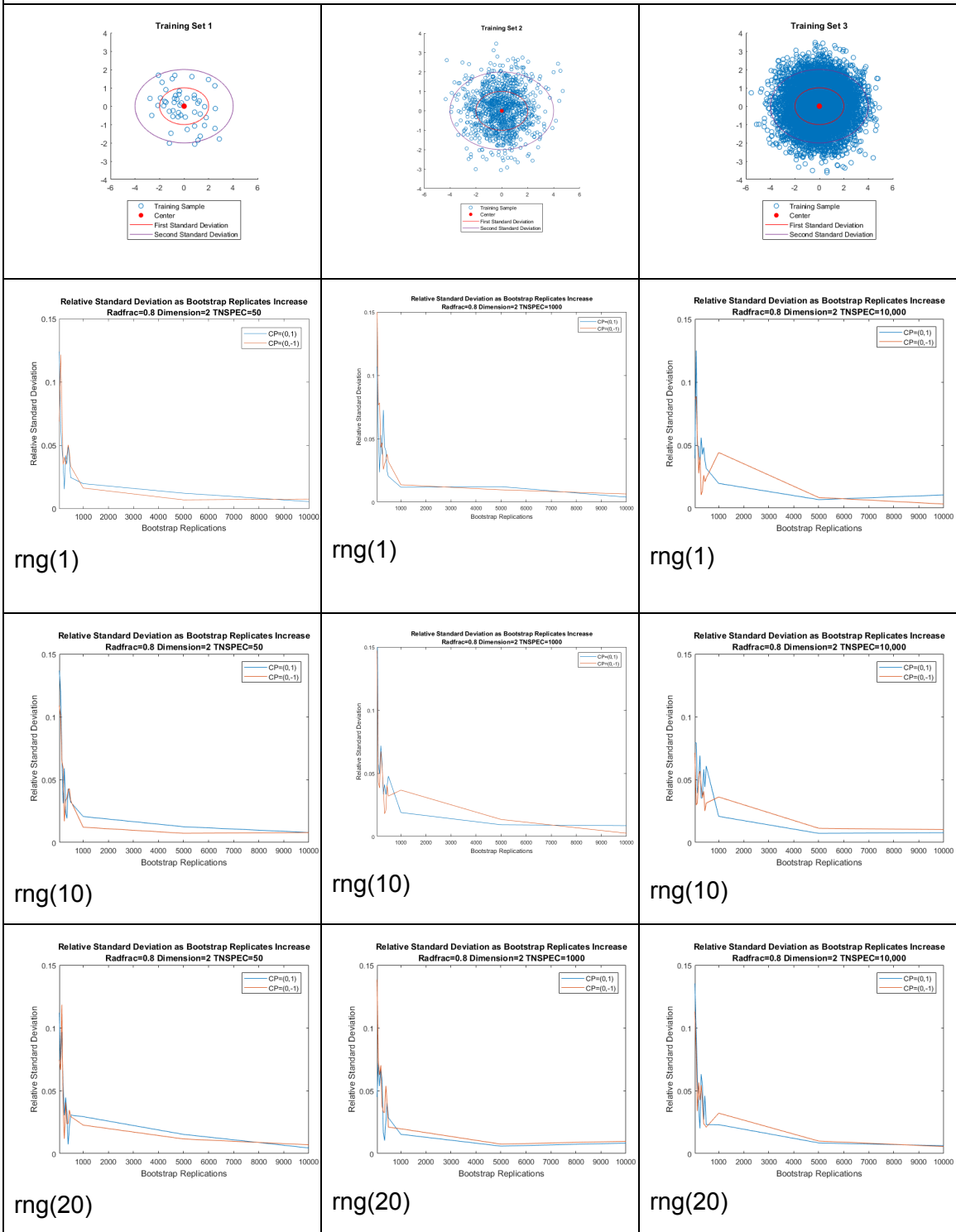


Number of Training Points Does Not Significantly Affect RSD

Figure 31 below shows elliptical clusters with the same parameters but different numbers of generated points. RSD calculations were run in triplicate for each training spectrum using a different random number generator seed for the bootstrapping process.

Changing the training set size does not appear to significantly affect relative standard deviation and is not a significant predictor of RSD. Increasing the number of bootstrap replicates even marginally drastically reduces the RSD. Bootstrap replicate number is a far more important factor in the RSD than number of training points.

Figure 31: Relative Standard Deviation as Bootstrap Replicates Increase: Effect of Training Set Size and Random Bootstrap Replicates



Relative Standard Deviation as TNSPEC and Dimension Increase:

RSD is not significantly affected by the training spectrum size or dimension increase. Neither of these are significant predictors of RSD.

Figure 32: Relative Standard Deviation as Training Set Size and Dimension Increase

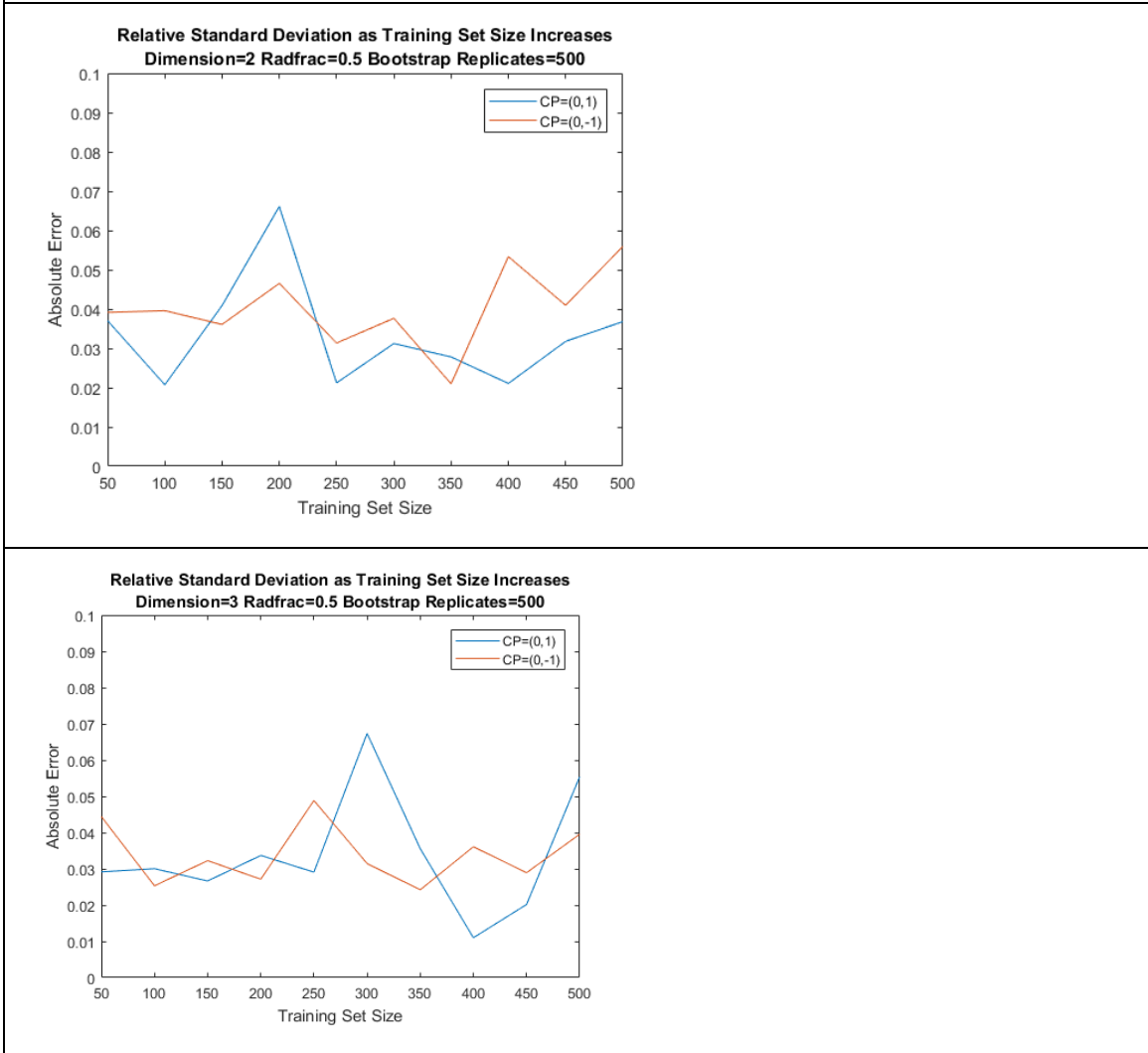
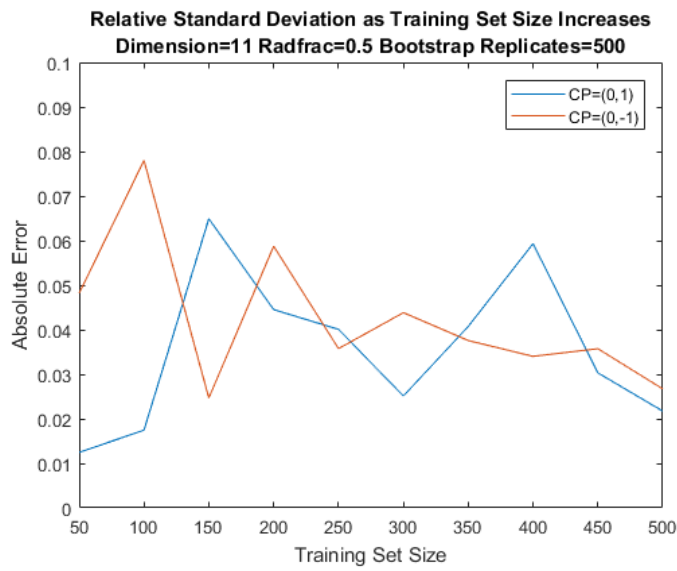
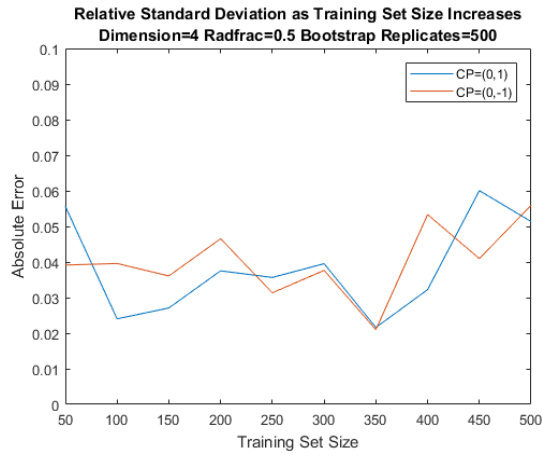


Figure 32 Continued



Mahalanobis Comparison

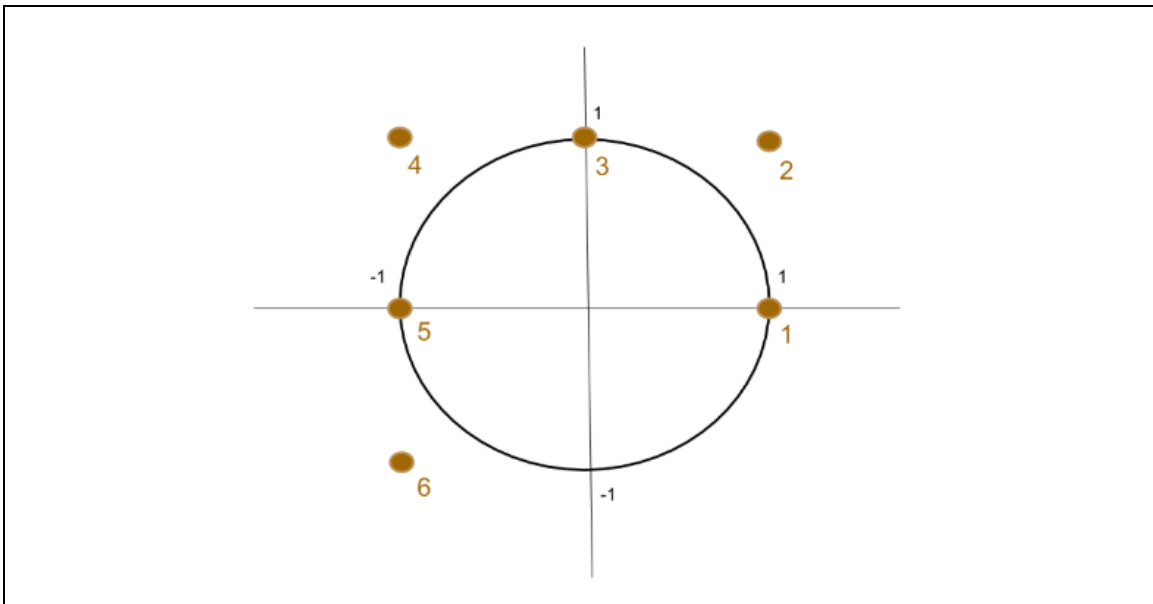


Figure 33: Standard Circle Used for Mahalanobis vs. BEST Comparison³

When the number of samples greatly exceeds the number of dimensions (independent variables), the BEST and the Mahalanobis metrics produce similar answers.

Theoretically, the on-axis points (1, 3, and 5) should be 1 SD from the center while the off-axis points (2, 4, and 6) should be 1.414 SD away.

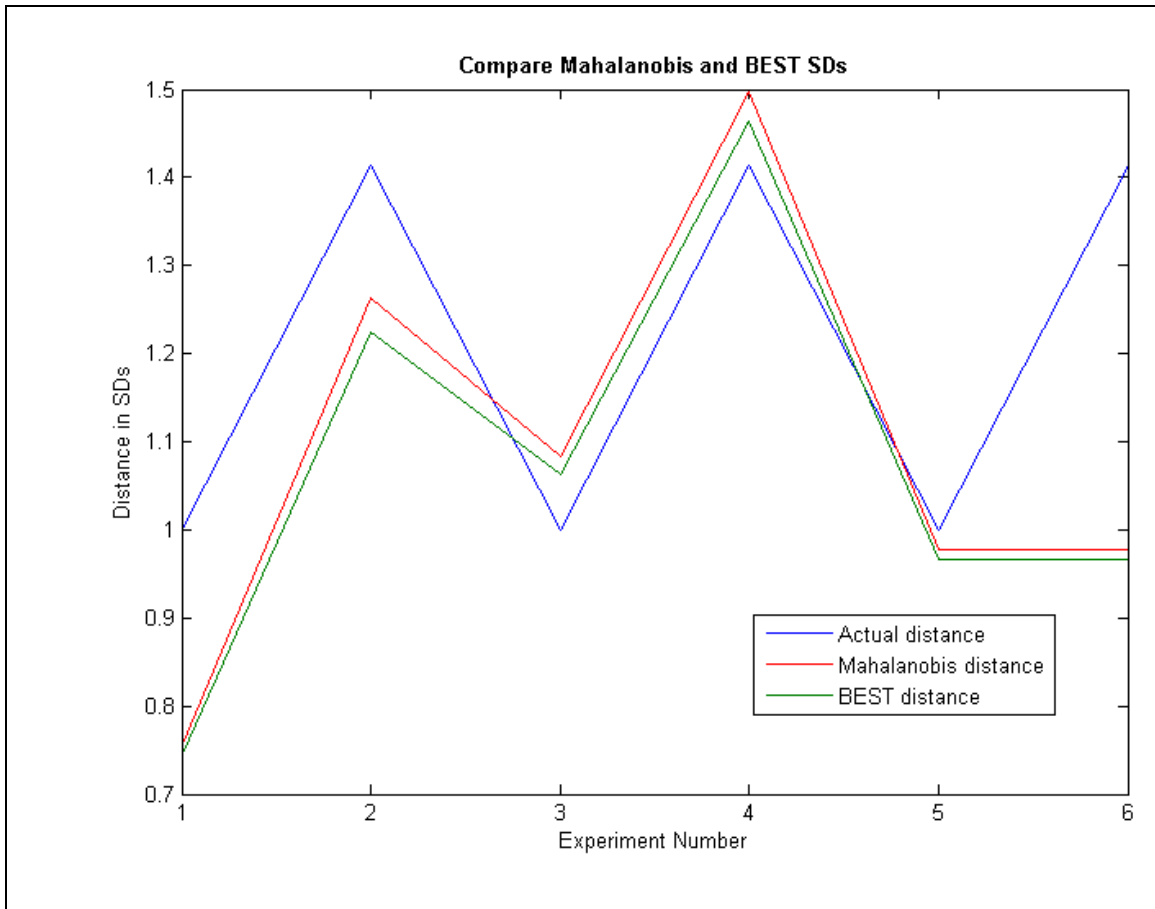


Figure 34: Comparison of Mahalanobis and BEST to Actual Distance when Dimension is Low and Number of Samples is High

With only 2 dimensions (independent variables) and 100 samples from a synthetic $N(0,1)$ distribution, the accuracy of the BEST metric is about the same as the Mahalanobis metric. The error in the red and green lines comes from the 100 randomly selected samples poorly representing the $N(0,1)$ distribution. The variation from run to run largely results from differences in the training set.

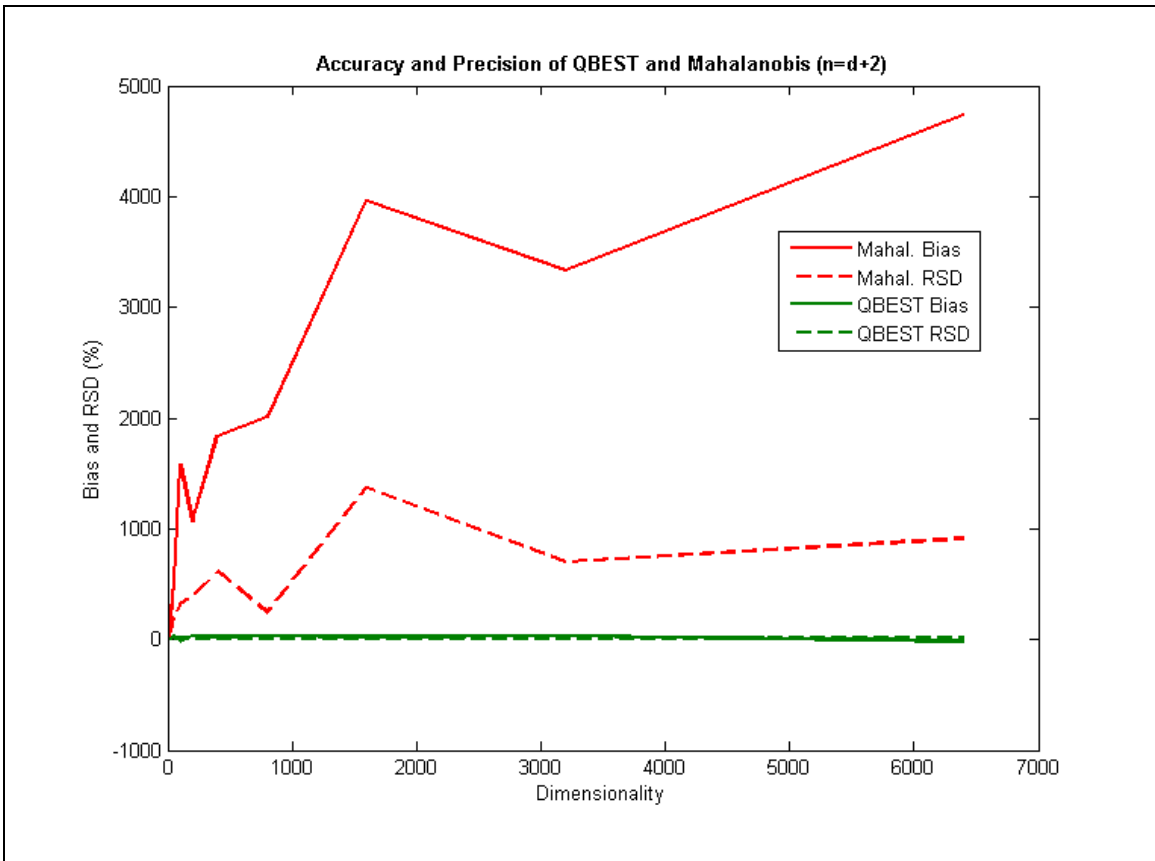


Figure 35: Accuracy and Precision of the BEST and Mahalanobis Metrics Using an $N(0,1)$ Synthetic Data Set

In this experiment, the number of samples (rows) always exceeded the number of dimensions (columns) by 2. The Mahalanobis metric can be very wrong.

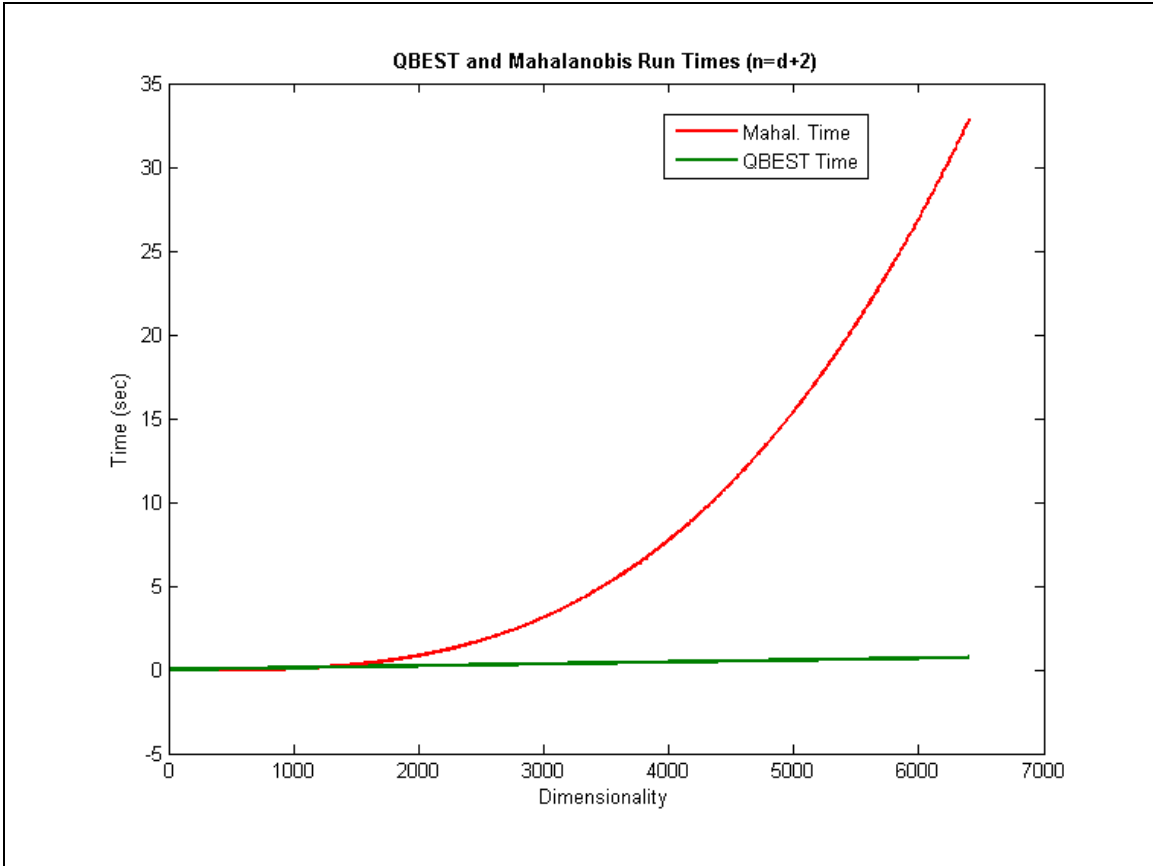


Figure 36: Run Time of Mahalanobis and QBEST as Dimension Increases
 The running time of the BEST calculation is n^2 faster than the Mahalanobis even without parallelization.

Discussion:

Equal Precision for Small and Large Datasets

Relative error is slightly decreased when the number of training points input are increased, but precision as measured by the relative standard deviation of the output is unaffected. This indicates that QBEST is capable of calculating multivariate standard deviation units with predictable error for both small and large datasets. QBEST is ideal for extrapolation from small datasets, as it is able to process these with nearly the same precision as large datasets.

Influence of Bootstrap Replication

For elliptical data, bias remains constant regardless of the the number of bootstrap replications, but relative standard deviation of the algorithm's output is decreased by increasing the number of bootstrap replications run on the data. This indicates that the

number of bootstrap replications does not significantly affect the accuracy of the algorithm's measurements, but by increasing the number of bootstrap replications, the user can ensure greater precision and a more consistent output from run to run.

Influence of Dimension

For data with very large numbers of dimensions, it is important that the training set size be increased to offset increasing bias.

Influence of Radfrac

For this elliptically-distributed data that has no concavities, bias remains constant regardless of the fraction of points used in generating the hypercylinder. The effect of radfrac, or fraction of points used to generate the hypercylinder on relative standard deviation is low, as is to be expected with symmetrically-dispersed data. In other words, it is safe to use a radfrac of 1.0, all of the data points in the cluster, for elliptical data.

Relative Error

Bias remains constant regardless of the the number of bootstrap replications performed or the fraction of points used in generating the hypercylinder (when considering elliptical data that has no concavities).

Increasing the size of the training spectrum appears to decrease the bias slightly, though it is difficult to say how much of this decrease is due to the inborn error of the training spectrum generation by the `mvrnd()` command and how much of the decrease is due to increased competency when the algorithm bootstrap replicates from a larger training set of data. Increasing the number of dimensions of the datapoints in question while holding other factors constant appears to slightly increase bias.

Relative Standard Deviation (RSD)

The algorithm's precision was calculated by dividing the standard deviation of the six runs by the average sds of the six runs. Relative standard deviation is decreased by increasing the number of bootstrap replications run on the data. Increasing the size of the training spectrum or increasing the number of dimensions of the data appears to have no significant effect on the relative standard deviation of the output of the algorithm. This indicates that the algorithm handles with equivalent precision increasingly multivariate datasets, as well as datasets for which more or less training points were input.

The effect of radfrac, or fraction of points used to generate the hypercylinder on relative standard deviation is low in proportion to the effect of bootstrap replication for data that is elliptical, or lacking in concavities. Non-linear functions, such as those seen in acoustic data, can create data clusters with concavities (clusters that appear shaped like the letter "c"). In these instances, the influence of radfrac should theoretically be quite great.

Further research should be done on data clusters including concavities to determine the nature of this effect.

Influence of Training Spectrum

The most important factor that influences relative standard deviation of the QBEST output is the quality of the training set. The particular points chosen for the training set can cause a great deal of variability in the relative standard deviation of the output. This is to be expected, because good results require good data input.

CONCLUSION

Summary of Chapters

Chapter 1

Summary: Chapter 1 introduced the QBEST method and compared it in general to the Mahalanobis method.

QBEST is a statistical algorithm for discriminant cluster analysis, and is the basis for qb.m, a program written in Matlab (and currently being translated into Python 2.7) for testing whether a test sample varies from a training set, library of spectra, or other multivariate data. A bootstrap replication method is used to generate a hypercluster from training spectra samples and the QB program calculates the probability that the test sample is the same as the training samples by measuring the distance of the test sample from the center of the hypercluster in multidimensional skew-adjusted nonparametric standard deviations. The method allows for differential directional weight of different variables.

Chapter 2

Summary: Chapter 2 outlined the method of research for the following chapters and discussed the application of quantile bootstrap (QB) methods to the problem of estimating the No Observed Adverse Effect Level (NOAEL) of a New Molecular Entity (NME) to anticipate a safe starting dose for beginning clinical trials. An estimate of the NOAEL from the extended QB method (called the QNOAEL) can be calculated using multiple disparate studies in the literature and/or from laboratory experiments. The QNOAEL of ellagic acid was calculated using this method in Chapter 4, and is the ADI (Acceptable Daily Intake).

Significance: The QNOAEL is similar in some ways to the Benchmark Dose (BMD) and is superior to the BMD in others. The Benchmark Dose method is currently widely used in toxicological research. Results are used in a meta-analysis simulation based on nonparametric cluster analysis methods to calculate confidence levels on the difference between the Effect and the No Effect studies. The QNOAEL simulation generates an intuitive curve that is comparable to the dose-response curve. The QNOAEL of ellagic acid (EA) was calculated in Chapter 4 for clinical trials of its use as a component therapeutic agent (in BSN476) for treating Chikungunya infections. This is the first application of QB to the problem of NOAEL estimation for a drug.

Chapter 3

Summary: Chapter 3 describes the determination of the Estimated Daily Intake (EDI) of ellagic acid. The purpose of this paper was to determine how much ellagic acid is consumed daily by the average American to serve as a baseline safe dose for human clinical trials. The United States Food and Drug Administration (FDA) describes food stuffs as “Generally Recognized as Safe” (GRAS) at some level (the ADI or less). For example, many GRAS products are classified as dietary supplements, which require less testing than drugs. Products that are GRAS typically do not require the manufacturers to repeatedly conduct the same battery of toxicity testing and clinical trials that drugs need to have to be FDA-approved.

Ellagic acid is a major component of BSN476. It is also a naturally-occurring polyphenol found in fruits and nuts. In this paper, the mean EDI of ellagic acid was determined to be 69.58 micrograms/person/day or 1.05 micrograms/kg body weight/day. The 90th percentile EDI was determined to be 258.33 µg/person/day or 3.89 µg/kg body weight/day, far less than the ADI and QNOAEL. This provides support for starting any human clinical trial of BSN476 at a daily dose containing a comparable amount of ellagic acid to the 95th percentile of ellagic acid consumption. Ellagic acid is safe in the food supply at a dose of 3.89 micrograms/kg body weight/day, so as long as there are no interactions between other components of the drug, there should be no concerns of toxicity to humans. Having demonstrated the expected daily intake of ellagic acid, there is little reason for any initial toxicity tests of ellagic acid before studies proceed to clinical trials.

Significance: The 90th percentile daily intake of a chemical occurring naturally in the food supply should be indicative of a safe dose for PK/PD. This number provides evidence for the level of a compound as Generally Recognized as Safe and represents an upper threshold for the starting dose in human clinical trials. This data analysis indicated that the 90th percentile expected daily intake of ellagic acid is 3.89 micrograms/kg body weight/day. Future PK human trials therefore should be started at a dose equal to or less than 3.89 micrograms/kg body weight/day.

Chapter 4

Summary: In Chapter 4, the acceptable daily intake has been extrapolated from animals studies treated with the QBEST to yield a QNOAEL of 32.8 mg/kg body weight/day. The ADI reinforces the assertion that conducting a PK trial below the EDI should be a safe approach to formulation development. Chapter 3 demonstrated that ellagic acid is a safe component of the food supply up to a dose of 3.89 µg/kg body weight/day (which was determined as a safe dose because it is the 90th percentile daily intake of ellagic acid from the food supply). In this chapter, the acceptable daily intake of ellagic acid was calculated to be 32.8129 mg/kg/day. Human clinical studies should be started at a

value below the 90th percentile estimated daily intake, well below the QNOAEL. Taken together, these two chapters form the basis for an exploratory IND filing at the FDA to expedite formulation testing for human clinical trials.

Significance: More than simply providing this specific opportunity to expedite the testing of BSN476 and to avoid early costly toxicity testing of a formulation, this chapter is important because it specifically provides proof of concept that QBEST and discriminant cluster analysis can be applied to obtain QNOAELs. QBEST is superior to the current standard practice of safety data analysis. It allows for the meta-analysis of data that has already been generated by numerous small studies, increasing the statistical power of small studies by bootstrap replication from numerous small studies. The QBEST method of generating QNOAELs clearly solves numerous scientific issues encountered when calculating a traditional NOAEL. QBEST can be used to determine with statistical confidence limits the highest dose of a drug that will cause no adverse effect; traditional NOAELs are determined by simply increasing an administered dose by arbitrary increments until an adverse effect is noted. Whatever highest dose was administered that is less than the lowest adverse effect dose is considered to be the NOAEL. NOAELs can never be determined to be anything other than an exact dose which was administered, so the NOAEL determination is imprecise. Furthermore, it is difficult to meta-analyze NOAELs, so to increase the precision of a NOAEL test, a completely new test would be required.

The current alternate method to safety determination is the Benchmark Dose Method (BMD). The BMD is determined by administering increasing doses of a pharmaceutical until a specific physiological response is elicited. The benchmark dose is considered to be the dose at which it is likely that a given response will occur. The BMD can be used to meta-analyze multiple studies, but can generate confusing and non-intuitive response curve when there are conflicting results of studies. QBEST solves this problem by generating a single QNOAEL from an array of studies in a simulation. Finally, QNOAELs remove the subject bias from study weighting that accompanies any meta-analysis. The bootstrapping algorithm causes there to be no need to subjectively assign weights to studies. QBEST naturally causes the typical studies to outweigh the outlier studies through a simple rule of probability.

Chapter 5

Summary: Chapter 5 covered the Round-the-Compass Rose experiment which validates that the QBEST is capable of handling non-spherical distributions of data, like that used in the ADI calculation, accurately and reproducibly. It also proves the concept that this algorithm can be used for very large datasets even on a personal laptop computer, unlike the Mahalanobis distance.

The purpose of this experiment was to explore sources of bias in the output of the QB program. To that end, synthetic data with known parameters were generated and the output number of distances in standard deviation units calculated by the QB program was analyzed for consistency and accuracy. Furthermore, symmetrical elliptical clusters of data were analyzed in order to determine whether compass location around the cluster significantly affected consistency or accuracy of the algorithm's output.

It was determined that the most significant factor that affects the algorithm's accuracy is the training spectrum used, which makes intuitive sense. Accurate results require good input data. The training set must accurately reflect the population that it seeks to represent. Average bias is largely unaffected by the number of bootstrap replicates used. Percent error is largely unaffected by the training set size. QBEST is capable of calculating with accuracy even when input parameters are at the extremes. Relative standard deviation of the algorithm's outputs is reduced when the number of bootstrap replicates is increased.

Significance: The specific aim of Chapter 5 is to evaluate the accuracy and precision of the QB Simulation. The differential directional weight is a key attribute of the QB program, as the direction of skew of a sample from the training sample may shed light on the nature of an impurity. This concept can be illustrated by the following: in an elliptical cluster of training points, there is a "long axis" - that is, an axis in which greater deviation can be expected from the mean, as well as a "short axis" along which training points are tightly clustered and little variation is expected of a pure sample in keeping with the training samples. The QB program is able to analyze points along any axis of any n-dimensional cluster of training points and determine whether significant skew exists along any of those axes. Before this work, only data clusters having an equivalent standard deviation along all vectors - that is, circular, spherical, or similar n-dimensional data clusters - had been analyzed. This work was the first to analyze the performance of QBEST when faced with clusters having non-equal standard deviations along two axes. The differential handling of data on the left versus right-hand side of the center of the cluster is the most important improvement of QBEST over the Mahalanobis equation; this experiment paves the way for future experiments exploring data that are asymmetrically distributed on either side of the center (unlike the ellipses used in this experiment).

Future Applications:

QBEST represents a significant improvement over the Mahalanobis equation for the nonparametric analysis of multivariate data. The most notable improvements are (1) that QBEST calculates standard deviation units separately for points to one side or the other of center (asymmetrical determination of standard deviation units); (2) QBEST is mathematically capable of determining with accuracy the QBEST distance even when

the number of observations are fewer than the number of dimensions or variables of the data; and (3) QBEST requires d^2 less computing power than the Mahalanobis equation. These improvements are functionally important, respectively, because:

- 1) Calculating the symmetrical Mahalanobis distance introduces inaccuracy when the cluster being evaluated is asymmetrically distributed. The Mahalanobis equation is insufficient for evaluating data that are not symmetrically distributed along any given line. Real-world data does not tend to be symmetrically distributed. Furthermore, the tailing observed in data clusters can show important information about a dataset. It is inappropriate to ignore the distribution of points differently along the “left and right hand” vectors from the center of the data distribution to a given test point.
- 2) It is mathematically impossible to utilize the Mahalanobis equation to determine a Mahalanobis distance when the number of datapoints or observations are far less than the number of dimensions or variables measured describing the data. The Mahalanobis equation relies on matrix factorization that cannot be performed in this instance. This means that the Mahalanobis equation cannot be utilized to evaluate, for example, all of the data gathered by a handful of clinical trials with one hundred endpoints each; the infrared spectra of a hundred duplicate chemical samples taken at ten-thousand wavelengths; or the massive amounts of data processed by an artificial intelligence program that it uses to make decisions. When the number of variables is close to the number of observations, but still exceeds the number of observations, QBEST is often more accurate. As the number of observations approaches the number of variables, the Mahalanobis distance becomes increasingly inaccurate. The QBEST Equation is capable of solving these mathematical problems and as yet unimagined others as our world rapidly becomes dependent on larger and more complicated massive data sets. In this way, QBEST is the statistical equation of the future.
- 3) Even if the number of observations required by the Mahalanobis equation is met, the number of dimensions of data that can be analyzed in a reasonable amount of time is severely limited. The Mahalanobis equation is an order of d^3 algorithm, which means that for every d dimensions increased in the data, the computer will need d^3 more computing power to solve for the Mahalanobis distance. In comparison, QBEST is an order of d algorithm. Computing power equates to run time for a given processor. Simply stated, a QBEST distance that could be run in an afternoon on a typical personal computer would have to be run on a supercomputer to be solved with the Mahalanobis equation in the same amount of time. In Chapter 5, it was determined that a typical laptop personal computer can calculate the QBEST distance for data with 6×10^6 dimensions in approximately twenty minutes or less. It is difficult to imagine any current pharmaceutical application which would require 6×10^6 dimensions of data, but this is an emerging problem in the physical and computer sciences. However, as medical monitoring becomes more advanced, an increasing amount of data will

be generated and stored, particularly in the case of biological parameters which are continuously monitored. The analysis of increasingly massively multivariate data is a serious emerging problem in statistics. QBEST provides the best solution to this problem at present.

Applications in Pharmaceutical Science:

As pharmaceutical science and computer science progress, Big Data continue to get bigger. Unfortunately, computing power is not improving as quickly as the capability to generate information, leaving scientists and businessmen alike with an information surplus. Before QBEST, there was not an algorithm capable of discriminant cluster analysis on massively multivariate data that could be run on a personal computer. In practical terms, QBEST can process with accuracy a handful of instances of million-dimensional data. This capability will allow scientists at startup research companies and university researchers, including students, to perform the kind of calculations that traditionally required the use of a supercomputer.

At present, most pharmaceutical data are generally not considered to fall under the umbrella of Big Data, or databases consisting of hundreds of thousands of datapoints with thousands of variables, but the trend of technology in medicine is moving toward Big Data. Currently, insurance databases and hospital databases do fall into the category of big data. Pharmaceutical companies, health care providers, and public policy makers draw on these massive databases to make generalizations about populations and enact policy changes. The push towards personalized medicine will also require the analysis of big data problems increasingly in the future.

Big Data in Health Insurance

Major insurance providers now complement their own massive databases by purchasing additional databases from commercial vendors. This strategy is necessary to be competitive in the modern market². The decisions that insurers make regarding care has a tremendous impact on the treatment chosen for a given patient, particularly as the industry pushes wellness incentives as a cost-saving measure.

Humana's proprietary algorithm, the Anvita Engine™, analyzes basic wellness measurements, such as weight, heart rate, BMI, etc. measured over time and makes predictions about possible chronic conditions that the patient could develop, based on patterns of disease progression noted in a training sample of patients^{5,6}. The algorithm then places flags in the patient's online medical file that appear to any physician who bills the insurance^{5,6}. These flags suggest that the doctor perform various routine tests or recommend preventative medicine for the patient. If performed, the treating physician is paid a fee by the insurer^{5,6}. This is only one of the first generation of algorithms storing and comparing massively multivariate patient data to make treatment decisions for a

patient. These databases grow larger and more sophisticated each year. As the scope of data analysis increases, it is necessary to have statistical algorithms capable of keeping pace while making accurate decisions for patient care.

On the surface, algorithms such as Humana's Anvita Engine appear to work using only information recorded by a patient's doctor. In reality, major insurers pool information from all of their patients for training sets, and actually purchase databases of outside information from commercial vendors and compile this information when making predictions for a given patient. Some examples of the types of databases bought by major insurers are as follows:

- Store loyalty card information - Major grocery store chains maintain records of grocery purchases. Healthy and unhealthy foods, cigarettes, alcohol, pregnancy tests, prophylactics, prescriptions, over-the-counter medicines, and pharmacy purchases made at grocery store chains using the store loyalty or discount card are stored in the corporate database. Grocery store chains such as Kroger, Safeway and Target use this information when sending out personalized coupons and deals. There is a great deal of money to be made in predicting certain medical conditions, and then in marketing to that prediction. Target has openly admitted to predicting which of its customers are pregnant as early as possible so that it can mail coupons for large-ticket baby necessities, such as cribs, before the shopper has the opportunity to purchase those items from a competitor⁴. In fact, Target's loyalty card algorithm is so accurate that it has been documented to have predicted a woman's pregnancy before her family was made aware of it⁴. A Target spokesperson has confirmed that Target's shopper algorithm can predict not only pregnancy, but also the trimester of a customer's pregnancy with accuracy from a single shopping visit⁴. Pregnancy is not the only condition that a store chain has the impetus to monitor; consider that major retailers of exercise equipment, orthotics, and prescription medications also collect shopper data.
- Social Media - Some insurance companies are purchasing databases from social media sites such as Facebook. One Facebook algorithm keeps track of a user's mood based off of whether his posts seem to be positive or negative^{7,8,9}. In a recent scandal, Facebook was accused of performing unethical research on unwilling subjects by altering the number of positive or negative posts a user saw on his news feed and then monitoring the user's mood over time^{7,8,9}. Social media posts have the potential to provide incredible insight into the risk of insuring a potential policyholder, as they include information about the person's habits, including drinking, tobacco use, recreational drug use, mental health, sexual activity and orientation, current health conditions, and all of the aforementioned information as it relates to the user's relatives. Harnessing Facebook's mood algorithm could prove extremely valuable in the treatment of depression, suicidal tendencies, and other mental disorders.
- Phone Applications or "Apps" - Numerous applications exist for smartphone users wishing to monitor their personal health. Some are free, while others are

paid services that collect and analyze daily health information. Various insurers have contracts with various app makers to receive databases of these user reports².

- Commercial Fitness Devices - According to a recent report, John Hancock has a standing contract with Vitality that integrates wellness initiatives into its life insurance policies². Policyholders can earn discounts on their life insurance policies if they register internet-connected commercial wellness devices, such as FitBits². In this way, John Hancock is using big data analysis when analyzing basic vital signs of policy holders and comparing them to standards of healthy and at-risk patients².
- Workplace Wellness Programs - Workplace wellness programs are company-instituted programs that encourage employees to increase their fitness¹. These programs are largely unregulated with very lax privacy laws, as they are not subject to HIPAA or to the provisions of Obamacare that prevent premium discrimination on the basis of pre-existing conditions¹. Increasingly, workplace wellness program data are being used to increase the costs of company insurance by up to 30% for overweight or otherwise unhealthy non-tobacco using employees¹. Associations such as the American Diabetes Association warn that under the guise of giving monetary incentive for healthy employees, employers are legally penalizing people with chronic conditions such as diabetes by increasing the cost of their health insurance, leading to decreased access to medical care for people with chronic and pre-chronic conditions¹.

Algorithms owned by data-mining companies piece together and re-identify consumer information, attempting to pair consumers to their respective identities as insurance customers. This is an easy task in the case of store loyalty cards, social media accounts, and personal health devices, since these things are tied to consumer names, addresses, and/or phone numbers. However, for those databases of truly de-identified data, there are algorithms capable of re-identifying particular consumers. In this way, the insurance company stores a comprehensive database of not only strict health information, but also related dietary health, wellness, and behavioral risk information for each of its customers. There is a large push in the industry now to develop newer and more powerful methods of analyzing and using this data to increase profits, either by adjusting policy payments for at-risk policy holders, or by implementing wellness programs that include preventative medicine and health screenings. Additionally, de-identified but localized data can shed light on the health of particular populations, such as the alcohol consumption across a particular zip code, or the purchases of populations of various racial compositions.

Big Data in Personalized Medicine

Personalized medicine, the process of tailoring treatment to the individual patient on the basis of metrics relating to that patient, is also an emerging necessity in the industry. Some pharmaceutical devices have already begun to move toward the storing of continuous health data for patients; this is most clearly seen in the blood glucose monitoring systems and paired insulin pumps that have been developed for Type I diabetics. Recall from Chapter 1 that any variable with continuous monitoring can be represented as a multivariate data point with coordinates at each time point (or wavelength in the case of spectral data). As more health endpoint monitors become continuous (such as heart rate monitors, blood content monitors, physical exercise monitors, and neurological activity monitors), more data will be generated for each individual patient, which can be utilized in care decisions. Even now, Fitbits™ and smart devices continuously monitor basic vital signs. Most of these devices have their own apps that can be synced with physician's offices, but there is a large push to make a universal record keeping system that can store the information generated by all of these devices and make comparisons across the patient population.

This massive amount of information is useless, however, unless baselines are established for a particular patient and deviations from that baseline are monitored. QBEST answers the important question: Does the current state differ significantly from the baseline training set of data? QBEST allows a chunk of daily data - either in blood glucose patterns or general vital signs - to serve as a training set, while the present measurements could be used as a test set. QBEST would then detect whether there was a significant difference between the spectra of patient's current status and the blood glucose spectra corresponding to the last finger stick. It could also compare a significantly different result with training sets of known glucose levels or disease states and suggest what problem was developing in the patient.

Big Data in Safety Analysis

Finally, QBEST has the potential to revolutionize analysis of toxicology data. Risk is an important factor in the regulation of pharmaceuticals. Initially, determinations of risk vs. benefit form the basis of FDA decisions regarding the approval and marketing of a new pharmaceutical, including its on-label usage, dose, and black box warnings. Post-marketing, risk plays strongly into civil suits when an injury has occurred, allegedly caused by a pharmaceutical agent. QBEST offers a method for the statistical determination of risk. As demonstrated in Chapter 4, QBEST can be used to determine a 98% confidence limit QNOEL of a drug, or a dose that would cause no observed effects 98% of the time if the same studies were to be repeated. Pre-marketing, QBEST can be used to inform executive decisions regarding the safety of a pharmaceutical and, as demonstrated in this work, can help to clarify the therapeutic window.

Post-marketing, QBEST can assist in retrospective analysis of what was known about the safety of a pharmaceutical agent at the time it was administered (a key component of pharmaceutical and other toxic exposure lawsuits) by simply performing an analysis with all data published at the time of the incident.

Applications in Computer Science and Robotics

Future applications of the QBEST will likely involve Artificial Intelligence and adaptive resonance. When does a neural net need to be re-trained? QBEST could provide an answer. With the QBEST, it is now possible to solve mathematical problems not even dreamed of before. QBEST's efficient computation makes it an ideal algorithm that can compute six million bootstrap replications, find significant differences in data with one million dimensions, or process 5 million training data points on a cheap laptop computer. This convenient method of quickly analyzing health data will become integral in personalized medicine.

Weaknesses of QBEST and Further Work that Must Be Done

Though QBEST represents a significant improvement over the Mahalanobis equation and fulfills a similar purpose, there are still some datasets which would be handled very ineffectively by QBEST. Nonlinear data sets present potential problems. Take for example the figure below:

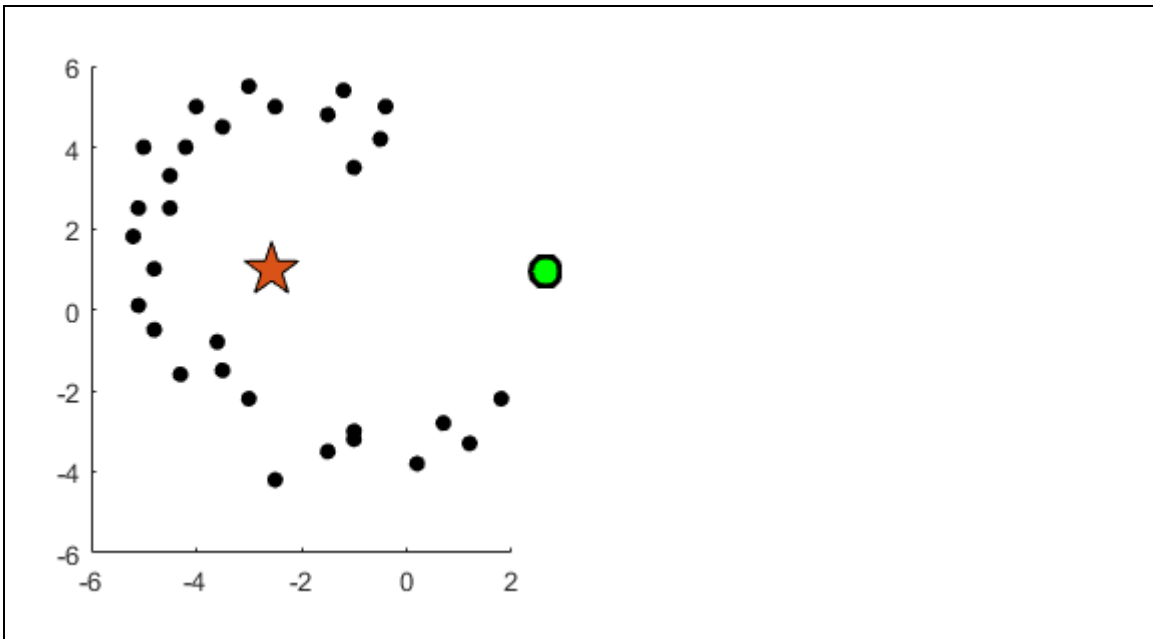


Figure 37: An Example of a Poor Dataset for Application of QBEST

The above data cluster has an obvious concavity. The red star marks the center of the cluster as calculated by QBEST and the green point is a theoretical test point. Note that if a QBEST “rubber ruler with a nail in the center” were drawn from the center to the green test point, no other points would fall along the ruler unless the width of the ruler (radfrac) was very great.

In the above example, QBEST would generate a poor QBEST distance between the calculated center and the green test point. There is not a linear relationship between the x axis and the y axis. Use of QBEST is not advisable when the test point falls on the other side of a concavity from center in the data. It is also not advisable to use QBEST when an obvious visual pattern in the data would indicate that there are zones where points should not ever fall in space (for example, if data were following a sine curve, there would be obvious concavities in the data following a clear visual pattern). These types of nonlinear patterns are difficult for the algorithm to handle because QBEST assumes that the center of the data cluster actually falls within the data cluster. Before using QBEST, the analyst should graph the data and visually inspect the clusters. As with any statistical analysis method, the analyst should plot all combinations of variables against one another and determine that nothing contraindicates use of the QBEST method.

It has been suggested by a colleague that QBEST could be improved upon by defining the center of the data cluster differently¹⁰. Rather than finding the average coordinate value on each axis and defining that point as the center (as was done in the example above), it is possible to define coordinate space as the area inside of a multidimensional figure¹⁰. Then, the calculated center of the data would be forced to fall within the defined

coordinate space¹⁰. In this way, the center of the data could never be calculated to fall within a concavity of the figure¹⁰. This method is in keeping with the Principal Curve Method, which calculates a center curve through a gray zone and measures points back orthogonally³. A similar definition of coordinate space could be used in the QBEST distance calculation. QBEST is not currently written to do this, but could be. However, it would probably be better to use a central curve (or a 1-D manifold) as in Hastie et al³. See Figure 38, below.

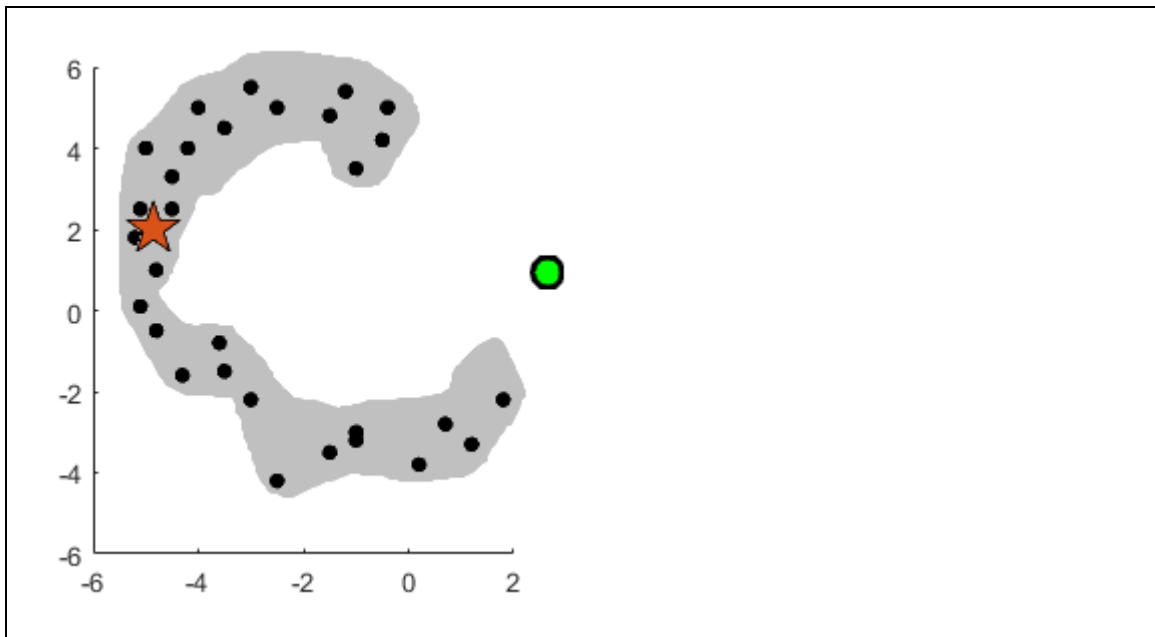


Figure 38: An Example of Coordinate Space Redefined for Purposes of Calculating the Center of a Data Cluster

It should be noted, however, that QBEST can in fact be used to calculate a QBEST distance in the given example (Figure 37). The test spectrum is defined as the point (2,1). The QBEST distance is 2.2976 standard deviation units, with 2.8302 standard deviation units of skew.

APPENDICES:

Appendix A: Equations

The Mahalanobis equation:

The Mahalanobis distance of an observation $\vec{x} = (x_1, x_2, x_3 \dots x_N)^T$ from a set of

observations with mean $\vec{\mu} = (\mu_1, \mu_2, \mu_3 \dots \mu_N)^T$

and covariance matrix S is defined as:

$$D_M(\vec{x}) = \sqrt{(\vec{x} - \vec{\mu})^T S^{-1} (\vec{x} - \vec{\mu})}$$

Equation Source:

R. De Maesschalck, D. Jouan-Rimbaud, D.L. Massart. "The Mahalanobis distance." Chemometrics and Intelligent Laboratory Systems. Volume 50, Issue 1, 2000, Pages 1-18. ISSN 0169-7439. <[https://doi.org/10.1016/S0169-7439\(99\)00047-7](https://doi.org/10.1016/S0169-7439(99)00047-7)>

Appendix B: Computer Programs

QBEST Algorithm:

REPLICA

```
function [BTRAIN,CENTER]=replica(TNSPEC,B)
% TNSPEC=training spectra, B=number of replicates desired.
% REPLICA outputs BTRAIN replicates, and the center of the replicates in CENTER
% "Copyright 2003 Robert A. Lodder & Bin Dai"
[N,D]=size(TNSPEC);
BTRAIN=zeros(B,D);
CENTER=zeros(D,1);
BSAMP=zeros(N,D);
PICKS=rand(B*N,1);
index=find(PICKS==1);
PICKS(index)=0.9999;
PICKS=reshape(PICKS,B,N);
PICKS=fix(N*PICKS+1);
for l=1:B
BSAMP=TNSPEC(PICKS(l,:),:);
BTRAIN(l,:)=sum(BSAMP)/N;
end
BTRAIN;
CENTER=sum(BTRAIN)/B;
```

SOB

```
function [ECDF,TCDF,R]=SOB(BTRAIN,CTN,NTN,TESTSPEC)
% This routine takes 2 spectral groups (with = numbers of samples)
% and calculates an ECDF and TCDF for QQ plotting. The routine
% also returns a correlation coefficient between the 2 CDFs.
% Training replicates and center must be provided, along with
% test spectra. BNUM = the number of replicates used.
% Note: you may have to set EPSILON(1e-200) or greater to prevent
% underflow errors in high dimensional hyperspaces
% Copyright 2003 Dr. Robert.A.Lodder & Bin Dai
[NTEST,C]=size(TESTSPEC);
[BTEST,CTEST]=replica(TESTSPEC,NTN);
[M,C]=size(BTRAIN);
p=0.1;
CTNN=repmat(CTN,M,1);
ST=(sum(((BTRAIN-CTNN).^C)).^(1/C);
CTNN=repmat(CTN,NTN,1);
SX=(sum(((BTEST-CTNN).^C)).^(1/C);
M=2*M;
TRIM=fix((M*p)+1):fix(M-(M*p));
ECDF=[ST,SX];
TCDF=[ST,ST];
ECDF=sort(ECDF);
TCDF=sort(TCDF);
ECDF=ECDF(TRIM);
TCDF=TCDF(TRIM);
R=corrcoef(TCDF,ECDF)
```

Estimated Daily Intake Programs:

EDI

```
function [zarray] = EDI(serving_sizes,dose,age,day1,day2,mass)
%% This function solves for Expected Daily Intakes (EDIs) using the NHANES and
FNDDS databases.
% This program requires that the "NHANES basic startup workspace.mat" workspace be
present.
% The programs searchnzero, searchfoods.m, and searchsizes.m must all be in
% the working directory path.
% serving sizes = in rows to target_codes, 1st col.=food codes, 2nd
%          col. = serving sizes in g
% dose is a scalar variable specified by the client
% age, day1, day2, and mass are all in the "NHANES basic startup workspace.mat"
workspace
% zarray returns the mean mass of all ages just to put something into the output variable
and avoid an error
%
% Start with all patient sequence numbers in age groups
% Time to make food code tables for each group
%
% loop through Day1 and check each row to see if the SEQN in the first column is in the
set of SEQN for each age group
% if so, leave the row as is. if not, then replace the row with all zeros and eliminate the
zeros later
% get all foods consumed by each age group
target_codes = serving_sizes(:,1); % target_codes is a vector food codes of length =
number of rows in serving_sizes
% find the infant groups
ind_age0_1 = find((age(:,2)>0)&(age(:,2)<13)) ;
age0_1=age(ind_age0_1,1);
disp('Number of infants aged 0-1')
size(ind_age0_1)
ind_age1_2 = find((age(:,2)>14)&(age(:,2)<24));
age1_2=age(ind_age1_2,1);
disp('Number of toddlers aged 1-2')
size(ind_age1_2)
ind_age2_5 = find((age(:,2)>23)&(age(:,2)<61)); % create indexes for age groups, ages
are given in months, not years
age2_5=age(ind_age2_5,1);
disp('Number of subjects aged 2-5')
```

```

size(age2_5)
ind_age6_12 = find((age(:,2)>62)&(age(:,2)<145));
age6_12=age(ind_age6_12,1);
disp('Number of subjects aged 6-12')
size(age6_12)
ind_age13_19 = find((age(:,2)>146)&(age(:,2)<229));
age13_19=age(ind_age13_19,1);
disp('Number of subjects aged 13-19')
size(age13_19)
ind_age20_up = find((age(:,2)>229));
age20_up=age(ind_age20_up,1);
disp('Number of subjects aged 20 and up')
size(ind_age20_up)
% find all ages > 2 years
ind_age2up = find((age(:,2)>23)&(age(:,2)<2000));
age2up=age(ind_age2up,1);
disp('Total number of subjects aged 2 and up')
size(ind_age2up)
% Day 1, delete entries not in specified age group
day1_age0_1 = searchnzero(age0_1,day1);
day1_age1_2 = searchnzero(age1_2,day1);
day1_age2_5 = searchnzero(age2_5,day1);
day1_age6_12 = searchnzero(age6_12,day1);
day1_age13_19 = searchnzero(age13_19,day1);
day1_age20_up = searchnzero(age20_up,day1);
day1_age2up = searchnzero(age2up,day1);
% Day 2
day2_age0_1 = searchnzero(age0_1,day2);
day2_age1_2 = searchnzero(age1_2,day2);
day2_age2_5 = searchnzero(age2_5,day2);
day2_age6_12 = searchnzero(age6_12,day2);
day2_age13_19 = searchnzero(age13_19,day2);
day2_age20_up = searchnzero(age20_up,day2);
day2_age2up = searchnzero(age2up,day2);
% throw away all food codes for each age's dietary data that are not in the targeted food
code list
%
day1_fc_age0_1 = searchfoods(day1_age0_1,target_codes);
day1_fc_age1_2 = searchfoods(day1_age1_2,target_codes);
day1_fc_age2_5 = searchfoods(day1_age2_5,target_codes);
day1_fc_age6_12 = searchfoods(day1_age6_12,target_codes);
day1_fc_age13_19 = searchfoods(day1_age13_19,target_codes);
day1_fc_age20_up = searchfoods(day1_age20_up,target_codes);

```

```

day1_fc_age2up = searchfoods(day1_age2up,target_codes);
%
day2_fc_age0_1 = searchfoods(day2_age0_1,target_codes);
day2_fc_age1_2 = searchfoods(day2_age1_2,target_codes);
day2_fc_age2_5 = searchfoods(day2_age2_5,target_codes);
day2_fc_age6_12 = searchfoods(day2_age6_12,target_codes);
day2_fc_age13_19 = searchfoods(day2_age13_19,target_codes);
day2_fc_age20_up = searchfoods(day2_age20_up,target_codes);
day2_fc_age2up = searchfoods(day2_age2up,target_codes);
%
disp('Number of subjects eating target food on Day 1 aged 0-1')
size(day1_fc_age0_1)
disp('Number of subjects eating target food on Day 1 aged 1-2')
size(day1_fc_age1_2)
disp('Number of subjects eating target food on Day 1 aged 2-5')
size(day1_fc_age2_5)
disp('Number of subjects eating target food on Day 1 aged 6-12')
size(day1_fc_age6_12)
disp('Number of subjects eating target food on Day 1 aged 13-19')
size(day1_fc_age13_19)
disp('Number of subjects eating target food on Day 1 aged 20 and up')
size(day1_fc_age20_up)
disp('Number of subjects eating target food on Day 1 aged 2 and up')
size(day1_fc_age2up)
%
disp('Number of subjects eating target food on Day 2 aged 0-1')
size(day2_fc_age0_1)
disp('Number of subjects eating target food on Day 2 aged 1-2')
size(day2_fc_age1_2)
disp('Number of subjects eating target food on Day 2 aged 2-5')
size(day2_fc_age2_5)
disp('Number of subjects eating target food on Day 2 aged 6-12')
size(day2_fc_age6_12)
disp('Number of subjects eating target food on Day 2 aged 13-19')
size(day2_fc_age13_19)
disp('Number of subjects eating target food on Day 2 aged 20 and up')
size(day2_fc_age20_up)
disp('Number of subjects eating target food on Day 2 aged 2 and up')
size(day2_fc_age2up)
% scale food consumption of each food in grams to number of servings
% now do Day 1
serv_day1_age0_1 = searchsizes(day1_fc_age0_1,serving_sizes);
serv_day1_age1_2 = searchsizes(day1_fc_age1_2,serving_sizes);

```

```

serv_day1_age2_5 = searchsizes(day1_fc_age2_5,serving_sizes);
serv_day1_age6_12 = searchsizes(day1_fc_age6_12,serving_sizes);
serv_day1_age13_19 = searchsizes(day1_fc_age13_19,serving_sizes);
serv_day1_age20_up = searchsizes(day1_fc_age20_up,serving_sizes);
serv_day1_age2up = searchsizes(day1_fc_age2up,serving_sizes);
% now do Day 2
serv_day2_age0_1 = searchsizes(day2_fc_age0_1,serving_sizes);
serv_day2_age1_2 = searchsizes(day2_fc_age1_2,serving_sizes);
serv_day2_age2_5 = searchsizes(day2_fc_age2_5,serving_sizes);
serv_day2_age6_12 = searchsizes(day2_fc_age6_12,serving_sizes);
serv_day2_age13_19 = searchsizes(day2_fc_age13_19,serving_sizes);
serv_day2_age20_up = searchsizes(day2_fc_age20_up,serving_sizes);
serv_day2_age2up = searchsizes(day2_fc_age2up,serving_sizes);
% now calculate mean Expected Daily Intakes (EDIs) using the mass of additive that the
client wants to add to each food.
% Day 1
disp('Mean Expected Daily Intakes (EDIs)')
mean0_1day1 = mean(serv_day1_age0_1(:,4))*dose; % mg EDI
mean1_2day1 = mean(serv_day1_age1_2(:,4))*dose; % mg EDI
mean2_5day1 = mean(serv_day1_age2_5(:,4))*dose; % mg EDI
mean6_12day1 = mean(serv_day1_age6_12(:,4))*dose; % mg EDI
mean13_19day1 = mean(serv_day1_age13_19(:,4))*dose; % mg EDI
mean20_upday1 = mean(serv_day1_age20_up(:,4))*dose; % mg EDI
mean2upday1 = mean(serv_day1_age2up(:,4))*dose; % mg EDI
% now do day 2
mean0_1day2 = mean(serv_day2_age0_1(:,4))*dose; % mg EDI
mean1_2day2 = mean(serv_day2_age1_2(:,4))*dose; % mg EDI
mean2_5day2 = mean(serv_day2_age2_5(:,4))*dose; % mg EDI
mean6_12day2 = mean(serv_day2_age6_12(:,4))*dose; % mg EDI
mean13_19day2 = mean(serv_day2_age13_19(:,4))*dose; % mg EDI
mean20_upday2 = mean(serv_day2_age20_up(:,4))*dose; % mg EDI
mean2upday2 = mean(serv_day2_age2up(:,4))*dose; % mg EDI
% to make histograms of exposure data here, add a breakpoint and use
% hist(). For example, copy and paste
% bins=round(sqrt(size(serv_day2_age2up(:,4))))
% bins=bins(1)
% hist(serv_day2_age2up(:,4),bins)
%
% average day 1 and day 2
disp('Mean EDI ages 0 to 1')
medi0_1=(mean0_1day1+mean0_1day2)/2
disp('Mean EDI ages 1-2')
medi1_2=(mean1_2day1+mean1_2day2)/2

```

```

disp('Mean EDI ages 2-5')
medi2_5=(mean2_5day1+mean2_5day2)/2
disp('Mean EDI ages 6-12')
medi6_12=(mean6_12day1+mean6_12day2)/2
disp('Mean EDI ages 13-19')
medi13_19=(mean13_19day1+mean13_19day2)/2
disp('Mean EDI ages 20 and up')
medi20up=(mean20_upday1+mean20_upday2)/2
disp('Mean EDI ages 2 and up')
medi2up=(mean2upday1+mean2upday2)/2
% for 90th percentile calculation of Day 1, produces sorted doses in each age group
sort_day1_age0_1 = sort(serv_day1_age0_1(:,4))*dose;
sort_day1_age1_2 = sort(serv_day1_age1_2(:,4))*dose;
sort_day1_age2_5 = sort(serv_day1_age2_5(:,4))*dose;
sort_day1_age6_12 = sort(serv_day1_age6_12(:,4))*dose;
sort_day1_age13_19 = sort(serv_day1_age13_19(:,4))*dose;
sort_day1_age20_up = sort(serv_day1_age20_up(:,4))*dose;
sort_day1_age2up = sort(serv_day1_age2up(:,4))*dose;
%
sort_day1_age0_1_d1 = sort_day1_age0_1(round(.9*size(sort_day1_age0_1)));
sort_day1_age1_2_d1 = sort_day1_age1_2(round(.9*size(sort_day1_age1_2)));
sort_day1_age2_5_d1 = sort_day1_age2_5(round(.9*size(sort_day1_age2_5))); % 90th
percentile
sort_day1_age6_12_d1 = sort_day1_age6_12(round(.9*size(sort_day1_age6_12)));
sort_day1_age13_19_d1 = sort_day1_age13_19(round(.9*size(sort_day1_age13_19)));
sort_day1_age20_up_d1 = sort_day1_age20_up(round(.9*size(sort_day1_age20_up)));
sort_day1_age2up_d1 = sort_day1_age2up(round(.9*size(sort_day1_age2up)));
%
% Day 2 for 90th percentile calculation, produces sorted doses in each age group
sort_day2_age0_1 = sort(serv_day2_age0_1(:,4))*dose;
sort_day2_age1_2 = sort(serv_day2_age1_2(:,4))*dose;
sort_day2_age2_5 = sort(serv_day2_age2_5(:,4))*dose;
sort_day2_age6_12 = sort(serv_day2_age6_12(:,4))*dose;
sort_day2_age13_19 = sort(serv_day2_age13_19(:,4))*dose;
sort_day2_age20_up = sort(serv_day2_age20_up(:,4))*dose;
sort_day2_age2up = sort(serv_day2_age2up(:,4))*dose;
%
sort_day2_age0_1_d2 = sort_day2_age0_1(round(.9*size(sort_day2_age0_1)));
sort_day2_age1_2_d2 = sort_day2_age1_2(round(.9*size(sort_day2_age1_2)));
sort_day2_age2_5_d2 = sort_day2_age2_5(round(.9*size(sort_day2_age2_5))); % 90th
percentile
sort_day2_age6_12_d2 = sort_day2_age6_12(round(.9*size(sort_day2_age6_12)));
sort_day2_age13_19_d2 = sort_day2_age13_19(round(.9*size(sort_day2_age13_19)));

```



```

sort_day2_age20_up_d2 = sort_day2_age20_up(round(.9*size(sort_day2_age20_up)));
sort_day2_age2up_d2 = sort_day2_age2up(round(.9*size(sort_day2_age2up)));
%
% calculate means of day 1 and day 2 90th percentiles
disp('Mean 90th percentiles')
M90_age0_1 = (sort_day1_age0_1_d1 + sort_day2_age0_1_d2)/2
M90_age1_2 = (sort_day1_age1_2_d1 + sort_day2_age1_2_d2)/2
M90_age2_5 = (sort_day1_age2_5_d1 + sort_day2_age2_5_d2)/2
M90_age6_12 = (sort_day1_age6_12_d1 + sort_day2_age6_12_d2)/2
M90_age13_19 = (sort_day1_age13_19_d1 + sort_day2_age13_19_d2)/2
M90_age20_up = (sort_day1_age20_up_d1 + sort_day2_age20_up_d2)/2
M90_age2up = (sort_day1_age2up_d1 + sort_day2_age2up_d2)/2
% now do body mass calculations
% find masses of each age group
mass0_1 = searchnzero(age0_1,mass);
mass1_2 = searchnzero(age1_2,mass);
mass2_5 = searchnzero(age2_5,mass);
mass6_12 = searchnzero(age6_12,mass);
mass13_19 = searchnzero(age13_19,mass);
mass20_up = searchnzero(age20_up,mass);
mass2up = searchnzero(age2up,mass);
% calculate the mean masses of each age group
disp('Mean mass of each age group')
age_group_masses =
[mean(mass0_1(:,2)),mean(mass1_2(:,2)),mean(mass2_5(:,2)),mean(mass6_12(:,2)),m
ean(mass13_19(:,2)),mean(mass20_up(:,2)),mean(mass2up(:,2))]
allmass_ages=[mean(mass0_1(:,2));mean(mass1_2(:,2));mass2_5(:,2);mass6_12(:,2);m
ass13_19(:,2);mass20_up(:,2);mean(mass2up(:,2))];
disp('Total number of subjects in all mass/age groups')
size(allmass_ages)
disp('Mean weight of all mass/age groups')
mean(allmass_ages)
disp('Percent Users')
disp('age 0-1')
(size(day1_fc_age0_1)/size(ind_age0_1))*100
disp('age 1-2')
(size(day1_fc_age1_2)/size(ind_age1_2))*100
disp('age 2-5')
(size(day1_fc_age2_5)/size(ind_age2_5))*100
disp('age 6-12')
(size(day1_fc_age6_12)/size(ind_age6_12))*100
disp('age 13-19')
(size(day1_fc_age13_19)/size(ind_age13_19))*100

```

```

disp('age 20+')
(size(day1_fc_age20_up)/size(ind_age20_up))*100
disp('age 2+')
(size(day1_fc_age2up)/size(ind_age2up))*100
% now adjust the mean EDI and 90th %-tile for weight
disp('Weight-based Mean EDIs')
disp('age 0-1')
medi0_1/mean(mass0_1(:,2))
disp('age 1-2')
medi1_2/mean(mass1_2(:,2))
disp('age 2-5')
medi2_5/mean(mass2_5(:,2))
disp('age 6-12')
medi6_12/mean(mass6_12(:,2))
disp('age 13-19')
medi13_19/mean(mass13_19(:,2))
disp('age 20+')
medi20up/mean(mass20_up(:,2))
disp('age 2+')
medi2up/mean(mass2up(:,2))
%
disp('Weight-based 90th Percentile EDIs')
disp('age 0-1')
M90_age0_1/mean(mass0_1(:,2))
disp('age 1-2')
M90_age1_2/mean(mass1_2(:,2))
disp('age 2-5')
M90_age2_5/mean(mass2_5(:,2))
disp('age 6-12')
M90_age6_12/mean(mass6_12(:,2))
disp('age 13-19')
M90_age13_19/mean(mass13_19(:,2))
disp('age 20+')
M90_age20_up/mean(mass20_up(:,2))
disp('age 2+')
M90_age2up/mean(mass2up(:,2))
% At this point it would be good to print out the food codes and serving
% sizes used to produce these data. Also the dose of proposed additive.
disp('Dose of Proposed Additive')
dose
disp('Food Codes and Serving Sizes Used to Produce These Data')
serving_sizes
zarray = mean(allmass_ages); % just to put something into the output variable and

```

avoid an error
end

searchfoodcodes

```
function [foodnums,longfooddesc] = searchfoodcodes(targ_codes)
% This function takes a 1-D list food codes (targ_codes) targeted for an EDI
% calculation in NHANES and zeros out rows (foods) in the food code list
% (food codes.xls, must be in path) that are not in the targeted food code list
[num,txt,row] = xlsread('food codes.xls'); % read in food codes
    % column 1=numbers, 2=TEXT CAPS, 3=long text
c=size(num); % setting counter maximum
c=c(1); % first number is loop size needed
zarray = txt(:,3); % fill output array with long food descriptors
for i=1:c % loop over all food codes
    if any(num(i,1) == targ_codes); % compare subject SEQNs to target codes
        a=1; % if match, do nothing, leave line there
    else
        zarray(i,1) = cellstr('X0'); % else zero it
    end
end
unkeep=find(strcmp(zarray(:,1),'X0')); % delete these foods in the food code list
tu=transpose(unkeep); % vectors must be the same type, row vector
k=zeros(1,c); % create indices for set of all food codes
k=k+(1:c); % indices are now a row vector
keep=setdiff(k,tu); % keep the set of foods we are not deleting
zarray = zarray(keep,:); % by deleting lines with x zeros
longfooddesc = zarray; % return the long food descriptors
foodnums = num(keep); % return the corresponding food codes
```

searchfoods

```
function zarray = searchfoods(age_diet,targ_codes)
% This function takes the food groups (targ_codes) targeted for an EDI
% calculation in NHANES and zeros out rows (foods) in a dietary age group
% (age_diet) that are not in the targeted food code list
zarray = age_diet;    % initialize output array
c=size(age_diet(:,2)); % setting counter maximum
c=c(1);              % first number is loop size needed
for i=1:c
    if any(age_diet(i,2) == targ_codes); % compare subject SEQNs
        a=1;          % do nothing
    else
        zarray(i,:) = 0;
    end
end
keep=find(zarray(:,2) ~= 0); % keep foods in the food code list
zarray = zarray(keep,:);
```

searchnzero

```
function zarray = searchnzero(age_group, dietary_data)
% This function takes the SEQN numbers for an age group for an EDI
% calculation in NHANES and zeros out rows not in that age group
zarray = dietary_data; % initialize output array
c=size(dietary_data(:,1)); % setting counter maximum
c=c(1); % first number is loop size needed
for i=1:c
    if any(dietary_data(i,1) == age_group); % compare subject SEQNs
        a=1; % do nothing
    else
        zarray(i,:) = 0;
    end
end
keep=find(zarray(:,1) ~= 0);
zarray = zarray(keep,:);
```

searchsizes

```
function zarray = searchsizes(age_diet_fc,serving_sizes)
% This function takes the food groups (targ_codes) targeted for an EDI
% calculation in NHANES and zeros out rows (foods) in a dietary age group
% (age_diet) that are not in the targeted food code list
zarray = age_diet_fc;    % initialize output array
c=size(age_diet_fc(:,2)); % setting counter maximum
c=c(1);                % first number is loop size needed
for i=1:c
    if any(age_diet_fc(i,2) == serving_sizes(:,1)); % compare subject SEQNs
        [l,loc_fc]=ismember(age_diet_fc(i,2),serving_sizes(:,1)); % find serving size
        zarray(i,4)=zarray(i,3)/serving_sizes(loc_fc,2); % scale g to servings
    end
end
end
```

ADI Programs:

Generate_ADI_translation_figure

```
%This program generates two
%elliptical clusters of data for demonstrating translation of clusters
%for ellagic acid ADI paper
mu=[0,0];
sigma=[2,1];
SavedTNSPEC=mvnrnd(mu,sigma,25);

figure
scatter(SavedTNSPEC(:,1),SavedTNSPEC(:,2))
hold on
mu=[3,3];
sigma=[2,1];
SavedTNSPEC=mvnrnd(mu,sigma,25);
scatter(SavedTNSPEC(:,1),SavedTNSPEC(:,2))

%Graph Center
scatter(0,0,50,'b','filled')

%Graph Center
scatter(3,3,50,'r','filled')

legend('No Effects Set','Effects Set',...
'No Effects Set Center','Effects Set Center')
```


International Units from King-Bodansky Units

```
%Conversion of King Armstrong Units into International Units
%This program uses the math laid out by
%“International Enzyme Units and Isoenzyme Nomenclature.”
%Journal of Clinical Pathology 16.4 (1963): 391–393. Print.
%and the two-point definition of a line to predict the equivalent value of
%International Units of enzyme activity from King-Armstrong Units.

%This program requires the input of an integer KingArmstrong.

%The text defines that the commonly accepted range of normal values of
%3 to 13 K.A. units/100 ml. corresponds to about 20 to 90
%micromoles/min./litre.
%The equation of the line that runs through those points is  $y=7x-1$ 

micromoles_per_min_per_liter=7*(KingArmstrong)-1;

%Bodansky Unit is defined by text as 0.535 micromol/min
Bodansky_Units=micromoles_per_min_per_liter/0.535;

%The text defines that the commonly accepted range of normal values of
%1 to 5 Bodansky units/100 ml. blood plasma corresponds to
%about 5.4 to 27 micromoles/min./litre.
%The equation of the line that runs through those points is  $y=5.4x$ 
International_Units=5.4*(Bodansky_Units)
```

Round the Compass Rose Experiment Programs:

Round the Compass Rose (Simplified)

%Round the Compass Rose

%Description:

%This program tests increasing bootstrap replicates increasing, number of
%dimensions (wavelengths), increasing size of training set, and
%differential proportion of points chosen to be inside of the
%hypercylinder (radfrac). Radfrac is tested with a given minimum
%radfrac, a step value, and a maximum value.
%Tests the performance of replica.m and qb.m on elliptical clusters of
%data.

%Parameters:

%z: maximum number of bootstrap replicates desired/50.
%dimension: the maximum number of dimensions (variables) wished to be
% tested
%Max_training: the maximum training set size desired/50, represented by "u"
%Radfrac_min: the minimum radfrac wished to be tested. Radfrac_min must be
% <=1.
%Radfrac_step: the step wished for testing radfrac
%Radfrac_max: the maximum radfrac wished to be tested. Radfrac_max must be
% <=1.
%Radfrac_min = 1, Radfrac_step=0 is an invalid input.

%CODE BEGINS HERE

%Start by initializing the random number generator.

rng('default');

rng(1);

sensitiv=0;

%Initialize the iteration counter for radfrac loops. Since radfrac is a
%non-integer value that can vary, it can't be used as an index for a cell
%array. Thus, radfrac_iteration_count is used for the index instead.

radfrac_iteration_count=0;

```

for radfrac= Radfrac_min : Radfrac_step : Radfrac_max
    %a123=1; %This line only exists for the purpose of utilizing
    %breakpoints.
    if Radfrac_min==1
        disp('Radfrac_min must not be 1.')
        break %Should resolve Radfrac_min!=1 problem and just
        %exit the loop.
    elseif Radfrac_step==0
        disp('Radfrac_step must not be 0')
        break
    elseif Radfrac_min==Radfrac_max
        disp('Radfrac_min may not equal Radfrac_max.')
        break
    else radfrac_iteration_count=radfrac_iteration_count+1;
    end

    for u=1:Max_training

        %Begin for-loop. This will allow for the increase of dimensions from 1
        %to dimensions with a step of one. Further uses of i are to concatenate
        %columns of zeroes and ensure the correct number of coordinates for mu (0
        %in all dimensions)and sigma (1 in all dimensions other than one).

        for v=0:dimension; %%this was originally from v=1:dimension so that the cell
            %indices could be x(v), but it has been modified to
            %x(v+1)this has been modified from v=0:dimension so that the cell indices work

            %Generate an ellipse
            mu=[0,0];
            mu = [mu zeros(size(mu,1),(v))];
            sigma=[2,1];
            sigma=[sigma ones(size(sigma,1),(v))];
            TNSPEC=mvnrnd(mu,sigma,50*u);compass_points=[0,1;(-2/(sqrt(5))),(2/(sqrt(5)));
            -2,0;(-2/(sqrt(5))),(2/(sqrt(5)));
            0,-1;(2/(sqrt(5))),(2/(sqrt(5))); 2,0; (2/(sqrt(5))),(2/(sqrt(5)))];

            %Add a column of zeros to compass_points indicating coordinate in
            %each of the the new dimensions
            compass_points=[compass_points zeros(size(compass_points,1),(v))];

            %Define expected standard deviations. CHANGE THIS LINE IF POINTS

```

```

% CHANGE.
expected_standard_deviation = [1;1.2649;2;1.2649;1;1.2649;2;1.2649];

%Test point at each point of the compass rose

for j=1:8
    newspec = [compass_points(j,:)];
    %This is the innermost iterative loop of the program.

    %Initialize storage matrices of sds and sds skew
    %(6+v) syntax increases the size of SDS_matrix to fit the
    %coordinate points of each dimension
    SDS_matrix=zeros(z,(6+(v)));
    SDS_matrix_2=zeros(z,(6+v));
    SDS_matrix_3=zeros(z,(6+v));
    SDS_matrix_4=zeros(z,(6+v));
    SDS_matrix_5=zeros(z,(6+v));
    SDS_matrix_6=zeros(z,(6+v));

%How number of bootstrap replicates affects SDS and SDS Skew at those
%points on the compass rose
%SDS_matrix contains [compass_points(j,:),B,sds,bias,sds skew] in that order
for i=1:z

    B=50*i;
    sds = nan; bias = nan; sds skew = nan;
    try
    [BTRAIN,CNTER]=replica(TNSPEC,B);
    [sds,sds skew,qr] = qb(TNSPEC,BTRAIN,newspec,CNTER,radfrac,sensitiv);
    bias=sds-expected_standard_deviation(j);
    catch ME
    end
    SDS_matrix((i),:)= [compass_points(j,:),B,sds,bias,sds skew];
end

%Repeat the process to generate SDS_matrix_2
for i=1:z
    B=50*i;
    sds = nan; bias = nan; sds skew = nan;
    try
    [BTRAIN,CNTER]=replica(TNSPEC,B);

```

```

[sds,sdskew,qrr] = qb(TNSPEC,BTRAIN,newspec,CENTER,radfrac,sensitiv);
bias=sds-expected_standard_deviation(j);
    catch ME

    end
    SDS_matrix_2((i),:)=[compass_points(j,:),B,sds,bias,sdskew];
end

%Repeat the process to generate SDS_matrix_3
for i=1:z
    B=50*i;
sds = nan; bias = nan; sdskew = nan;
    try
        [BTRAIN,CENTER]=replica(TNSPEC,B);
[sds,sdskew,qrr] = qb(TNSPEC,BTRAIN,newspec,CENTER,radfrac,sensitiv);
bias=sds-expected_standard_deviation(j);
        catch ME
            end
            SDS_matrix_3((i),:)=[compass_points(j,:),B,sds,bias,sdskew];
end

%Repeat the process to generate SDS_matrix_4
for i=1:z
    B=50*i;
sds = nan; bias = nan; sdskew = nan;
    try
        [BTRAIN,CENTER]=replica(TNSPEC,B);
[sds,sdskew,qrr] = qb(TNSPEC,BTRAIN,newspec,CENTER,radfrac,sensitiv);
bias=sds-expected_standard_deviation(j);
        catch ME
            end
            SDS_matrix_4((i),:)=[compass_points(j,:),B,sds,bias,sdskew];
end

%Repeat the process to generate SDS_matrix_5
for i=1:z
    B=50*i;
sds = nan; bias = nan; sdskew = nan;
    try
        [BTRAIN,CENTER]=replica(TNSPEC,B);
[sds,sdskew,qrr] = qb(TNSPEC,BTRAIN,newspec,CENTER,radfrac,sensitiv);
bias=sds-expected_standard_deviation(j);
        catch ME

```

```

    end
    SDS_matrix_5((i,:)=compass_points(j,:),B,sds,bias,sdskew];

end

%Repeat the process to generate SDS_matrix_6
for i=1:z
    B=50*i;
    sds = nan; bias = nan; sdskew = nan;
    try
        [BTRAIN,CNTER]=replica(TNSPEC,B);
        [sds,sdskew,qrr] = qb(TNSPEC,BTRAIN,newspec,CNTER,radfrac,sensitiv);
        bias=sds-expected_standard_deviation(j);
    catch ME
    end
    SDS_matrix_6((i,:)=compass_points(j,:),B,sds,bias,sdskew];
end

%***All storage has to happen here***
%This is the innermost iterative loop of the program.

%Generate a matrix of the average of the 6 runs.
%AVERAGE contains[compass_points(j,:),B,sds,bias,sdskew] in that order
%Note that AVERAGE increases in size as dimension increases.
%When dimension increases by one, one additional column is inserted in
%the three position. This can be expressed by the following:

    %sdskew is located in a column position equal to 6+v, or in
    %AVERAGE(:,6+v)
    %bias is located in (:,5+v)
    %sds is located in (:,4+v)
    %B is located in (:,3+v,)
    %compass points are located in(:,1:v+2)
AVERAGE=zeros(size(SDS_matrix));
AVERAGE=((SDS_matrix + SDS_matrix_2 + SDS_matrix_3 +SDS_matrix_4 +
SDS_matrix_5 + SDS_matrix_6))/6;

%Populate SDS_X_STDEV with the SD values of each run
%sds is located in (:,4:v)
SDS_1_STDEV=zeros(z,1);
SDS_1_STDEV(:,1)=SDS_matrix(:,(4+v));

```

```

SDS_2_STDEV=zeros(z,1);
SDS_2_STDEV(:,1)=SDS_matrix_2(:,(4+v));

SDS_3_STDEV=zeros(z,1);
SDS_3_STDEV(:,1)=SDS_matrix_3(:,(4+v));

SDS_4_STDEV=zeros(z,1);
SDS_4_STDEV(:,1)=SDS_matrix_4(:,(4+v));

SDS_5_STDEV=zeros(z,1);
SDS_5_STDEV(:,1)=SDS_matrix_5(:,(4+v));

SDS_6_STDEV=zeros(z,1);
SDS_6_STDEV(:,1)=SDS_matrix_6(:,(4+v));

%Populate SDS_X_STDEV_calc with SDS_X_STDEV - average for STDEV
SDS_1_STDEV_calc=zeros(z,1);
SDS_1_STDEV_calc(:,1)=((SDS_1_STDEV(:,1)-AVERAGE(:,(4+v))).^2);

SDS_2_STDEV_calc=zeros(z,1);
SDS_2_STDEV_calc(:,1)=((SDS_2_STDEV(:,1)-AVERAGE(:,(4+v))).^2);

SDS_3_STDEV_calc=zeros(z,1);
SDS_3_STDEV_calc(:,1)=((SDS_3_STDEV(:,1)-AVERAGE(:,(4+v))).^2);

SDS_4_STDEV_calc=zeros(z,1);
SDS_4_STDEV_calc(:,1)=((SDS_4_STDEV(:,1)-AVERAGE(:,(4+v))).^2);

SDS_5_STDEV_calc=zeros(z,1);
SDS_5_STDEV_calc(:,1)=((SDS_5_STDEV(:,1)-AVERAGE(:,(4+v))).^2);

SDS_6_STDEV_calc=zeros(z,1);
SDS_6_STDEV_calc(:,1)=((SDS_6_STDEV(:,1)-AVERAGE(:,(4+v))).^2);

% Sum the above and calculate STANDARD_DEVIATION=
FINAL_STDEV=zeros(z,1);
FINAL_STDEV=((((SDS_1_STDEV_calc + SDS_2_STDEV_calc + SDS_3_STDEV_calc
+ SDS_4_STDEV_calc + SDS_5_STDEV_calc + SDS_6_STDEV_calc)/6).^5);

%Store bias vs. B for each of the eight points of the ellipse,each TNSPEC,
%and each Radfrac

```

```

    %bias is located in AVERAGE(:,5+v)
    %B is located in AVERAGE(:,3+v,)
Bias_plot=zeros(z,7);
Bias_plot(:,1)=AVERAGE(:,(3+v));
Bias_plot(:,2)=AVERAGE(:,(5+v));
Bias_plot(:,3)=(v+2);
Bias_plot(:,4)=AVERAGE(:,1);
Bias_plot(:,5)=AVERAGE(:,2);
Bias_plot(:,6)=radfrac;
Bias_plot(:,7)=u*50;
%"NumVL{v+1}=Bias_plot" so that l v=0:dimension can be used at the
%beginning of the v loop
NumVL{radfrac_iteration_count,u,(v+1),j}=Bias_plot;
%Indexing is {radfrac number, training, dimension - 1, compass point
%number}

%B is located in Bias_plot(:,1)
%bias is located in Bias_plot(:,2)
%dimension is located in Bias_plot(:,3)
%First 2 coordinates of compass point located in Bias_plot(:,4) and (:,5)
%radfrac located in Bias_plot(:,6)
%Training set size located in Bias_plot(:,7)

%Store Relative Standard Deviation (average standard deviation/
%average)
%RSD holds [compass point coordinate, compass point coordinate,B,RSD of
%returned standard deviation, total number of dimensions (v+2), radfrac,
%Training Set Size]
RSD=zeros(z,7);
RSD(:,1)=AVERAGE(:,1);
RSD(:,2)=AVERAGE(:,2);
RSD(:,3)=AVERAGE(:,(v+3));
RSD(:,4)=(FINAL_STDEV(:,1)./AVERAGE(:,(v+4)));
RSD(:,5)=(v+2);
RSD(:,6)=radfrac;
RSD(:,7)=u*50;

%Hold all RSD plots for all runs
%Indexing is {Radfrac Iteration Count (must
%separately count Radfrac value),Training Set Size,dimension,Compass Point,}
RSD_NumVL{radfrac_iteration_count,u,(v+1),j}=RSD;

```


end
end
end
end

Regression Codes

RSD_Regression_2

%Regresses RSD as function of Bootstrapping Only
%Requires workspace variables NumVL and RSD_NumVL, Max_training,

%Indexing in NumVL Cell Matrix is as Follows:
%Indexing is {Radfrac Iteration Count (must separately count...
%Radfrac value), Training Set Size, dimension, Compass Point,}
%RSD_NumVL{radfrac_iteration_count,u,(v+1),j}=RSD;

%Within matrix:
%Store Relative Standard Deviation (average standard deviation/
%average)
%RSD holds [1.compass point coordinate, 2.compass point coordinate,...
%3.B, 4.RSD of returned standard deviation, ...
%5.total number of dimensions (v+2), 6.radfrac,
%7.Training Set Size]

%Make sure these exist.
NumVL;
RSD_NumVL;

```
out=[cat(1,RSD_NumVL{:})];  
x=size(out);  
w=x(:,1);  
y=zeros(x(:,1),1);
```

%Designate compass points - work in progress
newmatrix=[out,y];

%nonprincipal axis points
npa=newmatrix(:,1)==0;

%principal axis points
pa=newmatrix(:,1)==2;
opa=newmatrix(:,1)==-2;

%since diagonals have a calculation that determines the value
%it is hard to tell how many digits of precision
%therefore, diagonals are not equal to 0, 2, or -2.

```

%First find indices ~=0
dp=newmatrix(:,1)~=0;
dp2=newmatrix(dp,:);
%Now find indices in the ~=0 subset also ~= 2
fdp=dp2(:,1)~=2;
dp3=dp2(fdp,:);
%Last, find indices in that subset also ~= to -2
ffdp=dp3(:,1)~=-2;

nonprincipalaxis=[newmatrix(npa,:)];
principalaxis=[newmatrix(pa,:);newmatrix(opa,:)];
diagonalpoints=[dp3(ffdp,:)];

for i=1:3
    %try to regress output for each degree.
    degree=i

    %Regress Non-Principal
    disp('Non-Principal Axis Regression')
    npa_output=[nonprincipalaxis(:,4)];
    npa_input=[nonprincipalaxis(:,3)];
    %,nonprincipalaxis(:,5),...
    % nonprincipalaxis(:,6),nonprincipalaxis(:,7)];
    reg = MultiPolyRegress(npa_input,npa_output,degree)

    %Regress Principal
    disp('Principal Axis Regression')
    pa_output=[principalaxis(:,4)];
    pa_input=[principalaxis(:,3)];
    %,principalaxis(:,5),...
    % principalaxis(:,6),principalaxis(:,7)];
    reg = MultiPolyRegress(pa_input,pa_output,degree)

    %Regress Diagonal
    disp('Diagonal Regression')
    d_output=[diagonalpoints(:,4)];
    d_input=[diagonalpoints(:,3)];
    %,diagonalpoints(:,5),...
    % diagonalpoints(:,6),diagonalpoints(:,7)];

```

```
    reg = MultiPolyRegress(d_input,d_output,degree)
end
```

Plotting Codes:

RTCR_Defined_TNSPEC_Size

```
%Generate and save a training spectrum of a given size, then run and
%increase with different numbers of bootstrap
%replicates to determine whether TNSPEC is the cause of erratic RSD in
%samples
```

```
%TNSPEC was saved as SavedTNSPEC
%U HAS CHANGED. CHANGE INDEXING IN PLOT CODE
%RNG is now 20
```

```
%Parameters:
```

```
%z: maximum number of bootstrap replicates desired/50.
```

```
%dimension: the maximum number of dimensions (variables) wished to be
%      tested
```

```
%Max_training: the maximum training set size desired/50, represented by "u"
```

```
%Radfrac_min: the minimum radfrac wished to be tested. Radfrac_min must be
%      <=1.
```

```
%Radfrac_step: the step wished for testing radfrac
```

```
%Radfrac_max: the maximum radfrac wished to be tested. Radfrac_max must be
%      <=1.
```

```
%Radfrac_min = 1, Radfrac_step=0 is an invalid input.
```

```
%CODE BEGINS HERE
```

```
%Start by initializing the random number generator.
```

```
rng('default');
```

```
rng(20);
```

```
sensitiv=0;
```

```
%Initialize the iteration counter for radfrac loops. Since radfrac is a
%non-integer value that can vary, it can't be used as an index for a cell
%array. Thus, radfrac_iteration_count is used for the index instead.
```

```

radfrac_iteration_count=4;
radfrac=0.8;
u=200;
z=10;

%Begin for-loop. This will allow for the increase of dimensions from 1
%to dimensions with a step of one. Further uses of i are to concatenate
%columns of zeroes and ensure the correct number of coordinates for mu (0
%in all dimensions)and sigma (1 in all dimensions other than one).

v=0;

%Generate an ellipse
mu=[0,0];
mu = [mu zeros(size(mu,1),(v))];
sigma=[2,1];
sigma=[sigma ones(size(sigma,1),(v))];
TNSPEC=mvnrnd(mu,sigma,50*u);
SavedTNSPEC=TNSPEC
%Compass Points & Expected Standard Deviation

compass_points=[0,1;(-2/(sqrt(5))),(2/(sqrt(5))); -2,0;(-2/(sqrt(5))),(2/(sqrt(5)));...
0,-1;(2/(sqrt(5))),(2/(sqrt(5))); 2,0; (2/(sqrt(5))),(2/(sqrt(5)))];

expected_standard_deviation=[1;1.2649;2;1.2649;1;1.2649;2;1.2649];
%Test point at each point of the compass rose

for j=1:8
newspec = [compass_points(j,:)];
%This is the innermost iterative loop of the program.

%Initialize storage matrices of sds and sdsskew
%(6+v) syntax increases the size of SDS_matrix to fit the
%coordinate points of each dimension
SDS_matrix=zeros(z,(6+(v)));
SDS_matrix_2=zeros(z,(6+v));
SDS_matrix_3=zeros(z,(6+v));
SDS_matrix_4=zeros(z,(6+v));
SDS_matrix_5=zeros(z,(6+v));
SDS_matrix_6=zeros(z,(6+v));

```

```

%How number of bootstrap replicates affects SDS and SDS Skew at those
%points on the compass rose
%SDS_matrix contains [compass_points(j,:),B,sds,bias,sdskew] in that order
for i=1:z

    B=50*i;
    sds = nan; bias = nan; sdskew = nan;
    try
        [BTRAIN,CNTER]=replica(TNSPEC,B);
        [sds,sdskew,qrr] = qb(TNSPEC,BTRAIN,newspec,CNTER,radfrac,sensitiv);
        bias=sds-expected_standard_deviation(j);
    catch ME
    end
    SDS_matrix((i,:))=[compass_points(j,:),B,sds,bias,sdskew];
end

%Repeat the process to generate SDS_matrix_2
for i=1:z
    B=50*i;
    sds = nan; bias = nan; sdskew = nan;
    try
        [BTRAIN,CNTER]=replica(TNSPEC,B);
        [sds,sdskew,qrr] = qb(TNSPEC,BTRAIN,newspec,CNTER,radfrac,sensitiv);
        bias=sds-expected_standard_deviation(j);
    catch ME

    end
    SDS_matrix_2((i,:))=[compass_points(j,:),B,sds,bias,sdskew];
end

%Repeat the process to generate SDS_matrix_3
for i=1:z
    B=50*i;
    sds = nan; bias = nan; sdskew = nan;
    try
        [BTRAIN,CNTER]=replica(TNSPEC,B);
        [sds,sdskew,qrr] = qb(TNSPEC,BTRAIN,newspec,CNTER,radfrac,sensitiv);
        bias=sds-expected_standard_deviation(j);
    catch ME
    end
    SDS_matrix_3((i,:))=[compass_points(j,:),B,sds,bias,sdskew];
end

```

```

%Repeat the process to generate SDS_matrix_4
for i=1:z
    B=50*i;
    sds = nan; bias = nan; sdskew = nan;
    try
        [BTRAIN,CNTER]=replica(TNSPEC,B);
    [sds,sdskew,qrr] = qb(TNSPEC,BTRAIN,newspec,CNTER,radfrac,sensitiv);
    bias=sds-expected_standard_deviation(j);
    catch ME
    end
    SDS_matrix_4((i),:)=[compass_points(j,:),B,sds,bias,sdskew];
end

```

```

%Repeat the process to generate SDS_matrix_5
for i=1:z
    B=50*i;
    sds = nan; bias = nan; sdskew = nan;
    try
        [BTRAIN,CNTER]=replica(TNSPEC,B);
    [sds,sdskew,qrr] = qb(TNSPEC,BTRAIN,newspec,CNTER,radfrac,sensitiv);
    bias=sds-expected_standard_deviation(j);
    catch ME
    end
    SDS_matrix_5((i),:)=[compass_points(j,:),B,sds,bias,sdskew];

```

end

```

%Repeat the process to generate SDS_matrix_6
for i=1:z
    B=50*i;
    sds = nan; bias = nan; sdskew = nan;
    try
        [BTRAIN,CNTER]=replica(TNSPEC,B);
    [sds,sdskew,qrr] = qb(TNSPEC,BTRAIN,newspec,CNTER,radfrac,sensitiv);
    bias=sds-expected_standard_deviation(j);
    catch ME
    end
    SDS_matrix_6((i),:)=[compass_points(j,:),B,sds,bias,sdskew];
end

```

%***All storage has to happen here***

%This is the innermost iterative loop of the program.


```

%Generate a matrix of the average of the 6 runs.
%AVERAGE contains[compass_points(j,:),B,sds,bias,sdskew] in that order
%Note that AVERAGE increases in size as dimension increases.
%When dimension increases by one, one additional column is inserted in
%the three position. This can be expressed by the following:

    %sdskew is located in a column position equal to 6+v, or in
    %AVERAGE(:,6+v)
    %bias is located in (:,5+v)
    %sds is located in (:,4+v)
    %B is located in (:,3+v,)
    %compass points are located in(:,1:v+2)
AVERAGE=zeros(size(SDS_matrix));
AVERAGE=((SDS_matrix + SDS_matrix_2 + SDS_matrix_3 +SDS_matrix_4 +
SDS_matrix_5 + SDS_matrix_6))/6;

%Populate SDS_X_STDEV with the SD values of each run
%sds is located in (:,4:v)
SDS_1_STDEV=zeros(z,1);
SDS_1_STDEV(:,1)=SDS_matrix(:,(4+v));

SDS_2_STDEV=zeros(z,1);
SDS_2_STDEV(:,1)=SDS_matrix_2(:,(4+v));

SDS_3_STDEV=zeros(z,1);
SDS_3_STDEV(:,1)=SDS_matrix_3(:,(4+v));

SDS_4_STDEV=zeros(z,1);
SDS_4_STDEV(:,1)=SDS_matrix_4(:,(4+v));

SDS_5_STDEV=zeros(z,1);
SDS_5_STDEV(:,1)=SDS_matrix_5(:,(4+v));

SDS_6_STDEV=zeros(z,1);
SDS_6_STDEV(:,1)=SDS_matrix_6(:,(4+v));

%Populate SDS_X_STDEV_calc with SDS_X_STDEV - average for STDEV
SDS_1_STDEV_calc=zeros(z,1);
SDS_1_STDEV_calc(:,1)=((SDS_1_STDEV(:,1)-AVERAGE(:,(4+v))).^2);

SDS_2_STDEV_calc=zeros(z,1);

```

```

SDS_2_STDEV_calc(:,1)=((SDS_2_STDEV(:,1)-AVERAGE(:,(4+v))).^2);

SDS_3_STDEV_calc=zeros(z,1);
SDS_3_STDEV_calc(:,1)=((SDS_3_STDEV(:,1)-AVERAGE(:,(4+v))).^2);

SDS_4_STDEV_calc=zeros(z,1);
SDS_4_STDEV_calc(:,1)=((SDS_4_STDEV(:,1)-AVERAGE(:,(4+v))).^2);

SDS_5_STDEV_calc=zeros(z,1);
SDS_5_STDEV_calc(:,1)=((SDS_5_STDEV(:,1)-AVERAGE(:,(4+v))).^2);

SDS_6_STDEV_calc=zeros(z,1);
SDS_6_STDEV_calc(:,1)=((SDS_6_STDEV(:,1)-AVERAGE(:,(4+v))).^2);

% Sum the above and calculate STANDARD_DEVIATION=
FINAL_STDEV=zeros(z,1);
FINAL_STDEV=((SDS_1_STDEV_calc + SDS_2_STDEV_calc + SDS_3_STDEV_calc
+ SDS_4_STDEV_calc + SDS_5_STDEV_calc + SDS_6_STDEV_calc)/6).^5);

%Store bias vs. B for each of the eight points of the ellipse,each TNSPEC,
%and each Radfrac
    %bias is located in AVERAGE(:,5+v)
    %B is located in AVERAGE(:,3+v,)
Bias_plot=zeros(z,7);
Bias_plot(:,1)=AVERAGE(:,(3+v));
Bias_plot(:,2)=AVERAGE(:,(5+v));
Bias_plot(:,3)=(v+2);
Bias_plot(:,4)=AVERAGE(:,1);
Bias_plot(:,5)=AVERAGE(:,2);
Bias_plot(:,6)=radfrac;
Bias_plot(:,7)=u;
%"NumVL{v+1}=Bias_plot" so that l v=0:dimension can be used at the
%beginning of the v loop
NumVL{radfrac_iteration_count,u,(v+1),j}=Bias_plot;
%Indexing is {radfrac number, training, dimension - 1, compass point
%number}

%B is located in Bias_plot(:,1)
%bias is located in Bias_plot(:,2)
%dimension is located in Bias_plot(:,3)
%First 2 coordinates of compass point located in Bias_plot(:,4) and (:,5)

```

```

%radfrac located in Bias_plot(:,6)
%Training set size located in Bias_plot(:,7)

%Store Relative Standard Deviation (average standard deviation/
%average)
%RSD holds [compass point coordinate, compass point coordinate,B,RSD of
%returned standard deviation, total number of dimensions (v+2), radfrac,
%Training Set Size]
RSD=zeros(z,7);
RSD(:,1)=AVERAGE(:,1);
RSD(:,2)=AVERAGE(:,2);
RSD(:,3)=AVERAGE(:,(v+3));
RSD(:,4)=(FINAL_STDEV(:,1)./AVERAGE(:,(v+4)));
RSD(:,5)=(v+2);
RSD(:,6)=radfrac;
RSD(:,7)=u;

%Hold all RSD plots for all runs
%Indexing is {Radfrac Iteration Count (must
%separately count Radfrac value),Training Set Size,dimension,Compass Point,}
RSD_NumVL{radfrac_iteration_count,u,(v+1),j}=RSD;
end

%Plot Code

```

Plotting Relative Standard Deviation

```
%Plotting Relative Standard Deviation
%as Bootstrapping and Radfrac Increase

%This code plots bootstrapping vs. RSD as radfrac increases
%for run_TNSPEC code ONLY

%Changes: u has changed in the run_TNPSEC code from what it
%has in Round_the_Compass_Rose_simplified.m

%Indexing in NumVL Cell Matrix is as Follows:
%Indexing is {Radfrac Iteration Count (must separately count...
%Radfrac value),Training Set Size,dimension,Compass Point,}
%RSD_NumVL{radfrac_iteration_count,u,(v+1),j}=RSD;

%Within matrix:
%Store Relative Standard Deviation (average standard deviation/
%average)
%RSD holds [1.compass point coordinate, 2.compass point coordinate,...
%3.B, 4.RSD of returned standard deviation, ...
%5.total number of dimensions (v+2), 6.radfrac,
%7.Training Set Size]

figure('Name','4,1,1 npa')
r4_t50_d2_c01=[RSD_NumVL{4,200,1,1}]
plot(r4_t50_d2_c01(:,3),r4_t50_d2_c01(:,4));
hold on
r4_t50_d2_cneg01=[RSD_NumVL{4,200,1,5}];
plot(r4_t50_d2_cneg01(:,3),r4_t50_d2_cneg01(:,4));
title({'Relative Standard Deviation as Bootstrap Replicates Increase';...
'Radfrac=0.8 Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Relative Standard Deviation')
legend('CP=(0,1)','CP=(0,-1)')
hold off
```

Run TNSPEC Version 2

%Re-run premade Training spectrum with different numbers of bootstrap
%replicates to determine whether TNSPEC is the cause of erratic RSD in
%samples

%TNSPEC was saved as
%U HAS CHANGED. CHANGE INDEXING IN PLOT CODE

%Parameters:

%z: maximum number of bootstrap replicates desired/50.

%dimension: the maximum number of dimensions (variables) wished to be
% tested

%Max_training: the maximum training set size desired/50, represented by "u"

%Radfrac_min: the minimum radfrac wished to be tested. Radfrac_min must be
% <=1.

%Radfrac_step: the step wished for testing radfrac

%Radfrac_max: the maximum radfrac wished to be tested. Radfrac_max must be
% <=1.

%Radfrac_min = 1, Radfrac_step=0 is an invalid input.

%CODE BEGINS HERE

%Start by initializing the random number generator.

%Input this first:

```
%rng('default');
```

```
%rng(10);
```

```
sensitiv=0;
```

%Initialize the iteration counter for radfrac loops. Since radfrac is a
%non-integer value that can vary, it can't be used as an index for a cell
%array. Thus, radfrac_iteration_count is used for the index instead.

```
radfrac_iteration_count=4;
```

```
radfrac=0.8;
```

```
[u,other]=size(SavedTNSPEC);
```

```
z=10;
```

```

%Begin for-loop. This will allow for the increase of dimensions from 1
%to dimensions with a step of one. Further uses of i are to concatenate
%columns of zeroes and ensure the correct number of coordinates for mu (0
%in all dimensions)and sigma (1 in all dimensions other than one).

v=0;

%Generate an ellipse
mu=[0,0];
mu = [mu zeros(size(mu,1),(v))];
sigma=[2,1];
sigma=[sigma ones(size(sigma,1),(v))];
TNSPEC=SavedTNSPEC
%Compass Points & Expected Standard Deviation

compass_points=[0,1;(-2/(sqrt(5))),(2/(sqrt(5))); -2,0;(-2/(sqrt(5))),(2/(sqrt(5)));...
0,-1;(2/(sqrt(5))),(2/(sqrt(5))); 2,0; (2/(sqrt(5))),(2/(sqrt(5)))];

expected_standard_deviation=[1;1.2649;2;1.2649;1;1.2649;2;1.2649];
%Test point at each point of the compass rose

for j=1:8
newspec = [compass_points(j,:)];
%This is the innermost iterative loop of the program.

%Initialize storage matrices of sds and sds skew
%(6+v) syntax increases the size of SDS_matrix to fit the
%coordinate points of each dimension
SDS_matrix=zeros(z,(6+(v)));
SDS_matrix_2=zeros(z,(6+v));
SDS_matrix_3=zeros(z,(6+v));
SDS_matrix_4=zeros(z,(6+v));
SDS_matrix_5=zeros(z,(6+v));
SDS_matrix_6=zeros(z,(6+v));

%How number of bootstrap replicates affects SDS and SDS Skew at those
%points on the compass rose
%SDS_matrix contains [compass_points(j,:),B,sds,bias,sds skew] in that order
for i=1:z

B=50*i;
sds = nan; bias = nan; sds skew = nan;

```

```

    try
[BTRAIN,CNTER]=replica(TNSPEC,B);
[sds,sdskew,qrr] = qb(TNSPEC,BTRAIN,newspec,CNTER,radfrac,sensitiv);
bias=sds-expected_standard_deviation(j);
    catch ME
    end
    SDS_matrix((i,:)=compass_points(j,:),B,sds,bias,sdskew];
end

```

%Repeat the process to generate SDS_matrix_2

```

for i=1:z
    B=50*i;
sds = nan; bias = nan; sdskew = nan;
    try
[BTRAIN,CNTER]=replica(TNSPEC,B);
[sds,sdskew,qrr] = qb(TNSPEC,BTRAIN,newspec,CNTER,radfrac,sensitiv);
bias=sds-expected_standard_deviation(j);
    catch ME

    end
    SDS_matrix_2((i,:)=compass_points(j,:),B,sds,bias,sdskew];
end

```

%Repeat the process to generate SDS_matrix_3

```

for i=1:z
    B=50*i;
sds = nan; bias = nan; sdskew = nan;
    try
[BTRAIN,CNTER]=replica(TNSPEC,B);
[sds,sdskew,qrr] = qb(TNSPEC,BTRAIN,newspec,CNTER,radfrac,sensitiv);
bias=sds-expected_standard_deviation(j);
    catch ME
    end
    SDS_matrix_3((i,:)=compass_points(j,:),B,sds,bias,sdskew];
end

```

%Repeat the process to generate SDS_matrix_4

```

for i=1:z
    B=50*i;
sds = nan; bias = nan; sdskew = nan;
    try
[BTRAIN,CNTER]=replica(TNSPEC,B);

```

```

[sds,sdskew,qrr] = qb(TNSPEC,BTRAIN,newspec,CNTER,radfrac,sensitiv);
bias=sds-expected_standard_deviation(j);
    catch ME
    end
    SDS_matrix_4((i,:)=compass_points(j,:),B,sds,bias,sdskew];
end

```

```

%Repeat the process to generate SDS_matrix_5
for i=1:z
    B=50*i;
sds = nan; bias = nan; sdskew = nan;
    try
        [BTRAIN,CNTER]=replica(TNSPEC,B);
[sds,sdskew,qrr] = qb(TNSPEC,BTRAIN,newspec,CNTER,radfrac,sensitiv);
bias=sds-expected_standard_deviation(j);
        catch ME
        end
        SDS_matrix_5((i,:)=compass_points(j,:),B,sds,bias,sdskew];

end

```

```

%Repeat the process to generate SDS_matrix_6
for i=1:z
    B=50*i;
sds = nan; bias = nan; sdskew = nan;
    try
        [BTRAIN,CNTER]=replica(TNSPEC,B);
[sds,sdskew,qrr] = qb(TNSPEC,BTRAIN,newspec,CNTER,radfrac,sensitiv);
bias=sds-expected_standard_deviation(j);
        catch ME
        end
        SDS_matrix_6((i,:)=compass_points(j,:),B,sds,bias,sdskew];

end

```

%***All storage has to happen here***
 %This is the innermost iterative loop of the program.

%Generate a matrix of the average of the 6 runs.
 %AVERAGE contains[compass_points(j,:),B,sds,bias,sdskew] in that order
 %Note that AVERAGE increases in size as dimension increases.
 %When dimension increases by one, one additional column is inserted in
 %the three position. This can be expressed by the following:


```

%sdskew is located in a column position equal to 6+v, or in
%AVERAGE(:,6+v)
%bias is located in (:,5+v)
%sds is located in (:,4+v)
%B is located in (:,3+v,)
%compass points are located in(:,1:v+2)
AVERAGE=zeros(size(SDS_matrix));
AVERAGE=((SDS_matrix + SDS_matrix_2 + SDS_matrix_3 +SDS_matrix_4 +
SDS_matrix_5 + SDS_matrix_6))/6;

%Populate SDS_X_STDEV with the SD values of each run
%sds is located in (:,4:v)
SDS_1_STDEV=zeros(z,1);
SDS_1_STDEV(:,1)=SDS_matrix(:,(4+v));

SDS_2_STDEV=zeros(z,1);
SDS_2_STDEV(:,1)=SDS_matrix_2(:,(4+v));

SDS_3_STDEV=zeros(z,1);
SDS_3_STDEV(:,1)=SDS_matrix_3(:,(4+v));

SDS_4_STDEV=zeros(z,1);
SDS_4_STDEV(:,1)=SDS_matrix_4(:,(4+v));

SDS_5_STDEV=zeros(z,1);
SDS_5_STDEV(:,1)=SDS_matrix_5(:,(4+v));

SDS_6_STDEV=zeros(z,1);
SDS_6_STDEV(:,1)=SDS_matrix_6(:,(4+v));

%Populate SDS_X_STDEV_calc with SDS_X_STDEV - average for STDEV
SDS_1_STDEV_calc=zeros(z,1);
SDS_1_STDEV_calc(:,1)=((SDS_1_STDEV(:,1)-AVERAGE(:,(4+v))).^2);

SDS_2_STDEV_calc=zeros(z,1);
SDS_2_STDEV_calc(:,1)=((SDS_2_STDEV(:,1)-AVERAGE(:,(4+v))).^2);

SDS_3_STDEV_calc=zeros(z,1);
SDS_3_STDEV_calc(:,1)=((SDS_3_STDEV(:,1)-AVERAGE(:,(4+v))).^2);

SDS_4_STDEV_calc=zeros(z,1);
SDS_4_STDEV_calc(:,1)=((SDS_4_STDEV(:,1)-AVERAGE(:,(4+v))).^2);

```

```

SDS_5_STDEV_calc=zeros(z,1);
SDS_5_STDEV_calc(:,1)=((SDS_5_STDEV(:,1)-AVERAGE(:,(4+v))).^2);

SDS_6_STDEV_calc=zeros(z,1);
SDS_6_STDEV_calc(:,1)=((SDS_6_STDEV(:,1)-AVERAGE(:,(4+v))).^2);

% Sum the above and calculate STANDARD_DEVIATION=
FINAL_STDEV=zeros(z,1);
FINAL_STDEV=(((SDS_1_STDEV_calc + SDS_2_STDEV_calc + SDS_3_STDEV_calc
+ SDS_4_STDEV_calc + SDS_5_STDEV_calc + SDS_6_STDEV_calc)/6).^5);

%Store bias vs. B for each of the eight points of the ellipse,each TNSPEC,
%and each Radfrac
    %bias is located in AVERAGE(:,5+v)
    %B is located in AVERAGE(:,3+v,)
Bias_plot=zeros(z,7);
Bias_plot(:,1)=AVERAGE(:,(3+v));
Bias_plot(:,2)=AVERAGE(:,(5+v));
Bias_plot(:,3)=(v+2);
Bias_plot(:,4)=AVERAGE(:,1);
Bias_plot(:,5)=AVERAGE(:,2);
Bias_plot(:,6)=radfrac;
Bias_plot(:,7)=u;
%"NumVL{v+1}=Bias_plot" so that l v=0:dimension can be used at the
%beginning of the v loop
NumVL{radfrac_iteration_count,u,(v+1),j}=Bias_plot;
%Indexing is {radfrac number, training, dimension - 1, compass point
%number}

%B is located in Bias_plot(:,1)
%bias is located in Bias_plot(:,2)
%dimension is located in Bias_plot(:,3)
%First 2 coordinates of compass point located in Bias_plot(:,4) and (:,5)
%radfrac located in Bias_plot(:,6)
%Training set size located in Bias_plot(:,7)

%Store Relative Standard Deviation (average standard deviation/
%average)
%RSD holds [compass point coordinate, compass point coordinate,B,RSD of

```

```

%returned standard deviation, total number of dimensions (v+2), radfrac,
%Training Set Size]
RSD=zeros(z,7);
RSD(:,1)=AVERAGE(:,1);
RSD(:,2)=AVERAGE(:,2);
RSD(:,3)=AVERAGE(:,(v+3));
RSD(:,4)=(FINAL_STDEV(:,1)./AVERAGE(:,(v+4)));
RSD(:,5)=(v+2);
RSD(:,6)=radfrac;
RSD(:,7)=u;

%Hold all RSD plots for all runs
%Indexing is {Radfrac Iteration Count (must
%separately count Radfrac value),Training Set Size,dimension,Compass Point,}
RSD_NumVL{radfrac_iteration_count,u,(v+1),j}=RSD;
end

```

```

%Plot Code

```

Plotting_Code_Increase_Radfrac_Bootstrap

%This code plots bootstrapping vs. bias as radfrac increases

%Indexing in NumVL Cell Matrix is as Follows:

%{radfrac number, training, dimension - 1, compass point
%number}

%Within matrix:

%B is located in Bias_plot(:,1)

%bias is located in Bias_plot(:,2)

%dimension is located in Bias_plot(:,3)

%First 2 coordinates of compass point located in Bias_plot(:,4) and (:,5)

%radfrac located in Bias_plot(:,6)

%Training set size located in Bias_plot(:,7)

%Set Axis Limits for All Graphs:

graphlimits=[50,500,-40,40];

figure('Name','1,1,1 npa')

r1_t50_d2_c01=[NumVL{1,1,1,1}]

%x 100 to make a percent error

r1_t50_d2_c01per=[r1_t50_d2_c01(:,:).*100]

%x is out of original plot, y out of x 100

plot(r1_t50_d2_c01(:,1),r1_t50_d2_c01per(:,2));

hold on

r1_t50_d2_cneg01=[NumVL{1,1,1,5}]

%x 100 to make a percent error

r1_t50_d2_cneg01per=[r1_t50_d2_cneg01(:,:).*100]

%x is out of original plot, y out of x 100

plot(r1_t50_d2_cneg01(:,1),r1_t50_d2_cneg01per(:,2));

title({'Average Percent Error as Bootstrap Replicates Increase';'Radfrac=0.5

Dimension=2 TNSPEC=50'})

xlabel('Bootstrap Replications')

ylabel('Percent Error')

legend('CP=(0,1)', 'CP=(0,-1)')

axis(graphlimits)

hold off

figure('Name','2,1,1 npa')

r2_t50_d2_c01=[NumVL{2,1,1,1}];

```

%x 100 to make a percent error
r2_t50_d2_c01per=[r2_t50_d2_c01(:,:).*100]
%x is out of original plot, y out of x 100
plot (r2_t50_d2_c01(:,1),r2_t50_d2_c01per(:,2));
    hold on
r2_t50_d2_cneg01=[NumVL{2,1,1,5}];
%x 100 to make a percent error
r2_t50_d2_cneg01per=[r2_t50_d2_cneg01(:,:).*100]
%x is out of original plot, y out of x 100
plot (r2_t50_d2_cneg01(:,1),r2_t50_d2_cneg01per(:,2));
title({'Average Percent Error as Bootstrap Replicates Increase';'Radfrac=0.6
Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Percent Error')
legend('CP=(0,1)','CP=(0,-1)')
axis(graphlimits)
hold off

```

```

figure('Name','3,1,1 npa')
r3_t50_d2_c01=[NumVL{3,1,1,1}]
%x 100 to make a percent error
r3_t50_d2_c01per=[r3_t50_d2_c01(:,:).*100]
%x is out of original plot, y out of x 100
plot (r3_t50_d2_c01(:,1),r3_t50_d2_c01per(:,2));

```

```

hold on
r3_t50_d2_cneg01=[NumVL{3,1,1,5}];
%x 100 to make a percent error
r3_t50_d2_cneg01per=[r3_t50_d2_cneg01(:,:).*100];
%x is out of original plot, y out of x 100
plot (r3_t50_d2_cneg01(:,1),r3_t50_d2_cneg01per(:,2));
title({'Average Percent Error as Bootstrap Replicates Increase';'Radfrac=0.7
Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Percent Error')
legend('CP=(0,1)','CP=(0,-1)')
axis(graphlimits)
hold off

```

```

figure('Name','4,1,1 npa')
r4_t50_d2_c01=[NumVL{4,1,1,1}]
%x 100 to make a percent error
r4_t50_d2_c01per=[r4_t50_d2_c01(:,:).*100]

```

```

%x is out of original plot, y out of x 100
plot (r4_t50_d2_c01(:,1),r4_t50_d2_c01per(:,2));
hold on
r4_t50_d2_cneg01=[NumVL{4,1,1,5}];
%x 100 to make a percent error
r4_t50_d2_cneg01per=[r4_t50_d2_cneg01(:,:).*100]
%x is out of original plot, y out of x 100
plot (r4_t50_d2_cneg01(:,1),r4_t50_d2_cneg01per(:,2));
title({'Average Percent Error as Bootstrap Replicates Increase';...
      'Radfrac=0.8 Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Percent Error')
legend('CP=(0,1)','CP=(0,-1)')
axis(graphlimits)
hold off

figure('Name','5,1,1 npa')
r5_t50_d2_c01=[NumVL{5,1,1,1}]
%x 100 to make a percent error
r5_t50_d2_c01per=[r5_t50_d2_c01(:,:).*100]
%x is out of original plot, y out of x 100
plot (r5_t50_d2_c01(:,1),r5_t50_d2_c01per(:,2));
hold on
r5_t50_d2_cneg01=[NumVL{5,1,1,5}]
%x 100 to make a percent error
r5_t50_d2_cneg01per=[r5_t50_d2_cneg01(:,:).*100];
%x is out of original plot, y out of x 100
plot (r5_t50_d2_cneg01(:,1),r5_t50_d2_cneg01per(:,2));
title({'Average Percent Error as Bootstrap Replicates Increase';...
      'Radfrac=0.9 Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Percent Error')
legend('CP=(0,1)','CP=(0,-1)')
axis(graphlimits)
hold off

%Generate Last Figure
figure('Name','6,1,1 npa')
r6_t50_d2_c01=[NumVL{6,1,1,1}]
%Out of matrix variables to assign, so is being multiplied
%x 100 to make a percent error
%x 100 to make a percent error
r6_t50_d2_c01per=[r6_t50_d2_c01(:,:).*100]

```

```

%x is out of original plot, y out of x 100
plot(r6_t50_d2_c01(:,1),r6_t50_d2_c01per(:,2))
hold on
r6_t50_d2_cneg01=[NumVL{6,1,1,5}];
%x 100 to make a percent error
r6_t50_d2_cneg01per=[r6_t50_d2_cneg01(:,:).*100];
%x is out of original plot, y out of x 100
plot(r6_t50_d2_cneg01(:,1),r6_t50_d2_cneg01per(:,2))
title({'Average Percent Error as Bootstrap Replicates Increase';...
      'Radfrac=1.0 Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Percent Error')
legend('CP=(0,1)','CP=(0,-1)')
axis(graphlimits)
hold off

```

Plotting_Code_Increase_TNSPEC_Dimension

%This code plots tnspec vs. bias as dimension increases

%Indexing in NumVL Cell Matrix is as Follows:

%{radfrac number, training, dimension - 1, compass point
%number}

%Within matrix:

%B is located in Bias_plot(:,1)

%bias is located in Bias_plot(:,2)

%dimension is located in Bias_plot(:,3)

%First 2 coordinates of compass point located in Bias_plot(:,4) and (:,5)

%radfrac located in Bias_plot(:,6)

%Training set size located in Bias_plot(:,7)

%Get all for (0,1) Compass Points

r1_t50_d2_c01=[NumVL{1,1,1,1}];

r1_t100_d2_c01=[NumVL{1,2,1,1}];

r1_t150_d2_c01=[NumVL{1,3,1,1}];

r1_t200_d2_c01=[NumVL{1,4,1,1}];

r1_t250_d2_c01=[NumVL{1,5,1,1}];

r1_t300_d2_c01=[NumVL{1,6,1,1}];

r1_t350_d2_c01=[NumVL{1,7,1,1}];

r1_t400_d2_c01=[NumVL{1,8,1,1}];

r1_t450_d2_c01=[NumVL{1,9,1,1}];

r1_t500_d2_c01=[NumVL{1,10,1,1}];

Increasing_TNSPEC=[r1_t50_d2_c01(10,:);r1_t100_d2_c01(10,:);...

r1_t150_d2_c01(10,:);r1_t200_d2_c01(10,:);r1_t250_d2_c01(10,:);...

r1_t300_d2_c01(10,:);r1_t350_d2_c01(10,:);r1_t400_d2_c01(10,:);...

r1_t450_d2_c01(10,:);r1_t500_d2_c01(10,:)]

%Get all (0,-1) Compass Points

r1_t50_d2_cneg01=[NumVL{1,1,1,5}];

r1_t100_d2_cneg01=[NumVL{1,2,1,5}];

r1_t150_d2_cneg01=[NumVL{1,3,1,5}];

r1_t200_d2_cneg01=[NumVL{1,4,1,5}];

r1_t250_d2_cneg01=[NumVL{1,5,1,5}];

r1_t300_d2_cneg01=[NumVL{1,6,1,5}];

r1_t350_d2_cneg01=[NumVL{1,7,1,5}];

r1_t400_d2_cneg01=[NumVL{1,8,1,5}];


```

r1_t450_d2_cneg01=[NumVL{1,9,1,5}];
r1_t500_d2_cneg01=[NumVL{1,10,1,5}];

Increasing_TNSPEC_neg=[r1_t50_d2_cneg01(10,:);r1_t100_d2_cneg01(10,:);...
    r1_t150_d2_cneg01(10,:);r1_t200_d2_cneg01(10,:);r1_t250_d2_cneg01(10,:);...
    r1_t300_d2_cneg01(10,:);r1_t350_d2_cneg01(10,:);r1_t400_d2_cneg01(10,:);...
    r1_t450_d2_cneg01(10,:);r1_t500_d2_cneg01(10,:)]

figure('Name','1,1,1 npa')
plot (Increasing_TNSPEC(:,7),(Increasing_TNSPEC(:,2).*100));
%gives relative error b/c true value is 1. Answer/1*100% = relative error

hold on
plot (Increasing_TNSPEC_neg(:,7),(Increasing_TNSPEC_neg(:,2).*100));
title({'Average Percent Error as Training Set Size Increases';'Dimension=2 Radfrac=0.5
Bootstrap Replicates=500'})
xlabel('Training Set Size')
ylabel('Percent Error')
legend('CP=(0,1)','CP=(0,-1)')
hold off

%Make second graph for dimension=3

r1_t50_d3_c01=[NumVL{1,1,2,1}]
r1_t100_d3_c01=[NumVL{1,2,2,1}]
r1_t150_d3_c01=[NumVL{1,3,2,1}];
r1_t200_d3_c01=[NumVL{1,4,2,1}];
r1_t250_d3_c01=[NumVL{1,5,2,1}];
r1_t300_d3_c01=[NumVL{1,6,2,1}];
r1_t350_d3_c01=[NumVL{1,7,2,1}];
r1_t400_d3_c01=[NumVL{1,8,2,1}];
r1_t450_d3_c01=[NumVL{1,9,2,1}];
r1_t500_d3_c01=[NumVL{1,10,2,1}];

Increasing_TNSPEC2=[r1_t50_d3_c01(10,:);r1_t100_d3_c01(10,:);...
    r1_t150_d3_c01(10,:);r1_t200_d3_c01(10,:);r1_t250_d3_c01(10,:);...
    r1_t300_d3_c01(10,:);r1_t350_d3_c01(10,:);r1_t400_d3_c01(10,:);...
    r1_t450_d3_c01(10,:);r1_t500_d3_c01(10,:)]

r1_t50_d3_cneg01=[NumVL{1,1,2,5}]
r1_t100_d3_cneg01=[NumVL{1,2,2,5}]
r1_t150_d3_cneg01=[NumVL{1,3,2,5}];

```

```

r1_t200_d3_cneg01=[NumVL{1,4,2,5}];
r1_t250_d3_cneg01=[NumVL{1,5,2,5}];
r1_t300_d3_cneg01=[NumVL{1,6,2,5}];
r1_t350_d3_cneg01=[NumVL{1,7,2,5}];
r1_t400_d3_cneg01=[NumVL{1,8,2,5}];
r1_t450_d3_cneg01=[NumVL{1,9,2,5}];
r1_t500_d3_cneg01=[NumVL{1,10,2,5}];

Increasing_TNSPEC2neg=[r1_t50_d3_cneg01(10,:);r1_t100_d3_cneg01(10,:);...
    r1_t150_d3_cneg01(10,:);r1_t200_d3_cneg01(10,:);r1_t250_d3_cneg01(10,:);...
    r1_t300_d3_cneg01(10,:);r1_t350_d3_cneg01(10,:);r1_t400_d3_cneg01(10,:);...
    r1_t450_d3_cneg01(10,:);r1_t500_d3_cneg01(10,:)]

figure('Name','1,1,2 npa')
plot (Increasing_TNSPEC2(:,7),(Increasing_TNSPEC2(:,2).*100));
hold on
plot (Increasing_TNSPEC2neg(:,7),(Increasing_TNSPEC2neg(:,2).*100));
title({'Average Percent Error as Training Set Size Increases';...
    'Dimension=3 Radfrac=0.5 Bootstrap Replicates=500'})
xlabel('Training Set Size')
ylabel('Percent Error')
legend('CP=(0,1)', 'CP=(0,-1)')
hold off

%Make a third graph for 4 dimensions
r1_t50_d4_c01=[NumVL{1,1,3,1}];
r1_t100_d4_c01=[NumVL{1,2,3,1}];
r1_t150_d4_c01=[NumVL{1,3,3,1}];
r1_t200_d4_c01=[NumVL{1,4,3,1}];
r1_t250_d4_c01=[NumVL{1,5,3,1}];
r1_t300_d4_c01=[NumVL{1,6,3,1}];
r1_t350_d4_c01=[NumVL{1,7,3,1}];
r1_t400_d4_c01=[NumVL{1,8,3,1}];
r1_t450_d4_c01=[NumVL{1,9,3,1}];
r1_t500_d4_c01=[NumVL{1,10,3,1}];

Increasing_TNSPEC3=[r1_t50_d4_c01(10,:);r1_t100_d4_c01(10,:);...
    r1_t150_d4_c01(10,:);r1_t200_d4_c01(10,:);r1_t250_d4_c01(10,:);...
    r1_t300_d4_c01(10,:);r1_t350_d4_c01(10,:);r1_t400_d4_c01(10,:);...
    r1_t450_d4_c01(10,:);r1_t500_d4_c01(10,:)]

r1_t50_d4_cneg01=[NumVL{1,1,3,5}];

```

```

r1_t100_d4_cneg01=[NumVL{1,2,3,5}];
r1_t150_d4_cneg01=[NumVL{1,3,3,5}];
r1_t200_d4_cneg01=[NumVL{1,4,3,5}];
r1_t250_d4_cneg01=[NumVL{1,5,3,5}];
r1_t300_d4_cneg01=[NumVL{1,6,3,5}];
r1_t350_d4_cneg01=[NumVL{1,7,3,5}];
r1_t400_d4_cneg01=[NumVL{1,8,3,5}];
r1_t450_d4_cneg01=[NumVL{1,9,3,5}];
r1_t500_d4_cneg01=[NumVL{1,10,3,5}];

Increasing_TNSPEC3neg=[r1_t50_d4_cneg01(10,:);r1_t100_d4_cneg01(10,:);...
    r1_t150_d4_cneg01(10,:);r1_t200_d4_cneg01(10,:);r1_t250_d4_cneg01(10,:);...
    r1_t300_d4_cneg01(10,:);r1_t350_d4_cneg01(10,:);r1_t400_d4_cneg01(10,:);...
    r1_t450_d4_cneg01(10,:);r1_t500_d4_cneg01(10,:)]

figure('Name','1,1,3 npa')
plot (Increasing_TNSPEC3neg(:,7),(Increasing_TNSPEC3(:,2).*100));
hold on
plot (Increasing_TNSPEC_neg(:,7),(Increasing_TNSPEC_neg(:,2).*100));
title({'Average Percent Error as Training Set Size Increases';...
    'Dimension=4 Radfrac=0.5 Bootstrap Replicates=500'})
xlabel('Training Set Size')
ylabel('Percent Error')
legend('CP=(0,1)','CP=(0,-1)')
hold off

%Make a graph for 11 dimensions
r1_t50_d11_c01=[NumVL{1,1,10,1}];
r1_t100_d11_c01=[NumVL{1,2,10,1}];
r1_t150_d11_c01=[NumVL{1,3,10,1}];
r1_t200_d11_c01=[NumVL{1,4,10,1}];
r1_t250_d11_c01=[NumVL{1,5,10,1}];
r1_t300_d11_c01=[NumVL{1,6,10,1}];
r1_t350_d11_c01=[NumVL{1,7,10,1}];
r1_t400_d11_c01=[NumVL{1,8,10,1}];
r1_t450_d11_c01=[NumVL{1,9,10,1}];
r1_t500_d11_c01=[NumVL{1,10,10,1}];

Increasing_TNSPEC11=[r1_t50_d11_c01(10,:);r1_t100_d11_c01(10,:);...
    r1_t150_d11_c01(10,:);r1_t200_d11_c01(10,:);r1_t250_d11_c01(10,:);...
    r1_t300_d11_c01(10,:);r1_t350_d11_c01(10,:);r1_t400_d11_c01(10,:);...
    r1_t450_d11_c01(10,:);r1_t500_d11_c01(10,:)]

```

```

r1_t50_d11_cneg01=[NumVL{1,1,10,5}];
r1_t100_d11_cneg01=[NumVL{1,2,10,5}];
r1_t150_d11_cneg01=[NumVL{1,3,10,5}];
r1_t200_d11_cneg01=[NumVL{1,4,10,5}];
r1_t250_d11_cneg01=[NumVL{1,5,10,5}];
r1_t300_d11_cneg01=[NumVL{1,6,10,5}];
r1_t350_d11_cneg01=[NumVL{1,7,10,5}];
r1_t400_d11_cneg01=[NumVL{1,8,10,5}];
r1_t450_d11_cneg01=[NumVL{1,9,10,5}];
r1_t500_d11_cneg01=[NumVL{1,10,10,5}];

```

```

Increasing_TNSPEC11neg=[r1_t50_d11_cneg01(10,:);r1_t100_d11_cneg01(10,:);...
    r1_t150_d11_cneg01(10,:);r1_t200_d11_cneg01(10,:);r1_t250_d11_cneg01(10,:);...
    r1_t300_d11_cneg01(10,:);r1_t350_d11_cneg01(10,:);r1_t400_d11_cneg01(10,:);...
    r1_t450_d11_cneg01(10,:);r1_t500_d11_cneg01(10,:)]

```

```

figure('Name','1,1,10 npa')
plot (Increasing_TNSPEC11(:,7),(Increasing_TNSPEC11(:,2).*100));
hold on
plot (Increasing_TNSPEC11neg(:,7),(Increasing_TNSPEC11neg(:,2).*100));
title({'Average Percent Error as Training Set Size Increases';'Dimension=11 Radfrac=0.5
Bootstrap Replicates=500'})
xlabel('Training Set Size')
ylabel('Percent Error')
legend('CP=(0,1)', 'CP=(0,-1)')
hold off

```

Plotting_Code_Increase_Radfrac_Bootstrap

%This code plots bootstrapping vs. bias as radfrac increases

%Indexing in NumVL Cell Matrix is as Follows:

%{radfrac number, training, dimension - 1, compass point
%number}

%Within matrix:

%B is located in Bias_plot(:,1)

%bias is located in Bias_plot(:,2)

%dimension is located in Bias_plot(:,3)

%First 2 coordinates of compass point located in Bias_plot(:,4) and (:,5)

%radfrac located in Bias_plot(:,6)

%Training set size located in Bias_plot(:,7)

%Set Axis Limits for All Graphs:

graphlimits=[50,500,-40,40];

figure('Name','1,1,1 npa')

r1_t50_d2_c01=[NumVL{1,1,1,1}]

%x 100 to make a percent error

r1_t50_d2_c01per=[r1_t50_d2_c01(:,:).*100]

%x is out of original plot, y out of x 100

plot (r1_t50_d2_c01(:,1),r1_t50_d2_c01per(:,2));

hold on

r1_t50_d2_cneg01=[NumVL{1,1,1,5}]

%x 100 to make a percent error

r1_t50_d2_cneg01per=[r1_t50_d2_cneg01(:,:).*100]

%x is out of original plot, y out of x 100

plot (r1_t50_d2_cneg01(:,1),r1_t50_d2_cneg01per(:,2));

title({'Average Percent Error as Bootstrap Replicates Increase';'Radfrac=0.5

Dimension=2 TNSPEC=50'})

xlabel('Bootstrap Replications')

ylabel('Percent Error')

legend('CP=(0,1)','CP=(0,-1)')

axis(graphlimits)

hold off

figure('Name','2,1,1 npa')

r2_t50_d2_c01=[NumVL{2,1,1,1}];

```

%x 100 to make a percent error
r2_t50_d2_c01per=[r2_t50_d2_c01(:,:).*100]
%x is out of original plot, y out of x 100
plot (r2_t50_d2_c01(:,1),r2_t50_d2_c01per(:,2));
    hold on
r2_t50_d2_cneg01=[NumVL{2,1,1,5}];
%x 100 to make a percent error
r2_t50_d2_cneg01per=[r2_t50_d2_cneg01(:,:).*100]
%x is out of original plot, y out of x 100
plot (r2_t50_d2_cneg01(:,1),r2_t50_d2_cneg01per(:,2));
title({'Average Percent Error as Bootstrap Replicates Increase';'Radfrac=0.6
Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Percent Error')
legend('CP=(0,1)', 'CP=(0,-1)')
axis(graphlimits)
hold off

```

```

figure('Name','3,1,1 npa')
r3_t50_d2_c01=[NumVL{3,1,1,1}]
%x 100 to make a percent error
r3_t50_d2_c01per=[r3_t50_d2_c01(:,:).*100]
%x is out of original plot, y out of x 100
plot (r3_t50_d2_c01(:,1),r3_t50_d2_c01per(:,2));

```

```

hold on
r3_t50_d2_cneg01=[NumVL{3,1,1,5}];
%x 100 to make a percent error
r3_t50_d2_cneg01per=[r3_t50_d2_cneg01(:,:).*100];
%x is out of original plot, y out of x 100
plot (r3_t50_d2_cneg01(:,1),r3_t50_d2_cneg01per(:,2));
title({'Average Percent Error as Bootstrap Replicates Increase';'Radfrac=0.7
Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Percent Error')
legend('CP=(0,1)', 'CP=(0,-1)')
axis(graphlimits)
hold off

```

```

figure('Name','4,1,1 npa')
r4_t50_d2_c01=[NumVL{4,1,1,1}]
%x 100 to make a percent error
r4_t50_d2_c01per=[r4_t50_d2_c01(:,:).*100]

```

```

%x is out of original plot, y out of x 100
plot (r4_t50_d2_c01(:,1),r4_t50_d2_c01per(:,2));
hold on
r4_t50_d2_cneg01=[NumVL{4,1,1,5}];
%x 100 to make a percent error
r4_t50_d2_cneg01per=[r4_t50_d2_cneg01(:,:).*100]
%x is out of original plot, y out of x 100
plot (r4_t50_d2_cneg01(:,1),r4_t50_d2_cneg01per(:,2));
title({'Average Percent Error as Bootstrap Replicates Increase';...
      'Radfrac=0.8 Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Percent Error')
legend('CP=(0,1)','CP=(0,-1)')
axis(graphlimits)
hold off

figure('Name','5,1,1 npa')
r5_t50_d2_c01=[NumVL{5,1,1,1}]
%x 100 to make a percent error
r5_t50_d2_c01per=[r5_t50_d2_c01(:,:).*100]
%x is out of original plot, y out of x 100
plot (r5_t50_d2_c01(:,1),r5_t50_d2_c01per(:,2));
hold on
r5_t50_d2_cneg01=[NumVL{5,1,1,5}]
%x 100 to make a percent error
r5_t50_d2_cneg01per=[r5_t50_d2_cneg01(:,:).*100];
%x is out of original plot, y out of x 100
plot (r5_t50_d2_cneg01(:,1),r5_t50_d2_cneg01per(:,2));
title({'Average Percent Error as Bootstrap Replicates Increase';...
      'Radfrac=0.9 Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Percent Error')
legend('CP=(0,1)','CP=(0,-1)')
axis(graphlimits)
hold off

%Generate Last Figure
figure('Name','6,1,1 npa')
r6_t50_d2_c01=[NumVL{6,1,1,1}]
%Out of matrix variables to assign, so is being multiplied
%x 100 to make a percent error
%x 100 to make a percent error
r6_t50_d2_c01per=[r6_t50_d2_c01(:,:).*100]

```

```

%x is out of original plot, y out of x 100
plot(r6_t50_d2_c01(:,1),r6_t50_d2_c01per(:,2))
hold on
r6_t50_d2_cneg01=[NumVL{6,1,1,5}];
%x 100 to make a percent error
r6_t50_d2_cneg01per=[r6_t50_d2_cneg01(:,:).*100];
%x is out of original plot, y out of x 100
plot(r6_t50_d2_cneg01(:,1),r6_t50_d2_cneg01per(:,2))
title({'Average Percent Error as Bootstrap Replicates Increase';...
      'Radfrac=1.0 Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Percent Error')
legend('CP=(0,1)','CP=(0,-1)')
axis(graphlimits)
hold off

```


Plotting_Code_Increase_Bootstrap_Radfrac

%This code plots bootstrapping vs. bias as radfrac increases

%Indexing in NumVL Cell Matrix is as Follows:

%{radfrac number, training, dimension - 1, compass point
%number}

%Within matrix:

%B is located in Bias_plot(:,1)

%bias is located in Bias_plot(:,2)

%dimension is located in Bias_plot(:,3)

%First 2 coordinates of compass point located in Bias_plot(:,4) and (:,5)

%radfrac located in Bias_plot(:,6)

%Training set size located in Bias_plot(:,7)

%Set Axis Limits for All Graphs:

graphlimits=[50,500,-40,40];

figure('Name','1,1,1 npa')

r1_t50_d2_c01=[NumVL{1,1,1,1}]

%x 100 to make a percent error

r1_t50_d2_c01per=[r1_t50_d2_c01(:,:).*100]

%x is out of original plot, y out of x 100

plot(r1_t50_d2_c01(:,1),r1_t50_d2_c01per(:,2));

hold on

r1_t50_d2_cneg01=[NumVL{1,1,1,5}]

%x 100 to make a percent error

r1_t50_d2_cneg01per=[r1_t50_d2_cneg01(:,:).*100]

%x is out of original plot, y out of x 100

plot(r1_t50_d2_cneg01(:,1),r1_t50_d2_cneg01per(:,2));

title({'Average Percent Error as Bootstrap Replicates Increase';'Radfrac=0.5

Dimension=2 TNSPEC=50'})

xlabel('Bootstrap Replications')

ylabel('Percent Error')

legend('CP=(0,1)', 'CP=(0,-1)')

axis(graphlimits)

hold off

figure('Name','2,1,1 npa')

```

r2_t50_d2_c01=[NumVL{2,1,1,1}];
%x 100 to make a percent error
r2_t50_d2_c01per=[r2_t50_d2_c01(:,:).*100]
%x is out of original plot, y out of x 100
plot (r2_t50_d2_c01(:,1),r2_t50_d2_c01per(:,2));
    hold on
r2_t50_d2_cneg01=[NumVL{2,1,1,5}];
%x 100 to make a percent error
r2_t50_d2_cneg01per=[r2_t50_d2_cneg01(:,:).*100]
%x is out of original plot, y out of x 100
plot (r2_t50_d2_cneg01(:,1),r2_t50_d2_cneg01per(:,2));
title({'Average Percent Error as Bootstrap Replicates Increase';'Radfrac=0.6
Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Percent Error')
legend('CP=(0,1)', 'CP=(0,-1)')
axis(graphlimits)
hold off

```

```

figure('Name','3,1,1 npa')
r3_t50_d2_c01=[NumVL{3,1,1,1}]
%x 100 to make a percent error
r3_t50_d2_c01per=[r3_t50_d2_c01(:,:).*100]
%x is out of original plot, y out of x 100
plot (r3_t50_d2_c01(:,1),r3_t50_d2_c01per(:,2));

hold on
r3_t50_d2_cneg01=[NumVL{3,1,1,5}];
%x 100 to make a percent error
r3_t50_d2_cneg01per=[r3_t50_d2_cneg01(:,:).*100];
%x is out of original plot, y out of x 100
plot (r3_t50_d2_cneg01(:,1),r3_t50_d2_cneg01per(:,2));
title({'Average Percent Error as Bootstrap Replicates Increase';'Radfrac=0.7
Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Percent Error')
legend('CP=(0,1)', 'CP=(0,-1)')
axis(graphlimits)
hold off

```

```

figure('Name','4,1,1 npa')
r4_t50_d2_c01=[NumVL{4,1,1,1}]
%x 100 to make a percent error

```

```

r4_t50_d2_c01per=[r4_t50_d2_c01(:,:).*100]
%x is out of original plot, y out of x 100
plot (r4_t50_d2_c01(:,1),r4_t50_d2_c01per(:,2));
hold on
r4_t50_d2_cneg01=[NumVL{4,1,1,5}];
%x 100 to make a percent error
r4_t50_d2_cneg01per=[r4_t50_d2_cneg01(:,:).*100]
%x is out of original plot, y out of x 100
plot (r4_t50_d2_cneg01(:,1),r4_t50_d2_cneg01per(:,2));
title({'Average Percent Error as Bootstrap Replicates Increase';...
'Radfrac=0.8 Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Percent Error')
legend('CP=(0,1)', 'CP=(0,-1)')
axis(graphlimits)
hold off

figure('Name','5,1,1 npa')
r5_t50_d2_c01=[NumVL{5,1,1,1}]
%x 100 to make a percent error
r5_t50_d2_c01per=[r5_t50_d2_c01(:,:).*100]
%x is out of original plot, y out of x 100
plot (r5_t50_d2_c01(:,1),r5_t50_d2_c01per(:,2));
hold on
r5_t50_d2_cneg01=[NumVL{5,1,1,5}]
%x 100 to make a percent error
r5_t50_d2_cneg01per=[r5_t50_d2_cneg01(:,:).*100];
%x is out of original plot, y out of x 100
plot (r5_t50_d2_cneg01(:,1),r5_t50_d2_cneg01per(:,2));
title({'Average Percent Error as Bootstrap Replicates Increase';...
'Radfrac=0.9 Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Percent Error')
legend('CP=(0,1)', 'CP=(0,-1)')
axis(graphlimits)
hold off

%Generate Last Figure
figure('Name','6,1,1 npa')
r6_t50_d2_c01=[NumVL{6,1,1,1}]
%Out of matrix variables to assign, so is being multiplied
%x 100 to make a percent error
%x 100 to make a percent error

```

```

r6_t50_d2_c01per=[r6_t50_d2_c01(:,:).*100]
%x is out of original plot, y out of x 100
plot(r6_t50_d2_c01(:,1),r6_t50_d2_c01per(:,2))
hold on
r6_t50_d2_cneg01=[NumVL{6,1,1,5}];
%x 100 to make a percent error
r6_t50_d2_cneg01per=[r6_t50_d2_cneg01(:,:).*100];
%x is out of original plot, y out of x 100
plot(r6_t50_d2_cneg01(:,1),r6_t50_d2_cneg01per(:,2))
title({'Average Percent Error as Bootstrap Replicates Increase';...
      'Radfrac=1.0 Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Percent Error')
legend('CP=(0,1)', 'CP=(0,-1)')
axis(graphlimits)
hold off

```

Plotting_Code_Increase_TNSPEC_Dimension

%This code plots tnspec vs. bias as dimension increases

%Indexing in NumVL Cell Matrix is as Follows:

%{radfrac number, training, dimension - 1, compass point
%number}

%Within matrix:

%B is located in Bias_plot(:,1)

%bias is located in Bias_plot(:,2)

%dimension is located in Bias_plot(:,3)

%First 2 coordinates of compass point located in Bias_plot(:,4) and (:,5)

%radfrac located in Bias_plot(:,6)

%Training set size located in Bias_plot(:,7)

%Get all for (0,1) Compass Points

r1_t50_d2_c01=[NumVL{1,1,1,1}];
r1_t100_d2_c01=[NumVL{1,2,1,1}];
r1_t150_d2_c01=[NumVL{1,3,1,1}];
r1_t200_d2_c01=[NumVL{1,4,1,1}];
r1_t250_d2_c01=[NumVL{1,5,1,1}];
r1_t300_d2_c01=[NumVL{1,6,1,1}];
r1_t350_d2_c01=[NumVL{1,7,1,1}];
r1_t400_d2_c01=[NumVL{1,8,1,1}];
r1_t450_d2_c01=[NumVL{1,9,1,1}];
r1_t500_d2_c01=[NumVL{1,10,1,1}];

Increasing_TNSPEC=[r1_t50_d2_c01(10,:);r1_t100_d2_c01(10,:);...
r1_t150_d2_c01(10,:);r1_t200_d2_c01(10,:);r1_t250_d2_c01(10,:);...
r1_t300_d2_c01(10,:);r1_t350_d2_c01(10,:);r1_t400_d2_c01(10,:);...
r1_t450_d2_c01(10,:);r1_t500_d2_c01(10,:)]

%Get all (0,-1) Compass Points

r1_t50_d2_cneg01=[NumVL{1,1,1,5}];
r1_t100_d2_cneg01=[NumVL{1,2,1,5}];
r1_t150_d2_cneg01=[NumVL{1,3,1,5}];
r1_t200_d2_cneg01=[NumVL{1,4,1,5}];
r1_t250_d2_cneg01=[NumVL{1,5,1,5}];
r1_t300_d2_cneg01=[NumVL{1,6,1,5}];
r1_t350_d2_cneg01=[NumVL{1,7,1,5}];
r1_t400_d2_cneg01=[NumVL{1,8,1,5}];

```

r1_t450_d2_cneg01=[NumVL{1,9,1,5}];
r1_t500_d2_cneg01=[NumVL{1,10,1,5}];

Increasing_TNSPEC_neg=[r1_t50_d2_cneg01(10,:);r1_t100_d2_cneg01(10,:);...
    r1_t150_d2_cneg01(10,:);r1_t200_d2_cneg01(10,:);r1_t250_d2_cneg01(10,:);...
    r1_t300_d2_cneg01(10,:);r1_t350_d2_cneg01(10,:);r1_t400_d2_cneg01(10,:);...
    r1_t450_d2_cneg01(10,:);r1_t500_d2_cneg01(10,:)]

figure('Name','1,1,1 npa')
plot (Increasing_TNSPEC(:,7),(Increasing_TNSPEC(:,2).*100));
%gives relative error b/c true value is 1. Answer/1*100% = relative error

hold on
plot (Increasing_TNSPEC_neg(:,7),(Increasing_TNSPEC_neg(:,2).*100));
title({'Average Percent Error as Training Set Size Increases';'Dimension=2 Radfrac=0.5
Bootstrap Replicates=500'})
xlabel('Training Set Size')
ylabel('Percent Error')
legend('CP=(0,1)','CP=(0,-1)')
hold off

%Make second graph for dimension=3

r1_t50_d3_c01=[NumVL{1,1,2,1}]
r1_t100_d3_c01=[NumVL{1,2,2,1}]
r1_t150_d3_c01=[NumVL{1,3,2,1}];
r1_t200_d3_c01=[NumVL{1,4,2,1}];
r1_t250_d3_c01=[NumVL{1,5,2,1}];
r1_t300_d3_c01=[NumVL{1,6,2,1}];
r1_t350_d3_c01=[NumVL{1,7,2,1}];
r1_t400_d3_c01=[NumVL{1,8,2,1}];
r1_t450_d3_c01=[NumVL{1,9,2,1}];
r1_t500_d3_c01=[NumVL{1,10,2,1}];

Increasing_TNSPEC2=[r1_t50_d3_c01(10,:);r1_t100_d3_c01(10,:);...
    r1_t150_d3_c01(10,:);r1_t200_d3_c01(10,:);r1_t250_d3_c01(10,:);...
    r1_t300_d3_c01(10,:);r1_t350_d3_c01(10,:);r1_t400_d3_c01(10,:);...
    r1_t450_d3_c01(10,:);r1_t500_d3_c01(10,:)]

r1_t50_d3_cneg01=[NumVL{1,1,2,5}]
r1_t100_d3_cneg01=[NumVL{1,2,2,5}]
r1_t150_d3_cneg01=[NumVL{1,3,2,5}];

```

```

r1_t200_d3_cneg01=[NumVL{1,4,2,5}];
r1_t250_d3_cneg01=[NumVL{1,5,2,5}];
r1_t300_d3_cneg01=[NumVL{1,6,2,5}];
r1_t350_d3_cneg01=[NumVL{1,7,2,5}];
r1_t400_d3_cneg01=[NumVL{1,8,2,5}];
r1_t450_d3_cneg01=[NumVL{1,9,2,5}];
r1_t500_d3_cneg01=[NumVL{1,10,2,5}];

Increasing_TNSPEC2neg=[r1_t50_d3_cneg01(10,:);r1_t100_d3_cneg01(10,:);...
    r1_t150_d3_cneg01(10,:);r1_t200_d3_cneg01(10,:);r1_t250_d3_cneg01(10,:);...
    r1_t300_d3_cneg01(10,:);r1_t350_d3_cneg01(10,:);r1_t400_d3_cneg01(10,:);...
    r1_t450_d3_cneg01(10,:);r1_t500_d3_cneg01(10,:)]

figure('Name','1,1,2 npa')
plot (Increasing_TNSPEC2(:,7),(Increasing_TNSPEC2(:,2).*100));
hold on
plot (Increasing_TNSPEC2neg(:,7),(Increasing_TNSPEC2neg(:,2).*100));
title({'Average Percent Error as Training Set Size Increases';...
    'Dimension=3 Radfrac=0.5 Bootstrap Replicates=500'})
xlabel('Training Set Size')
ylabel('Percent Error')
legend('CP=(0,1)', 'CP=(0,-1)')
hold off

%Make a third graph for 4 dimensions
r1_t50_d4_c01=[NumVL{1,1,3,1}];
r1_t100_d4_c01=[NumVL{1,2,3,1}];
r1_t150_d4_c01=[NumVL{1,3,3,1}];
r1_t200_d4_c01=[NumVL{1,4,3,1}];
r1_t250_d4_c01=[NumVL{1,5,3,1}];
r1_t300_d4_c01=[NumVL{1,6,3,1}];
r1_t350_d4_c01=[NumVL{1,7,3,1}];
r1_t400_d4_c01=[NumVL{1,8,3,1}];
r1_t450_d4_c01=[NumVL{1,9,3,1}];
r1_t500_d4_c01=[NumVL{1,10,3,1}];

Increasing_TNSPEC3=[r1_t50_d4_c01(10,:);r1_t100_d4_c01(10,:);...
    r1_t150_d4_c01(10,:);r1_t200_d4_c01(10,:);r1_t250_d4_c01(10,:);...
    r1_t300_d4_c01(10,:);r1_t350_d4_c01(10,:);r1_t400_d4_c01(10,:);...
    r1_t450_d4_c01(10,:);r1_t500_d4_c01(10,:)]

r1_t50_d4_cneg01=[NumVL{1,1,3,5}];

```

```

r1_t100_d4_cneg01=[NumVL{1,2,3,5}];
r1_t150_d4_cneg01=[NumVL{1,3,3,5}];
r1_t200_d4_cneg01=[NumVL{1,4,3,5}];
r1_t250_d4_cneg01=[NumVL{1,5,3,5}];
r1_t300_d4_cneg01=[NumVL{1,6,3,5}];
r1_t350_d4_cneg01=[NumVL{1,7,3,5}];
r1_t400_d4_cneg01=[NumVL{1,8,3,5}];
r1_t450_d4_cneg01=[NumVL{1,9,3,5}];
r1_t500_d4_cneg01=[NumVL{1,10,3,5}];

Increasing_TNSPEC3neg=[r1_t50_d4_cneg01(10,:);r1_t100_d4_cneg01(10,:);...
    r1_t150_d4_cneg01(10,:);r1_t200_d4_cneg01(10,:);r1_t250_d4_cneg01(10,:);...
    r1_t300_d4_cneg01(10,:);r1_t350_d4_cneg01(10,:);r1_t400_d4_cneg01(10,:);...
    r1_t450_d4_cneg01(10,:);r1_t500_d4_cneg01(10,:)]

figure('Name','1,1,3 npa')
plot (Increasing_TNSPEC3neg(:,7),(Increasing_TNSPEC3(:,2).*100));
hold on
plot (Increasing_TNSPEC_neg(:,7),(Increasing_TNSPEC_neg(:,2).*100));
title({'Average Percent Error as Training Set Size Increases';...
    'Dimension=4 Radfrac=0.5 Bootstrap Replicates=500'})
xlabel('Training Set Size')
ylabel('Percent Error')
legend('CP=(0,1)','CP=(0,-1)')
hold off

%Make a graph for 11 dimensions
r1_t50_d11_c01=[NumVL{1,1,10,1}];
r1_t100_d11_c01=[NumVL{1,2,10,1}];
r1_t150_d11_c01=[NumVL{1,3,10,1}];
r1_t200_d11_c01=[NumVL{1,4,10,1}];
r1_t250_d11_c01=[NumVL{1,5,10,1}];
r1_t300_d11_c01=[NumVL{1,6,10,1}];
r1_t350_d11_c01=[NumVL{1,7,10,1}];
r1_t400_d11_c01=[NumVL{1,8,10,1}];
r1_t450_d11_c01=[NumVL{1,9,10,1}];
r1_t500_d11_c01=[NumVL{1,10,10,1}];

Increasing_TNSPEC11=[r1_t50_d11_c01(10,:);r1_t100_d11_c01(10,:);...
    r1_t150_d11_c01(10,:);r1_t200_d11_c01(10,:);r1_t250_d11_c01(10,:);...
    r1_t300_d11_c01(10,:);r1_t350_d11_c01(10,:);r1_t400_d11_c01(10,:);...
    r1_t450_d11_c01(10,:);r1_t500_d11_c01(10,:)]

```



```

r1_t50_d11_cneg01=[NumVL{1,1,10,5}];
r1_t100_d11_cneg01=[NumVL{1,2,10,5}];
r1_t150_d11_cneg01=[NumVL{1,3,10,5}];
r1_t200_d11_cneg01=[NumVL{1,4,10,5}];
r1_t250_d11_cneg01=[NumVL{1,5,10,5}];
r1_t300_d11_cneg01=[NumVL{1,6,10,5}];
r1_t350_d11_cneg01=[NumVL{1,7,10,5}];
r1_t400_d11_cneg01=[NumVL{1,8,10,5}];
r1_t450_d11_cneg01=[NumVL{1,9,10,5}];
r1_t500_d11_cneg01=[NumVL{1,10,10,5}];

Increasing_TNSPEC11neg=[r1_t50_d11_cneg01(10,:);r1_t100_d11_cneg01(10,:);...
    r1_t150_d11_cneg01(10,:);r1_t200_d11_cneg01(10,:);r1_t250_d11_cneg01(10,:);...
    r1_t300_d11_cneg01(10,:);r1_t350_d11_cneg01(10,:);r1_t400_d11_cneg01(10,:);...
    r1_t450_d11_cneg01(10,:);r1_t500_d11_cneg01(10,:)]

figure('Name','1,1,10 npa')
plot (Increasing_TNSPEC11(:,7),(Increasing_TNSPEC11(:,2).*100));
hold on
plot (Increasing_TNSPEC11neg(:,7),(Increasing_TNSPEC11neg(:,2).*100));
title({'Average Percent Error as Training Set Size Increases';'Dimension=11 Radfrac=0.5
Bootstrap Replicates=500'})
xlabel('Training Set Size')
ylabel('Percent Error')
legend('CP=(0,1)', 'CP=(0,-1)')
hold off

```

Plotting_Code_Increase_TNSPEC_Dimension

%This code plots tnspec vs. bias as dimension increases

%Indexing in NumVL Cell Matrix is as Follows:

%{radfrac number, training, dimension - 1, compass point
%number}

%Within matrix:

%B is located in Bias_plot(:,1)

%bias is located in Bias_plot(:,2)

%dimension is located in Bias_plot(:,3)

%First 2 coordinates of compass point located in Bias_plot(:,4) and (:,5)

%radfrac located in Bias_plot(:,6)

%Training set size located in Bias_plot(:,7)

%Get all for (0,1) Compass Points

r1_t50_d2_c01=[NumVL{1,1,1,1}];

r1_t100_d2_c01=[NumVL{1,2,1,1}];

r1_t150_d2_c01=[NumVL{1,3,1,1}];

r1_t200_d2_c01=[NumVL{1,4,1,1}];

r1_t250_d2_c01=[NumVL{1,5,1,1}];

r1_t300_d2_c01=[NumVL{1,6,1,1}];

r1_t350_d2_c01=[NumVL{1,7,1,1}];

r1_t400_d2_c01=[NumVL{1,8,1,1}];

r1_t450_d2_c01=[NumVL{1,9,1,1}];

r1_t500_d2_c01=[NumVL{1,10,1,1}];

Increasing_TNSPEC=[r1_t50_d2_c01(10,:);r1_t100_d2_c01(10,:);...

r1_t150_d2_c01(10,:);r1_t200_d2_c01(10,:);r1_t250_d2_c01(10,:);...

r1_t300_d2_c01(10,:);r1_t350_d2_c01(10,:);r1_t400_d2_c01(10,:);...

r1_t450_d2_c01(10,:);r1_t500_d2_c01(10,:)]

%Get all (0,-1) Compass Points

r1_t50_d2_cneg01=[NumVL{1,1,1,5}];

r1_t100_d2_cneg01=[NumVL{1,2,1,5}];

r1_t150_d2_cneg01=[NumVL{1,3,1,5}];

r1_t200_d2_cneg01=[NumVL{1,4,1,5}];

r1_t250_d2_cneg01=[NumVL{1,5,1,5}];

r1_t300_d2_cneg01=[NumVL{1,6,1,5}];

r1_t350_d2_cneg01=[NumVL{1,7,1,5}];

```

r1_t400_d2_cneg01=[NumVL{1,8,1,5}];
r1_t450_d2_cneg01=[NumVL{1,9,1,5}];
r1_t500_d2_cneg01=[NumVL{1,10,1,5}];

Increasing_TNSPEC_neg=[r1_t50_d2_cneg01(10,:);r1_t100_d2_cneg01(10,:);...
    r1_t150_d2_cneg01(10,:);r1_t200_d2_cneg01(10,:);r1_t250_d2_cneg01(10,:);...
    r1_t300_d2_cneg01(10,:);r1_t350_d2_cneg01(10,:);r1_t400_d2_cneg01(10,:);...
    r1_t450_d2_cneg01(10,:);r1_t500_d2_cneg01(10,:)]

figure('Name','1,1,1 npa')
plot (Increasing_TNSPEC(:,7),(Increasing_TNSPEC(:,2).*100));
%gives relative error b/c true value is 1. Answer/1*100% = relative error

hold on
plot (Increasing_TNSPEC_neg(:,7),(Increasing_TNSPEC_neg(:,2).*100));
title({'Average Percent Error as Training Set Size Increases';'Dimension=2 Radfrac=0.5
Bootstrap Replicates=500'})
xlabel('Training Set Size')
ylabel('Percent Error')
legend('CP=(0,1)','CP=(0,-1)')
hold off

%Make second graph for dimension=3

r1_t50_d3_c01=[NumVL{1,1,2,1}]
r1_t100_d3_c01=[NumVL{1,2,2,1}]
r1_t150_d3_c01=[NumVL{1,3,2,1}];
r1_t200_d3_c01=[NumVL{1,4,2,1}];
r1_t250_d3_c01=[NumVL{1,5,2,1}];
r1_t300_d3_c01=[NumVL{1,6,2,1}];
r1_t350_d3_c01=[NumVL{1,7,2,1}];
r1_t400_d3_c01=[NumVL{1,8,2,1}];
r1_t450_d3_c01=[NumVL{1,9,2,1}];
r1_t500_d3_c01=[NumVL{1,10,2,1}];

Increasing_TNSPEC2=[r1_t50_d3_c01(10,:);r1_t100_d3_c01(10,:);...
    r1_t150_d3_c01(10,:);r1_t200_d3_c01(10,:);r1_t250_d3_c01(10,:);...
    r1_t300_d3_c01(10,:);r1_t350_d3_c01(10,:);r1_t400_d3_c01(10,:);...
    r1_t450_d3_c01(10,:);r1_t500_d3_c01(10,:)]

r1_t50_d3_cneg01=[NumVL{1,1,2,5}]
r1_t100_d3_cneg01=[NumVL{1,2,2,5}]

```

```

r1_t150_d3_cneg01=[NumVL{1,3,2,5}];
r1_t200_d3_cneg01=[NumVL{1,4,2,5}];
r1_t250_d3_cneg01=[NumVL{1,5,2,5}];
r1_t300_d3_cneg01=[NumVL{1,6,2,5}];
r1_t350_d3_cneg01=[NumVL{1,7,2,5}];
r1_t400_d3_cneg01=[NumVL{1,8,2,5}];
r1_t450_d3_cneg01=[NumVL{1,9,2,5}];
r1_t500_d3_cneg01=[NumVL{1,10,2,5}];

```

```

Increasing_TNSPEC2neg=[r1_t50_d3_cneg01(10,:);r1_t100_d3_cneg01(10,:);...
    r1_t150_d3_cneg01(10,:);r1_t200_d3_cneg01(10,:);r1_t250_d3_cneg01(10,:);...
    r1_t300_d3_cneg01(10,:);r1_t350_d3_cneg01(10,:);r1_t400_d3_cneg01(10,:);...
    r1_t450_d3_cneg01(10,:);r1_t500_d3_cneg01(10,:)]

```

```

figure('Name','1,1,2 npa')
plot (Increasing_TNSPEC2(:,7),(Increasing_TNSPEC2(:,2).*100));
hold on
plot (Increasing_TNSPEC2neg(:,7),(Increasing_TNSPEC2neg(:,2).*100));
title({'Average Percent Error as Training Set Size Increases';...
    'Dimension=3 Radfrac=0.5 Bootstrap Replicates=500'})
xlabel('Training Set Size')
ylabel('Percent Error')
legend('CP=(0,1)','CP=(0,-1)')
hold off

```

```

%Make a third graph for 4 dimensions
r1_t50_d4_c01=[NumVL{1,1,3,1}];
r1_t100_d4_c01=[NumVL{1,2,3,1}];
r1_t150_d4_c01=[NumVL{1,3,3,1}];
r1_t200_d4_c01=[NumVL{1,4,3,1}];
r1_t250_d4_c01=[NumVL{1,5,3,1}];
r1_t300_d4_c01=[NumVL{1,6,3,1}];
r1_t350_d4_c01=[NumVL{1,7,3,1}];
r1_t400_d4_c01=[NumVL{1,8,3,1}];
r1_t450_d4_c01=[NumVL{1,9,3,1}];
r1_t500_d4_c01=[NumVL{1,10,3,1}];

```

```

Increasing_TNSPEC3=[r1_t50_d4_c01(10,:);r1_t100_d4_c01(10,:);...
    r1_t150_d4_c01(10,:);r1_t200_d4_c01(10,:);r1_t250_d4_c01(10,:);...
    r1_t300_d4_c01(10,:);r1_t350_d4_c01(10,:);r1_t400_d4_c01(10,:);...
    r1_t450_d4_c01(10,:);r1_t500_d4_c01(10,:)]

```

```

r1_t50_d4_cneg01=[NumVL{1,1,3,5}];
r1_t100_d4_cneg01=[NumVL{1,2,3,5}];
r1_t150_d4_cneg01=[NumVL{1,3,3,5}];
r1_t200_d4_cneg01=[NumVL{1,4,3,5}];
r1_t250_d4_cneg01=[NumVL{1,5,3,5}];
r1_t300_d4_cneg01=[NumVL{1,6,3,5}];
r1_t350_d4_cneg01=[NumVL{1,7,3,5}];
r1_t400_d4_cneg01=[NumVL{1,8,3,5}];
r1_t450_d4_cneg01=[NumVL{1,9,3,5}];
r1_t500_d4_cneg01=[NumVL{1,10,3,5}];

Increasing_TNSPEC3neg=[r1_t50_d4_cneg01(10,:);r1_t100_d4_cneg01(10,:);...
    r1_t150_d4_cneg01(10,:);r1_t200_d4_cneg01(10,:);r1_t250_d4_cneg01(10,:);...
    r1_t300_d4_cneg01(10,:);r1_t350_d4_cneg01(10,:);r1_t400_d4_cneg01(10,:);...
    r1_t450_d4_cneg01(10,:);r1_t500_d4_cneg01(10,:)]

figure('Name','1,1,3 npa')
plot (Increasing_TNSPEC3neg(:,7),(Increasing_TNSPEC3(:,2).*100));
hold on
plot (Increasing_TNSPEC_neg(:,7),(Increasing_TNSPEC_neg(:,2).*100));
title({'Average Percent Error as Training Set Size Increases';...
    'Dimension=4 Radfrac=0.5 Bootstrap Replicates=500'})
xlabel('Training Set Size')
ylabel('Percent Error')
legend('CP=(0,1)','CP=(0,-1)')
hold off

%Make a graph for 11 dimensions
r1_t50_d11_c01=[NumVL{1,1,10,1}];
r1_t100_d11_c01=[NumVL{1,2,10,1}];
r1_t150_d11_c01=[NumVL{1,3,10,1}];
r1_t200_d11_c01=[NumVL{1,4,10,1}];
r1_t250_d11_c01=[NumVL{1,5,10,1}];
r1_t300_d11_c01=[NumVL{1,6,10,1}];
r1_t350_d11_c01=[NumVL{1,7,10,1}];
r1_t400_d11_c01=[NumVL{1,8,10,1}];
r1_t450_d11_c01=[NumVL{1,9,10,1}];
r1_t500_d11_c01=[NumVL{1,10,10,1}];

Increasing_TNSPEC11=[r1_t50_d11_c01(10,:);r1_t100_d11_c01(10,:);...
    r1_t150_d11_c01(10,:);r1_t200_d11_c01(10,:);r1_t250_d11_c01(10,:);...
    r1_t300_d11_c01(10,:);r1_t350_d11_c01(10,:);r1_t400_d11_c01(10,:);...

```

```

r1_t450_d11_c01(10,:);r1_t500_d11_c01(10,:)]

r1_t50_d11_cneg01=[NumVL{1,1,10,5}];
r1_t100_d11_cneg01=[NumVL{1,2,10,5}];
r1_t150_d11_cneg01=[NumVL{1,3,10,5}];
r1_t200_d11_cneg01=[NumVL{1,4,10,5}];
r1_t250_d11_cneg01=[NumVL{1,5,10,5}];
r1_t300_d11_cneg01=[NumVL{1,6,10,5}];
r1_t350_d11_cneg01=[NumVL{1,7,10,5}];
r1_t400_d11_cneg01=[NumVL{1,8,10,5}];
r1_t450_d11_cneg01=[NumVL{1,9,10,5}];
r1_t500_d11_cneg01=[NumVL{1,10,10,5}];

Increasing_TNSPEC11neg=[r1_t50_d11_cneg01(10,:);r1_t100_d11_cneg01(10,:);...
    r1_t150_d11_cneg01(10,:);r1_t200_d11_cneg01(10,:);r1_t250_d11_cneg01(10,:);...
    r1_t300_d11_cneg01(10,:);r1_t350_d11_cneg01(10,:);r1_t400_d11_cneg01(10,:);...
    r1_t450_d11_cneg01(10,:);r1_t500_d11_cneg01(10,:)]

figure('Name','1,1,10 npa')
plot (Increasing_TNSPEC11(:,7),(Increasing_TNSPEC11(:,2).*100));
hold on
plot (Increasing_TNSPEC11neg(:,7),(Increasing_TNSPEC11neg(:,2).*100));
title({'Average Percent Error as Training Set Size Increases';'Dimension=11 Radfrac=0.5
Bootstrap Replicates=500'})
xlabel('Training Set Size')
ylabel('Percent Error')
legend('CP=(0,1)','CP=(0,-1)')
hold off

```

RSD_as_Bootstrap_Radfrac_Increase

```
%Plotting Relative Standard Deviation
%as Bootstrapping and Radfrac Increase

%This code plots bootstrapping vs. RSD as radfrac increases

%Indexing in NumVL Cell Matrix is as Follows:
%Indexing is {Radfrac Iteration Count (must separately count...
%Radfrac value), Training Set Size, dimension, Compass Point,}
%RSD_NumVL{radfrac_iteration_count,u,(v+1),j}=RSD;

%Within matrix:
%Store Relative Standard Deviation (average standard deviation/
%average)
%RSD holds [1.compass point coordinate, 2.compass point coordinate,...
%3.B, 4.RSD of returned standard deviation, ...
%5.total number of dimensions (v+2), 6.radfrac,
%7.Training Set Size]

figure('Name','1,1,1 npa')
r1_t50_d2_c01=[RSD_NumVL{1,1,1,1}];
plot(r1_t50_d2_c01(:,3),r1_t50_d2_c01(:,4));
hold on
r1_t50_d2_cneg01=[RSD_NumVL{1,1,1,5}];
plot(r1_t50_d2_cneg01(:,3),r1_t50_d2_cneg01(:,4));
title({'Relative Standard Deviation as Bootstrap Replicates Increase';Radfrac=0.5
Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Relative Standard Deviation')
legend('CP=(0,1)','CP=(0,-1)')
%axis(graphlimits)
hold off

figure('Name','2,1,1 npa')
r2_t50_d2_c01=[RSD_NumVL{2,1,1,1}];
plot(r2_t50_d2_c01(:,3),r2_t50_d2_c01(:,4));
hold on
r2_t50_d2_cneg01=[RSD_NumVL{2,1,1,5}];
```

```

plot (r2_t50_d2_cneg01(:,3),r2_t50_d2_cneg01(:,4));
title({'Relative Standard Deviation as Bootstrap Replicates Increase';Radfrac=0.6
Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Relative Standard Deviation')
legend('CP=(0,1)','CP=(0,-1)')
hold off

```

```

figure('Name','3,1,1 npa')
r3_t50_d2_c01=[RSD_NumVL{3,1,1,1}];
plot (r3_t50_d2_c01(:,3),r3_t50_d2_c01(:,4));

```

```

hold on
r3_t50_d2_cneg01=[RSD_NumVL{3,1,1,5}];
plot (r3_t50_d2_cneg01(:,3),r3_t50_d2_cneg01(:,4));
title({'Relative Standard Deviation as Bootstrap Replicates Increase';Radfrac=0.7
Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Relative Standard Deviation')
legend('CP=(0,1)','CP=(0,-1)')
hold off

```

```

figure('Name','4,1,1 npa')
r4_t50_d2_c01=[RSD_NumVL{4,1,1,1}];
plot (r4_t50_d2_c01(:,3),r4_t50_d2_c01(:,4));
hold on
r4_t50_d2_cneg01=[RSD_NumVL{4,1,1,5}];
plot (r4_t50_d2_cneg01(:,3),r4_t50_d2_cneg01(:,4));
title({'Relative Standard Deviation as Bootstrap Replicates Increase';...
'Radfrac=0.8 Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Relative Standard Deviation')
legend('CP=(0,1)','CP=(0,-1)')
hold off

```

```

figure('Name','5,1,1 npa')
r5_t50_d2_c01=[RSD_NumVL{5,1,1,1}];
plot (r5_t50_d2_c01(:,3),r5_t50_d2_c01(:,4));
hold on
r5_t50_d2_cneg01=[RSD_NumVL{5,1,1,5}];
plot (r5_t50_d2_cneg01(:,3),r5_t50_d2_cneg01(:,4));
title({'Relative Standard Deviation as Bootstrap Replicates Increase';...
'Radfrac=0.9 Dimension=2 TNSPEC=50'})

```



```
xlabel('Bootstrap Replications')
ylabel('Relative Standard Deviation')
legend('CP=(0,1)', '(0,-1)')
hold off
```

```
figure('Name','6,1,1 npa')
r6_t50_d2_c01=[RSD_NumVL{6,1,1,1}];
plot (r6_t50_d2_c01(:,3),r6_t50_d2_c01(:,4));
hold on
r6_t50_d2_cneg01=[RSD_NumVL{6,1,1,5}];
plot (r6_t50_d2_cneg01(:,3),r6_t50_d2_cneg01(:,4));
title({'Relative Standard Deviation as Bootstrap Replicates Increase';...
      'Radfrac=1.0 Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replications')
ylabel('Relative Standard Deviation')
legend('CP=(0,1)', 'CP=(0,-1)')
hold off
```

RSD_as_TNSPEC_Dimension_Increase

%Plotting Relative Standard Deviation
%as TNSPEC and Dimension Increase

%Indexing in NumVL Cell Matrix is as Follows:
%Indexing is {Radfrac Iteration Count (must separately count...
%Radfrac value), Training Set Size, dimension, Compass Point,}
%RSD_NumVL{radfrac_iteration_count,u,(v+1),j}=RSD;

%Within matrix:
%Store Relative Standard Deviation (average standard deviation/
%average)
%RSD holds [1.compass point coordinate, 2.compass point coordinate,...
%3.B, 4.RSD of returned standard deviation, ...
%5.total number of dimensions (v+2), 6.radfrac,
%7.Training Set Size]

%Set Axis Limits for Graphs
%axis(limits) syntax is axis(xmin xmax ymin ymax)
graphlimits=[50,500,0,0.1]

%Get all for (0,1) Compass Points
r1_t50_d2_c01=[RSD_NumVL{1,1,1,1}];
r1_t100_d2_c01=[RSD_NumVL{1,2,1,1}];
r1_t150_d2_c01=[RSD_NumVL{1,3,1,1}];
r1_t200_d2_c01=[RSD_NumVL{1,4,1,1}];
r1_t250_d2_c01=[RSD_NumVL{1,5,1,1}];
r1_t300_d2_c01=[RSD_NumVL{1,6,1,1}];
r1_t350_d2_c01=[RSD_NumVL{1,7,1,1}];
r1_t400_d2_c01=[RSD_NumVL{1,8,1,1}];
r1_t450_d2_c01=[RSD_NumVL{1,9,1,1}];
r1_t500_d2_c01=[RSD_NumVL{1,10,1,1}];

Increasing_TNSPEC=[r1_t50_d2_c01(10,:);r1_t100_d2_c01(10,:);...
r1_t150_d2_c01(10,:);r1_t200_d2_c01(10,:);r1_t250_d2_c01(10,:);...
r1_t300_d2_c01(10,:);r1_t350_d2_c01(10,:);r1_t400_d2_c01(10,:);...
r1_t450_d2_c01(10,:);r1_t500_d2_c01(10,:);

%Get all (0,-1) Compass Points
r1_t50_d2_cneg01=[RSD_NumVL{1,1,1,5}];

```

r1_t100_d2_cneg01=[RSD_NumVL{1,2,1,5}];
r1_t150_d2_cneg01=[RSD_NumVL{1,3,1,5}];
r1_t200_d2_cneg01=[RSD_NumVL{1,4,1,5}];
r1_t250_d2_cneg01=[RSD_NumVL{1,5,1,5}];
r1_t300_d2_cneg01=[RSD_NumVL{1,6,1,5}];
r1_t350_d2_cneg01=[RSD_NumVL{1,7,1,5}];
r1_t400_d2_cneg01=[RSD_NumVL{1,8,1,5}];
r1_t450_d2_cneg01=[RSD_NumVL{1,9,1,5}];
r1_t500_d2_cneg01=[RSD_NumVL{1,10,1,5}];

Increasing_TNSPEC_neg=[r1_t50_d2_cneg01(10,:);r1_t100_d2_cneg01(10,:);...
    r1_t150_d2_cneg01(10,:);r1_t200_d2_cneg01(10,:);r1_t250_d2_cneg01(10,:);...
    r1_t300_d2_cneg01(10,:);r1_t350_d2_cneg01(10,:);r1_t400_d2_cneg01(10,:);...
    r1_t450_d2_cneg01(10,:);r1_t500_d2_cneg01(10,:)];

figure('Name','1,1,1 npa')
plot (Increasing_TNSPEC(:,7),Increasing_TNSPEC(:,4));

hold on
plot (Increasing_TNSPEC_neg(:,7),Increasing_TNSPEC_neg(:,4));
title({'Relative Standard Deviation as Training Set Size Increases','Dimension=2
Radfrac=0.5 Bootstrap Replicates=500'})
xlabel('Training Set Size')
ylabel('Absolute Error')
legend('CP=(0,1)','CP=(0,-1)')
axis(graphlimits)
hold off

```

```

%Make second graph for dimension=3

```

```

r1_t50_d3_c01=[RSD_NumVL{1,1,2,1}];
r1_t100_d3_c01=[RSD_NumVL{1,2,2,1}];
r1_t150_d3_c01=[RSD_NumVL{1,3,2,1}];
r1_t200_d3_c01=[RSD_NumVL{1,4,2,1}];
r1_t250_d3_c01=[RSD_NumVL{1,5,2,1}];
r1_t300_d3_c01=[RSD_NumVL{1,6,2,1}];
r1_t350_d3_c01=[RSD_NumVL{1,7,2,1}];
r1_t400_d3_c01=[RSD_NumVL{1,8,2,1}];
r1_t450_d3_c01=[RSD_NumVL{1,9,2,1}];
r1_t500_d3_c01=[RSD_NumVL{1,10,2,1}];

Increasing_TNSPEC2=[r1_t50_d3_c01(10,:);r1_t100_d3_c01(10,:);...

```

```

r1_t150_d3_c01(10,:);r1_t200_d3_c01(10,:);r1_t250_d3_c01(10,:);...
r1_t300_d3_c01(10,:);r1_t350_d3_c01(10,:);r1_t400_d3_c01(10,:);...
r1_t450_d3_c01(10,:);r1_t500_d3_c01(10,:);

r1_t50_d3_cneg01=[RSD_NumVL{1,1,2,5}];
r1_t100_d3_cneg01=[RSD_NumVL{1,2,2,5}];
r1_t150_d3_cneg01=[RSD_NumVL{1,3,2,5}];
r1_t200_d3_cneg01=[RSD_NumVL{1,4,2,5}];
r1_t250_d3_cneg01=[RSD_NumVL{1,5,2,5}];
r1_t300_d3_cneg01=[RSD_NumVL{1,6,2,5}];
r1_t350_d3_cneg01=[RSD_NumVL{1,7,2,5}];
r1_t400_d3_cneg01=[RSD_NumVL{1,8,2,5}];
r1_t450_d3_cneg01=[RSD_NumVL{1,9,2,5}];
r1_t500_d3_cneg01=[RSD_NumVL{1,10,2,5}];

Increasing_TNSPEC2neg=[r1_t50_d3_cneg01(10,:);r1_t100_d3_cneg01(10,:);...
    r1_t150_d3_cneg01(10,:);r1_t200_d3_cneg01(10,:);r1_t250_d3_cneg01(10,:);...
    r1_t300_d3_cneg01(10,:);r1_t350_d3_cneg01(10,:);r1_t400_d3_cneg01(10,:);...
    r1_t450_d3_cneg01(10,:);r1_t500_d3_cneg01(10,:);

figure('Name','1,1,2 npa')
plot (Increasing_TNSPEC2(:,7),Increasing_TNSPEC2(:,4));
hold on
plot (Increasing_TNSPEC2neg(:,7),Increasing_TNSPEC2neg(:,4));
title({'Relative Standard Deviation as Training Set Size Increases';...
    'Dimension=3 Radfrac=0.5 Bootstrap Replicates=500'})
xlabel('Training Set Size')
ylabel('Absolute Error')
legend('CP=(0,1)','CP=(0,-1)')
axis(graphlimits)
hold off

%Make a third graph for 4 dimensions
r1_t50_d4_c01=[RSD_NumVL{1,1,3,1}];
r1_t100_d4_c01=[RSD_NumVL{1,2,3,1}];
r1_t150_d4_c01=[RSD_NumVL{1,3,3,1}];
r1_t200_d4_c01=[RSD_NumVL{1,4,3,1}];
r1_t250_d4_c01=[RSD_NumVL{1,5,3,1}];
r1_t300_d4_c01=[RSD_NumVL{1,6,3,1}];
r1_t350_d4_c01=[RSD_NumVL{1,7,3,1}];
r1_t400_d4_c01=[RSD_NumVL{1,8,3,1}];
r1_t450_d4_c01=[RSD_NumVL{1,9,3,1}];

```

```

r1_t500_d4_c01=[RSD_NumVL{1,10,3,1}];

Increasing_TNSPEC3=[r1_t50_d4_c01(10,:);r1_t100_d4_c01(10,:);...
    r1_t150_d4_c01(10,:);r1_t200_d4_c01(10,:);r1_t250_d4_c01(10,:);...
    r1_t300_d4_c01(10,:);r1_t350_d4_c01(10,:);r1_t400_d4_c01(10,:);...
    r1_t450_d4_c01(10,:);r1_t500_d4_c01(10,:)];

r1_t50_d4_cneg01=[RSD_NumVL{1,1,3,5}];
r1_t100_d4_cneg01=[RSD_NumVL{1,2,3,5}];
r1_t150_d4_cneg01=[RSD_NumVL{1,3,3,5}];
r1_t200_d4_cneg01=[RSD_NumVL{1,4,3,5}];
r1_t250_d4_cneg01=[RSD_NumVL{1,5,3,5}];
r1_t300_d4_cneg01=[RSD_NumVL{1,6,3,5}];
r1_t350_d4_cneg01=[RSD_NumVL{1,7,3,5}];
r1_t400_d4_cneg01=[RSD_NumVL{1,8,3,5}];
r1_t450_d4_cneg01=[RSD_NumVL{1,9,3,5}];
r1_t500_d4_cneg01=[RSD_NumVL{1,10,3,5}];

Increasing_TNSPEC3neg=[r1_t50_d4_cneg01(10,:);r1_t100_d4_cneg01(10,:);...
    r1_t150_d4_cneg01(10,:);r1_t200_d4_cneg01(10,:);r1_t250_d4_cneg01(10,:);...
    r1_t300_d4_cneg01(10,:);r1_t350_d4_cneg01(10,:);r1_t400_d4_cneg01(10,:);...
    r1_t450_d4_cneg01(10,:);r1_t500_d4_cneg01(10,:)];

figure('Name','1,1,3 npa')
plot (Increasing_TNSPEC3neg(:,7),Increasing_TNSPEC3(:,4));
hold on
plot (Increasing_TNSPEC_neg(:,7),Increasing_TNSPEC_neg(:,4));
title({'Relative Standard Deviation as Training Set Size Increases';...
    'Dimension=4 Radfrac=0.5 Bootstrap Replicates=500'})
xlabel('Training Set Size')
ylabel('Absolute Error')
legend('CP=(0,1)', 'CP=(0,-1)')
axis(graphlimits)
hold off

%Make a graph for 11 dimensions
r1_t50_d11_c01=[RSD_NumVL{1,1,10,1}];
r1_t100_d11_c01=[RSD_NumVL{1,2,10,1}];
r1_t150_d11_c01=[RSD_NumVL{1,3,10,1}];
r1_t200_d11_c01=[RSD_NumVL{1,4,10,1}];
r1_t250_d11_c01=[RSD_NumVL{1,5,10,1}];
r1_t300_d11_c01=[RSD_NumVL{1,6,10,1}];

```

```

r1_t350_d11_c01=[RSD_NumVL{1,7,10,1}];
r1_t400_d11_c01=[RSD_NumVL{1,8,10,1}];
r1_t450_d11_c01=[RSD_NumVL{1,9,10,1}];
r1_t500_d11_c01=[RSD_NumVL{1,10,10,1}];

```

```

Increasing_TNSPEC11=[r1_t50_d11_c01(10,:);r1_t100_d11_c01(10,:);...
    r1_t150_d11_c01(10,:);r1_t200_d11_c01(10,:);r1_t250_d11_c01(10,:);...
    r1_t300_d11_c01(10,:);r1_t350_d11_c01(10,:);r1_t400_d11_c01(10,:);...
    r1_t450_d11_c01(10,:);r1_t500_d11_c01(10,:);];

```

```

r1_t50_d11_cneg01=[RSD_NumVL{1,1,10,5}];
r1_t100_d11_cneg01=[RSD_NumVL{1,2,10,5}];
r1_t150_d11_cneg01=[RSD_NumVL{1,3,10,5}];
r1_t200_d11_cneg01=[RSD_NumVL{1,4,10,5}];
r1_t250_d11_cneg01=[RSD_NumVL{1,5,10,5}];
r1_t300_d11_cneg01=[RSD_NumVL{1,6,10,5}];
r1_t350_d11_cneg01=[RSD_NumVL{1,7,10,5}];
r1_t400_d11_cneg01=[RSD_NumVL{1,8,10,5}];
r1_t450_d11_cneg01=[RSD_NumVL{1,9,10,5}];
r1_t500_d11_cneg01=[RSD_NumVL{1,10,10,5}];

```

```

Increasing_TNSPEC11neg=[r1_t50_d11_cneg01(10,:);r1_t100_d11_cneg01(10,:);...
    r1_t150_d11_cneg01(10,:);r1_t200_d11_cneg01(10,:);r1_t250_d11_cneg01(10,:);...
    r1_t300_d11_cneg01(10,:);r1_t350_d11_cneg01(10,:);r1_t400_d11_cneg01(10,:);...
    r1_t450_d11_cneg01(10,:);r1_t500_d11_cneg01(10,:);];

```

```

figure('Name','1,1,10 npa')
plot(Increasing_TNSPEC11(:,7),Increasing_TNSPEC11(:,4));
hold on
plot(Increasing_TNSPEC11neg(:,7),Increasing_TNSPEC11neg(:,4));
title({'Relative Standard Deviation as Training Set Size Increases';'Dimension=11
Radfrac=0.5 Bootstrap Replicates=500'})
xlabel('Training Set Size')
ylabel('Absolute Error')
legend('CP=(0,1)','CP=(0,-1)')
axis(graphlimits)
hold off

```

Figures_Code

```
figure
mu=[0,0];
sigma=[2,1];
TNSPEC=mvnrnd(mu,sigma,100);
scatter(TNSPEC(:,1), TNSPEC(:,2))
hold on
%equation of an ellipse

%Graph Compass Points
compass_points=[0,1;(-2/(sqrt(5))),(2/(sqrt(5))); -2,0;(-2/(sqrt(5))),(2/(sqrt(5)));
0,-1;(2/(sqrt(5))),(2/(sqrt(5))); 2,0; (2/(sqrt(5))),(2/(sqrt(5)))];
scatter(compass_points(:,1),compass_points(:,2),50,'r','filled')

%Graph Center
scatter(0,0,50,'r','filled')

%Graph Ellipse
a=2; % horizontal radius
b=1; % vertical radius
x0=0; % x0,y0 ellipse centre coordinates
y0=0;
t=-pi:0.01:pi;
x=x0+a*cos(t);
y=y0+b*sin(t);
plot(x,y)

hold off

figure
mu=[0,0];
sigma=[1,1];
TNSPEC=mvnrnd(mu,sigma,100);
scatter(TNSPEC(:,1), TNSPEC(:,2))
hold on

%Graph center
scatter(0,0,'filled')
%Graph circle for 1 standard deviation
centers=[0,0];
```

```
radii=[1];  
viscircles(centers,radii)  
  
%Graph second standard deviation  
radii=[2];  
viscircles(centers,radii,'Color','b')  
hold off
```


Graphing_Code_2

%Requires workspace variables NumVL and RSD_NumVL, Max_training,

%Make sure these exist.

NumVL;

RSD_NumVL;

out=[cat(1,NumVL{:})];

x=size(out);

w=x(:,1);

y=zeros(x(:,1),1);

%Designate compass points - work in progress

newmatrix=[out,y];

%nonprincipal axis points

npa=newmatrix(:,4)==0;

%principal axis points

pa=newmatrix(:,4)==2;

opa=newmatrix(:,4)==-2;

%since diagonals have a calculation that determines the value

%it is hard to tell how many digits of precision

%therefore, diagonals are not equal to 0, 2, or -2.

%First find indices $\sim=0$

dp=newmatrix(:,4) $\sim=0$;

dp2=newmatrix(dp,:);

%Now find indices in the $\sim=0$ subset also $\sim= 2$

fdp=dp2(:,4) $\sim=2$;

dp3=dp2(fdp,:);

%Last, find indices in that subset also $\sim=$ to -2

ffdp=dp3(:,4) $\sim=-2$;

nonprincipalaxis=[newmatrix(npa,:)];

principalaxis=[newmatrix(pa,:);newmatrix(opa,:)];

diagonalpoints=[dp3(ffdp,:)];

%Regress Non-Principal

%B is located in Bias_plot(:,1)

%bias is located in Bias_plot(:,2)

```

%dimension is located in Bias_plot(:,3)
%First 2 coordinates of compass point located in Bias_plot(:,4) and (:,5)
%radfrac located in Bias_plot(:,6)
%Training set size located in Bias_plot(:,7)
%Columns are
%(bootstrap,bias,dimension,compass,compass,radfrac,training)
npa_output=[nonprincipalaxis(:,2)];

npa_input=[nonprincipalaxis(:,1),nonprincipalaxis(:,3),nonprincipalaxis(:,6),nonprincipalaxis(:,7)];
reg = MultiPolyRegress(npa_input,npa_output,5)

%Regress Principal
%B is located in Bias_plot(:,1)
%bias is located in Bias_plot(:,2)
%dimension is located in Bias_plot(:,3)
%First 2 coordinates of compass point located in Bias_plot(:,4) and (:,5)
%radfrac located in Bias_plot(:,6)
%Training set size located in Bias_plot(:,7)
%Columns are
%(bootstrap,bias,dimension,compass,compass,radfrac,training)
pa_output=[principalaxis(:,2)];
pa_input=[principalaxis(:,1),principalaxis(:,3),principalaxis(:,6),principalaxis(:,7)];
reg = MultiPolyRegress(pa_input,pa_output,5)

%Regress Diagonal
%B is located in Bias_plot(:,1)
%bias is located in Bias_plot(:,2)
%dimension is located in Bias_plot(:,3)
%First 2 coordinates of compass point located in Bias_plot(:,4) and (:,5)
%radfrac located in Bias_plot(:,6)
%Training set size located in Bias_plot(:,7)
%Columns are
%(bootstrap,bias,dimension,compass,compass,radfrac,training)
d_output=[diagonalpoints(:,2)];

d_input=[diagonalpoints(:,1),diagonalpoints(:,3),diagonalpoints(:,6),diagonalpoints(:,7)];
reg = MultiPolyRegress(d_input,d_output,5)

```

RTCR_Point_Bias2

```
%RTCR_Point_Bias2.m
```

```
%This program calculates the average percent error of RTCR outputs and graphs  
%the average percent error for each compass point at increasing numbers of  
%bootstrap replications.
```

```
%TNSPEC=mvrnd(mu,sigma,50*u);compass_points=[0,1;...  
%(-2/(sqrt(5))),(2/(sqrt(5)));...  
%-2,0;...  
%(-2/(sqrt(5))),(2/(sqrt(5)));...  
% 0,-1;...  
%(2/(sqrt(5))),(2/(sqrt(5)));...  
%2,0;...  
%(2/(sqrt(5))),(2/(sqrt(5)))];
```

```
A=NumVL{1,1,1,1}  
point_0_1=A;  
point_0_1(:,2)=( point_0_1(:,2)/(expected_standard_deviation(1))*100)
```

```
figure  
plot(point_0_1(:,1),point_0_1(:,2),'-r')  
hold on
```

```
A=NumVL{1,1,1,2}  
point2=A;  
point2(:,2)=( point2(:,2)/(expected_standard_deviation(1))*100)  
plot(point2(:,1),point2(:,2),'-b')  
hold on
```

```
A=NumVL{1,1,1,3};  
point3=A;  
point3(:,2)=( point3(:,2)/(expected_standard_deviation(1))*100)  
plot(point3(:,1),point3(:,2),'-k')  
hold on
```

```
A=NumVL{1,1,1,4}  
point4=A;  
point4(:,2)=( point4(:,2)/(expected_standard_deviation(1))*100)  
plot(point4(:,1),point4(:,2),'-ob')  
hold on
```

```

A=NumVL{1,1,1,5}
point5=A;
point5(:,2)=(point5(:,2))/(expected_standard_deviation(1))*100
plot(point5(:,1),point5(:,2),'-or')
hold on

A=NumVL{1,1,1,6}
point6=A;
point6(:,2)=(point6(:,2))/(expected_standard_deviation(1))*100
plot(point6(:,1),point6(:,2),'-sb')
hold on

A=NumVL{1,1,1,7}
point7=A;
point7(:,2)=(point7(:,2))/(expected_standard_deviation(1))*100
plot(point7(:,1),point7(:,2),'-ok')
hold on

A=NumVL{1,1,1,8}
point8=A;
point8(:,2)=(point8(:,2))/(expected_standard_deviation(1))*100
plot(point8(:,1),point8(:,2),'-pb')
hold off
legend('(0,1)', '(-2/(sqrt(5))), (2/(sqrt(5)))', '(-2,0)', ...,
        '(-2/(sqrt(5))), (-2/(sqrt(5)))', '(0,-1)', '(2/(sqrt(5))), (-2/(sqrt(5)))', ...,
        '(2,0)', '(2/(sqrt(5))), (2/(sqrt(5)))', 'Location', 'southoutside')
title({'Average Percent Error as Bootstrap Replicates Increase'; Radfrac=0.5
Dimension=2 TNSPEC=50'})
xlabel('Bootstrap Replicates')
ylabel('Average Percent Error of Six Runs')

```

REFERENCES:

Chapter 1 References:

1. Drennen, James K. et al. "Nondestructive Near-Infrared Analysis of Intact Tablets for Determination of Degradation Products." *Journal of Pharmaceutical Sciences* , Volume 79 , Issue 7 , 622 - 627
2. Heise, H. (2009). Donald A. Burns, Emil W. Ciurczak (Eds.): *Handbook of near-infrared analysis*, 3rd ed. *Analytical and Bioanalytical Chemistry*, 393(5), 314.
3. Lodder, RA. "Can You Find Intelligent Communications in Ultrahigh -Dimensional Big Data from Near-Infrared Optical SETI?" SoCIA 2018 (Reno, NV) April 2018
4. Lodder, R. A.; Hieftje, G. M. Detection of Capsule Tampering by Near-Infrared Reflectance Analysis. *Analytical Chemistry* 1987, 59 (15), 1921-1930.
5. Lodder, R. A.; Hieftje, G. M. Detection of Subpopulations in Near-Infrared Reflectance Analysis. *Applied Spectroscopy* 1988, 42 (8), 1500-1512.
6. Lodder, R. A. Method for detecting subpopulations in spectral analysis. 6-9-1992. Google Patents.
7. Robert A. Lodder and Gary M. Hieftje, "Quantile Analysis: A Method for Characterizing Data Distributions," *Appl. Spectrosc.* 42, 1512-1520 (1988)
8. Lodder, R. A.; Hieftje, G. M. "Quantile BEAST Attacks the False-Sample Problem in Near Infrared Reflectance Analysis." *Applied Spectroscopy* 1988, 42 (8), 1353-1365.
9. Lodder, R., & Hieftje, Gary M. (1988). "Solving the False-sample Problem in Near-infrared Reflectance Analysis." *ProQuest Dissertations and Theses*.
10. Margarida G.M.S. Cardoso, Isabel H. Themido, Fernando Moura Pires. "Evaluating a clustering solution: An application in the tourism market." *Intelligent Data Analysis*, Volume 3, Issue 6, 1999, Pages 491-510, ISSN 1088-467X
11. "What is Discriminant Analysis?" *Research Optimus*. Access Date: 9/20/2018 <<https://www.researchoptimus.com/article/what-is-discriminant-analysis.php>>

Chapter 2 References:

1. Cherian, S. S.; Walimbe, A. M.; Jadhav, S. M.; Gandhe, S. S.; Hundekar, S. L.; Mishra, A. C.; Arankalle, V. A. Evolutionary rates and timescale comparison of Chikungunya viruses inferred from the whole genome/E1 gene with special reference to the 2005 outbreak in the Indian subcontinent. 2009, 9 (1), 16-23.
2. Cochrane Handbook for Systematic Reviews of Interventions. Higgins JPT and Green, S. [Version 5.1.0]. 2011. The Cochrane Collaboration.
3. Davis, J. A.; Gift, J. S.; Zhao, Q. J. Introduction to benchmark dose methods and U.S. EPA's benchmark dose software (BMDS) version 2.1.1. Toxicology and Applied Pharmacology 2011, (254(2)), 181-191.
4. Esdaile, D. J. Principles, benefits and limitations of the NOAEL approach. Toxicology Letters 1994, 74 (Supplement 1), 23.
5. Espín, J.; Larrosa, M.; García-Conesa; Tomás-Barberán, F. A. Biological Significance of Urolithins, the Gut Microbial Ellagic Acid-Derived Metabolites: The Evidence So Far. Evidence-Based Complementary and Alternative Medicine 2013, 2013.
6. Kaur, P. e. a. Inhibition of Chikungunya Virus Replication by Harringtonine, a Novel Antiviral that Suppresses Viral Protein Expression. Antimicrobial Agents and Chemotherapy 2013, 57 (1), 155-167.
7. Lodder, R. A. Method for detecting subpopulations in spectral analysis. 6-9-1992. Google Patents.
8. Lodder, R. A.; Hieftje, G. M. Detection of Capsule Tampering by Near-Infrared Reflectance Analysis. Analytical Chemistry 1987, 59 (15), 1921-1930.
9. Lodder, R. A.; Hieftje, G. M. Detection of Subpopulations in Near-Infrared Reflectance Analysis. Applied Spectroscopy 1988, 42 (8), 1500-1512.
10. Lodder, R. A.; Hieftje, G. M. Quantile BEAST Attacks the False-Sample Problem in Near Infrared Reflectance Analysis. Applied Spectroscopy 1988, 42 (8), 1353-1365.
11. NIH U.S.National Library of Medicine: Tox Tutor . NOAEL and LOAEL. 2017. 7-17-2017.
12. Seppa, N.; Hirshfeld, J. Chikungunya is on the move. Science News . 6-25-2015. The Society for Science & the Public. 7-27-2017.
13. Shamieh, C. The V-model for the systems engineering process. In Systems Engineering for Dummies, Wiley Publishing, Inc.: Hoboken, NJ, 2011.
14. Tomás-Barberán, F. A.; García-Villalba, R.; González-Sarrías, A.; Selma, M. V.; Espín, J.; . Ellagic Acid Metabolism by Human Gut Microbiota: Consistent Observation of Three Urolithin Phenotypes in Intervention Trials, Independent of Food Source, Age, and Health Status. J. Agric. Food Chem. 2014, 62 (28), 6535-6538.
15. US Center for Disease Control and Prevention . Chikungunya in South America. 9-7-2016. 7-27-2017.
16. US Center for Disease Control and Prevention . Chikungunya Nowcast for the Americas. 7-27-2017. 7-27-2017.

17. US Center for Disease Control and Prevention . Chikungunya Virus: Geographic Distribution. 4-22-2016. 7-27-2017.
18. US Center for Disease Control and Prevention . Travel Notice: Chikungunya in the Caribbean. 9-17-2015. 7-27-2017.
19. World Health Organization Regional Office for South-East Asia . Surveillance and Outbreak Alert: Chikungunya. 2017. 7-27-2017.

Published Version of Chapter 2

20. Dickerson, Cynthia; Lodder, Robert A. "A Novel Statistical Approach to NOAEL: QBEST Applied to Dosing of Ellagic Acid." WebmedCentral TOXICOLOGY 2018,WMC005451.

Chapter 3 References:

1. Park, S., Kang, Y.: Dietary ellagic acid suppresses atherosclerotic lesion formation and vascular inflammation in apoE-deficient mice. *FASEB J.* 27(1), 861-23 (2013)
2. García-Niño, R.W., Zazueta, C.: Ellagic acid: pharmacological activities and molecular mechanisms involved in liver protection. *Pharmacol. Res.* 97, 84–103 (2015)
3. Chikungunya-Wikipedia, <https://en.wikipedia.org/wiki/Chikungunya>
4. Kaur, P., Thiruchelvan, M., Lee, R.C.H., Chen, H., Chen, K.C., Ng, M.L., Chu, J.J.H.: Inhibition of chikungunya virus replication by harringtonine, a novel antiviral that suppresses viral protein expression. *Antimicrob. Agents Chemother.* 57(1), 155–167 (2013)
5. Cerdá, B., et al.: Identification of urolithin A as a metabolite produced by human colon microflora from ellagic acid and related compounds. *J. Agric. Food Chem.* 53(14), 5571–5576 (2005)
6. Cerdá, B., et al.: The potent in vitro antioxidant ellagitannins from pomegranate juice are metabolised into bioavailable but poor antioxidant hydroxy-6H-dibenzopyran-6-one derivatives by the colonic microflora of healthy humans. *Eur. J. Nutr.* 43(4), 205–220 (2004)
7. Cerdá, B., Tomás-Barberán, F.A., Espín, J.C.: Metabolism of antioxidant and chemopreventive ellagitannins from strawberries, raspberries, walnuts, and oak-aged wine in humans: identification of biomarkers and individual variability. *J. Agric. Food Chem.* 53(2), 227–235 (2005)
8. Espín, J.C., et al.: Iberian pig as a model to clarify obscure points in the bioavailability and metabolism of ellagitannins in humans. *J. Agric. Food Chem.* 55(25), 10476–10485 (2007)
9. Mertens-Talcott, S.U., et al.: Absorption, metabolism, and antioxidant effects of pomegranate (*Punica granatum L.*) polyphenols after ingestion of a standardized extract in healthy human volunteers. *J. Agric. Food Chem.* 54(23), 8956–8961 (2006)
10. Seeram, N.P., Lee, R., Heber, D.: Bioavailability of ellagic acid in human plasma after

- consumption of ellagitannins from pomegranate (*Punica granatum* L.) juice. *Clin. Chim. Acta* 348(1), 63–68 (2004)
11. Seeram, N.P., et al.: Pomegranate juice ellagitannin metabolites are present in human plasma and some persist in urine for up to 48 hours. *J. Nutr.* 136(10), 2481–2485 (2006)
 12. Tomás-Barberán, F.A., Espín, J.C., García-Conesa, M.T.: Bioavailability and metabolism of ellagic acid and ellagitannins. *Chem. Biol. Ellagitannins* 7, 293–297 (2009)
 13. González-Barrio, R., et al.: UV and MS identification of urolithins and nasutins, the bioavailable metabolites of ellagitannins and ellagic acid in different mammals. *J. Agric. Food Chem.* 59(4), 1152–1162 (2011)
 14. Espín, J.C., et al.: Biological significance of urolithins, the gut microbial ellagic acid-derived metabolites: the evidence so far. *Evid. Based Complement. Altern. Med.* 2013, 1–15 (2013)
 15. Larrosa, M., et al.: Anti-inflammatory properties of a pomegranate extract and its metabolite urolithin-A in a colitis rat model and the effect of colon inflammation on phenolic metabolism. *J. Nutr. Biochem.* 21(8), 717–725 (2010)
 16. Ishimoto, H., et al.: In vivo anti-inflammatory and antioxidant properties of ellagitannin metabolite urolithin A. *Bioorg. Med. Chem. Lett.* 21(19), 5901–5904 (2011)
 17. Piwowarski, J.P., et al.: Role of human gut microbiota metabolism in the anti-inflammatory effect of traditionally used ellagitannin-rich plant materials. *J. Ethnopharmacol.* 155(1), 801–809 (2014)
 18. Adams, L.S., et al.: Pomegranate ellagitannin-derived compounds exhibit antiproliferative and antiaromatase activity in breast cancer cells in vitro. *Cancer Prevent. Res.* 3(1), 108–113 (2010)
 19. Seeram, N.P., et al.: In vitro antiproliferative, apoptotic and antioxidant activities of punicalagin, ellagic acid and a total pomegranate tannin extract are enhanced in combination with other polyphenols as found in pomegranate juice. *J. Nutr. Biochem.* 16(6), 360–367 (2005)
 20. Seeram, N.P., Aronson, W.J., Zhang, Y., Henning, S.M., Moro, A., Lee, R.P., Sartippour, M., Harris, D.M., Rettig, M., Suchard, M.A., Pantuck, A.J.: Pomegranate ellagitannin-derived metabolites inhibit prostate cancer growth and localize to the mouse prostate gland. *J. Agric. Food Chem.* 55(19), 7732–7737 (2007)
 21. Larrosa, M., et al.: Urolithins, ellagic acid-derived metabolites produced by human colonic microflora, exhibit estrogenic and antiestrogenic activities. *J. Agric. Food Chem.* 54(5), 1611–1620 (2006)
 22. Liu, W., et al.: Pomegranate phenolics inhibit formation of advanced glycation endproducts by scavenging reactive carbonyl species. *Food Funct.* 5(11), 2996–3004 (2014)
 23. Bialonska, D., et al.: Urolithins, intestinal microbial metabolites of pomegranate ellagitannins, exhibit potent antioxidant activity in a cell-based assay. *J. Agric. Food Chem.* 57(21), 10181–10186 (2009)
 24. Giménez-Bastida, J.A., et al.: Urolithins, ellagitannin metabolites produced by colon

- microbiota, inhibit quorum sensing in *Yersinia enterocolitica*: phenotypic response and associated molecular changes. *Food Chem.* 132(3), 1465–1474 (2012)
25. González-Barrio, R., et al.: Bioavailability of anthocyanins and ellagitannins following consumption of raspberries by healthy humans and subjects with an ileostomy. *J. Agric. Food Chem.* 58(7), 3933–3939 (2010)
26. Tomás-Barberán, F.A., et al.: Ellagic acid metabolism by human gut microbiota: consistent observation of three urolithin phenotypes in intervention trials, independent of food source, age, and health status. *J. Agric. Food Chem.* 62(28), 6535–6538 (2014)
27. Truchado, P., et al.: Strawberry processing does not affect the production and urinary excretion of urolithins, ellagic acid metabolites, in humans. *J. Agric. Food Chem.* 60(23), 5749–5754 (2011)
28. CDC 2006: Analytical and Reporting Guidelines: The National Health and Nutrition Examination Survey (NHANES). National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Maryland.
http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/nhanes_analytic_guidelines_dec_2005.pdf
29. USDA 2012: What We Eat In America (WWEIA), NHANES: overview. <http://www.ars.usda.gov/Services/docs.htm?docid=13793#release>. Accessed 29 Jan 2018
30. Bodner-Montville, J., Ahuja, J.K.C., Ingwersen, L.A., Haggerty, E.S., Enns, C.W., Perloff, B.P.: USDA food and nutrient database for dietary studies: released on the web. *J. Food Compos. Anal.* 19(Suppl. 1), S100–S107 (2006)
31. Hayes, A.W., Kruger, C.L. (eds.): *Hayes' Principles and Methods of Toxicology*, 6th edn, p. 631. CRC Press, Boca Raton (2014)

Chapter 4 References:

Sources Cited by Company Files:

1. Aiyer, Harini S., Kichambare, Sunita, & Gupta, Ramesh C. (2008). Prevention of Oxidative DNA Damage by Bioactive Berry Components. *Nutrition and Cancer*, 60, 36-42.
2. Aiyer, H., Srinivasan, C., & Gupta, R. (2008). Dietary Berries and Ellagic Acid Diminish Estrogen-Mediated Mammary Tumorigenesis in ACI rats. *Nutrition and Cancer*, 60(2), 227-234.
3. Amakura, Okada, Tsuji, & Tonogai. (2000). High-performance liquid chromatographic determination with photodiode array detection of ellagic acid in fresh and processed fruits. *Journal of Chromatography A*, 896(1), 87-93.
4. Amakura, Okada, Tsuji, & Tonogai. (2000). Determination of phenolic acids in fruit juices by isocratic column liquid chromatography. *Journal of Chromatography A*, 891(1), 183-188.

5. Da Silva, Calgarotto, Char, & Marangoni. (2008). Isolation and characterization of ellagic acid derivatives isolated from *Casearia sylvestris* SW aqueous extract with anti-PLA 2 activity. *Toxicol*, 52(6), 655-666.
6. Damas, J.; Remacle-Volon, G. "Congestion of Lymph Nodes by Ellagic Acid in the Rat." *Naunyn-Schmeideberg's Archives of Pharmacology*. V. 330, Supplement 1. March 1985. P. 223-224.
7. Damas, J., Remacle-Volon, G., & Adam, A. (1987). Mechanism of the congestion of lymph nodes induced by ellagic acid in rats. *Agents and Actions*, 22(3), 202-208.
8. Damas, J., Adam, A., Remacle-Volon, G., & Grek, V. (1987). Studies on the vascular and hematological changes induced by ellagic acid in rats. *Agents and Actions*, 22(3), 270-279.
9. Daniel, E., Ratnayake, S., Kinstle, T., & Stoner, G. (1991). The effects of pH and rat intestinal contents on the liberation of ellagic acid from purified and crude ellagitannins. *Journal of Natural Products*, 54(4), 946-52.
10. Devipriya, N., Sudheer, A., Vishwanathan, P., & Menon, V. (2008). Modulatory potential of ellagic acid, a natural plant polyphenol on altered lipid profile and lipid peroxidation status during alcohol-induced toxicity: A pathohistological study. *Journal of Biochemical and Molecular Toxicology*, 22(2), 101-112.
11. B. Doyle; L. A. Griffiths. "The metabolism of ellagic acid in the rat." *Xenobiotica*, 1980, vol. 10, no. 4, 247-256.
12. Hassoun, E., Vodhanel, J., Holden, B., & Abushaban, A. (2006). The Effects of Ellagic Acid and Vitamin E Succinate on Antioxidant Enzymes Activities and Glutathione Levels in Different Brain Regions of Rats After Subchronic Exposure to TCDD. *Journal of Toxicology and Environmental Health, Part A*, 69(5), 381-393.
13. Hassoun, Walter, Alsharif, & Stohs. (1997). Modulation of TCDD-induced fetotoxicity and oxidative stress in embryonic and placental tissues of C57BL/6J mice by vitamin E succinate and ellagic acid. *Toxicology*, 124(1), 27-37.
14. Hayeshi, Mutingwende, Mavengere, Masiyanise, & Mukanganyama. (2007). The inhibition of human glutathione S-transferases activity by plant polyphenolic compounds ellagic acid and curcumin. *Food and Chemical Toxicology*, 45(2), 286-295.
15. Ho, C., Lai, Y., Wang, D., Chen, Y., Lee, J., Tang, N., & Chung, J. (2005). Effects of ellagic acid by oral administration on distribution and metabolism of 2-aminofluorene in Sprague-Dawley rats. *In Vivo (Athens, Greece)*, 19(1), 143-56.

16. Hurley, R., Bohn, A., & Frank, A. (1995). The effect of ellagic acid on embryonic rat development. The American College of Veterinary Pathologists 46th Annual Meeting. (1995). *Veterinary Pathology*, 32(5), 545-603.
17. Mukhtar, H., Das, M., & Bickers, D. (1986). Inhibition of 3-methylcholanthrene-induced skin tumorigenicity in BALB/c mice by chronic oral feeding of trace amounts of ellagic acid in drinking water. *Cancer Research*, 46(5), 2262-5.
18. Shahrzad, & Bitsch. (1996). Determination of some pharmacologically active phenolic acids in juices by high-performance liquid chromatography. *Journal of Chromatography A*, 741(2), 223-231.
19. Tani, Sakurai, & Kondo. (1986). Effects of tannins and some related compounds on ethanol-induced gastric lesions in the rat. *Yakugaku Zasshi : Journal of the Pharmaceutical Society of Japan*, 106(4), 347-9.
20. Tasaki, Umemura, Maeda, Ishii, Okamura, Inoue, . . . Nishikawa. (2008). Safety assessment of ellagic acid, a food additive, in a subchronic toxicity study using F344 rats. *Food and Chemical Toxicology*, 46(3), 1119-1124.
21. Teel, Robert W.; Martin, Ronald M. "Disposition of the plant phenol ellagic acid in the mouse following oral administration by gavage." *Xenobiotica* Vol. 18, Iss. 4, 1988

ToxNet Sources:

22. Bohn AA; Frank, AA; Forsyth, CS; Kimm, KI; Stoner, GD. Distribution of 14C-ellagic acid following intrauterine injection. *Teratology* 1993 May 47(5):413
23. Castonguay, A., Boukharta, M., & Teel, R. (1998). Biodistribution of, antimutagenic efficacies in *Salmonella typhimurium* of, and inhibition of P450 activities by ellagic acid and one analogue. *Chemical Research in Toxicology*, 11(11), 1258-64.
24. Hazardous Substances Data Bank [Internet]. Bethesda (MD): National Library of Medicine (US); [Last Revision Date ????. cited 2018 July]. Ellagic Acid; Hazardous Substances Databank Number: ???; [about # screens]. CASRN: 476-66-4 Available from: <<https://toxnet.nlm.nih.gov/cgi-bin/sis/search2/f?./temp/~qv6le4:3>>
25. Chemical Carcinogenesis Research Information System (CCRIS) [Internet]. Bethesda (MD): National Library of Medicine (US); [Last Revision Date ????. cited 2018 July]. Ellagic Acid; Hazardous Substances Databank Number: ???; [about # screens]. CASRN: 476-66-4 Available from:

<<https://toxnet.nlm.nih.gov/cgi-bin/sis/search2/f?./temp/~qv6le4:4>>

26. "Ellagic Acid." The Carcinogenic Potency Project. Berkeley Labs. Last updated: October 3, 2007 Accessed 7/30/2018.

<<https://toxnet.nlm.nih.gov/cpdb/chempages/ELLAGIC%20ACID.html>>

27. Comparative Toxicogenomics Database (CTD) [Internet]. Bethesda (MD): National Library of Medicine (US); [Last Revision Date ????. cited 2018 July]. Ellagic Acid; Hazardous Substances Databank Number: ???; [about # screens]. CASRN: 476-66-4 Available from: <<https://toxnet.nlm.nih.gov/cgi-bin/sis/search2/f?./temp/~qv6le4:7>>

28. Haz-Map. [Internet]. Bethesda (MD): National Library of Medicine (US); [Last Revision Date ????. cited 2018 July]. Ellagic Acid; Hazardous Substances Databank CASRN: 476-66-4 Available from:

<https://hazmap.nlm.nih.gov/category-details?table=copytblagents&id=19805>

UK Library Search: "Ellagic Acid NOAEL"

29. Aguilera-Carbo, A., Augur, F., Prado-Barragan, C., Aguilar, L., & Favela-Torres, A. (2008). Extraction and analysis of ellagic acid from novel complex sources. *Chemical Papers*, 62(4), 440-444.

30. Bala, I., Bhardwaj, V., Hariharan, S., Sitterberg, J., Bakowsky, U., & Ravi Kumar, M. (2005). Design of biodegradable nanoparticles: A novel approach to encapsulating poorly soluble phytochemical ellagic acid. *Nanotechnology*, 16(12), 2819-2822.

31. Boniface, Ferreira, & Kaiser. (2016). Recent trends in phytochemistry, ethnobotany and pharmacological significance of *Alchornea cordifolia* (Schumach. & Thonn.) Muell. Arg. *Journal of Ethnopharmacology*, 191, 216-244.

32. Card, J., Jonaitis, T., Tafazoli, S., & Magnuson, B. (2011). An appraisal of the published literature on the safety and toxicity of food-related nanomaterials. *Critical Reviews in Toxicology*, 41(1), 20-49.

33. Chang, Yi, Chen, Wei-Fan, Lin, Kuan-Hung, Hsieh, Cheng-Ying, Chou, Duen-Suey, Lin, Li-Jyun, . . . Chang, Chao-Chien. (2013). Novel Bioactivity of Ellagic Acid in Inhibiting Human Platelet Activation. *Evidence-Based Complementary and Alternative Medicine*, 2013, 9.

34. Heilman, Andreux, Tran, Rinsch, & Blanco-Bose. (2017). Safety assessment of Urolithin A, a metabolite produced by the human gut microbiota upon dietary intake of plant derived ellagitannins and ellagic acid. *Food and Chemical Toxicology*, 108, 289-297.

35. Ismail, T., Calcabrini, C., Diaz, A., Fimognari, C., Turrini, E., Catanzaro, E., . . . Fang, J. (2016). Ellagitannins in Cancer Chemoprevention and Therapy. *Toxins*, 8(5), *Toxins*, 2016, Vol.8(5).
36. Makino-Wakagi, Yoshimura, Uzawa, Zaima, Moriyama, & Kawamura. (2012). Ellagic acid in pomegranate suppresses resistin secretion by a novel regulatory mechanism involving the degradation of intracellular resistin protein in adipocytes. *Biochemical and Biophysical Research Communications*, 417(2), 880-885.
37. Medjakovic, S., & Jungbauer, A. (2012). Pomegranate: A fruit that ameliorates metabolic syndrome. *Food & Function*, 4(1), 19-39.
38. Mishra, S., & Vinayak, M. (2014). Ellagic Acid Induces Novel and Atypical PKC Isoforms and Promotes Caspase-3 Dependent Apoptosis by Blocking Energy Metabolism. *Nutrition and Cancer*, 66(4), 675-681.
39. Mishra, S., & Vinayak, M. (2015). Role of ellagic acid in regulation of apoptosis by modulating novel and atypical PKC in lymphoma bearing mice. *BMC Complementary and Alternative Medicine*, 15, 281.
40. Moore, J., Yousef, M., & Tsiani, E. (2016). Anticancer Effects of Rosemary (*Rosmarinus officinalis* L.) Extract and Rosemary Extract Polyphenols. *Nutrients*, 8(11), <xocs:firstpage xmlns:xocs=""/>
41. Pavan, Damazo, Lemos, Adzu, Balogun, Arunachalam, & Martins. (2018). Evaluation of genotoxicity and subchronic toxicity of the standardized leaves infusion extract of *Copaifera malmei* Harms in experimental models. *Journal of Ethnopharmacology*, 211, 70-77.
42. Promsong, Chuenchitra, Saipin, Tewtrakul, Panichayupakaranant, Satthakarn, & Nittayananta. (2018). Ellagic acid inhibits HIV-1 infection in vitro: Potential role as a novel microbicide. *Oral Diseases*, 24(1-2), 249-252.
43. Reddy, B., Mullick, R., Kumar, A., Sudha, G., Srinivasan, N., & Das, S. (2014). Small molecule inhibitors of HCV replication from pomegranate. *Scientific Reports*, 4, 5411.
44. Rhomberg, L., Goodman, J., Haber, L., Dourson, M., Andersen, M., Klaunig, J., . . . Cohen, S. (2011). Linear low-dose extrapolation for noncancer health effects is the exception, not the rule. *Critical Reviews in Toxicology*, 41(1), 1-19.

Google Scholar Search: “ellagic acid noel”

45. Ahmed, T.; Setzer, W; Nabavi, S.F.; Orhan, I.E.; Braidy, N.; Sobarzo-Sanchez, E.; Nabavi, S.M. 2016. “Insights Into Effects of Ellagic Acid on the Nervous System: A Mini

Review". *Current Pharmaceutical Design*, 2016, 22, 1350-1360

46. Anitha, P., Priyadarsini, R., Kavitha, V., Thiyagarajan, K., & Nagini, P. (2013). Ellagic acid coordinately attenuates Wnt/ β -catenin and NF- κ B signaling pathways to induce intrinsic apoptosis in an animal model of oral oncogenesis. *European Journal of Nutrition*, 52(1), 75-84.
47. Beserra, A., Calegari, P., Souza, M., Dos Santos, R., Lima, J., Silva, R., . . . Martins, D. (2011). Gastroprotective and ulcer-healing mechanisms of ellagic acid in experimental rats. *Journal of Agricultural and Food Chemistry*, 59(13), 6957-65.
48. Bhandary, B. SATHEESH KUMAR, et al. "Acute and subacute toxicity study of the ethanol extracts of *Punica granatum* (Linn). Whole fruit and seeds and synthetic ellagic acid in swiss albino mice." *Asian J Pharm Clin Res* 6.4 (2013): 192-198.
49. Cerdá, B., Cerón, J., Tomás-Barberán, F., & Espín, J. (2003). Repeated oral administration of high doses of the pomegranate ellagitannin punicalagin to rats for 37 days is not toxic. *Journal of Agricultural and Food Chemistry*, 51(11), 3493-501.
50. Dinesh Dhingra, Ritu Chhillar. (2012) Antidepressant-like activity of ellagic acid in unstressed and acute immobilization-induced stressed mice, *Pharmacological Reports*, Volume 64, Issue 4, 2012, Pages 796-807, ISSN 1734-1140, <[https://doi.org/10.1016/S1734-1140\(12\)70875-7](https://doi.org/10.1016/S1734-1140(12)70875-7)> (<<http://www.sciencedirect.com/science/article/pii/S1734114012708757>>)
51. Engelke, L., Hamacher, A., Proksch, P., & Kassack, M. (2016). Ellagic Acid and Resveratrol Prevent the Development of Cisplatin Resistance in the Epithelial Ovarian Cancer Cell Line A2780. *Journal of Cancer*, 7(4), 353-363.
52. Espín, García-Conesa, & Tomás-Barberán. (2007). Nutraceuticals: Facts and fiction. *Phytochemistry*, 68(22), 2986-3008.
53. Faria, A., & Calhau, C. (2011). The Bioactivity of Pomegranate: Impact on Health and Disease. *Critical Reviews in Food Science and Nutrition*, 51(7), 626-634.
54. Ismail, Sestili, & Akhtar. (2012). Pomegranate peel and fruit extracts: A review of potential anti-inflammatory and anti-infective effects. *Journal of Ethnopharmacology*, 143(2), 397-405.
55. Łabieniec, & Gabryelak. (2006). Oxidatively modified proteins and DNA in digestive gland cells of the fresh-water mussel *Unio tumidus* in the presence of tannic acid and its derivatives. *Mut.Res.-Genetic Toxicology and Environmental Mutagenesis*, 603(1), 48-55.

56. Kang, Inhae Kang; Buckner, Teresa; Shay, Neil F; Gu, Liwei; Chung, Soonkyu. "Improvements in Metabolic Health with Consumption of Ellagic Acid and Subsequent Conversion into Urolithins: Evidence and Mechanisms." *Advances in Nutrition*, Volume 7, Issue 5, 1 September 2016, Pages 961–972, <<https://doi.org/10.3945/an.116.012575>>
57. Mandal S and Stoner GD: Inhibition of N-nitrosobenzylmethylamine-induced esophageal tumorigenesis in rats by ellagic acid. *Carcinogenesis* 11, 55-61, 1990
58. Muscat, J. E., Stellman, S. D. and Wynder, E. L. (1994), Nonsteroidal antiinflammatory drugs and colorectal cancer. *Cancer*, 74: 1847-1854.
doi:10.1002/1097-0142(19941001)74:7<1847::AID-CNCR2820740704>3.0.CO;2-#
59. Ndjonka, D., Abladam, E., Djafsia, B., Ajonina-Ekoti, I., Liebau, M., & Achukwi. (2014). Anthelmintic activity of phenolic acids from the axlewood tree *Anogeissus leiocarpus* on the filarial nematode *Onchocerca ochengi* and drug-resistant strains of the free-living nematode *Caenorhabditis elegans*. *Journal of Helminthology*, 88(4), 481-488.
60. Ndjonka, D., Bergmann, B., Agyare, C., Zimbres, F., Lüersen, M., Hensel, K., . . . Liebau, C. (2012). In vitro activity of extracts and isolated polyphenols from West African medicinal plants against *Plasmodium falciparum*. *Parasitology Research*, 111(2), 827-834.
61. Paller, C., Rudek, M., Zhou, X., Wagner, W., Hudson, T., Anders, N., . . . Carducci, M. (2015). A phase I study of muscadine grape skin extract in men with biochemically recurrent prostate cancer: Safety, tolerability, and dose determination. *Prostate*, 75(14), 1518-1525.
62. Viladomiu, M., Hontecillas, R., Lu, P., & Bassaganya-Riera, J. (2013). Preventive and Prophylactic Mechanisms of Action of Pomegranate Bioactive Constituents. *Evidence-Based Complementary and Alternative Medicine*, 2013, 18.
63. Wang, S. (2017). A food-grade self-nanoemulsifying delivery system for enhancing oral bioavailability of ellagic acid. *Journal of Functional Foods*, 207-215.
64. Wang, Ru-Feng; Yi , Ding; Ruining Liu; Lan Xiang; Lijun, Du. (2010). "Pomegranate: Constituents." *Bioactivities and Pharmacokinetics. Fruit Veg. Cereal Sci. Biotechnol.* 4.

Citations Made By Papers:

65. Adams, L.S., Seeram, N.P., Aggarwal, B.B., Takada, Y., Sand, D., & Heber, D. (2006). "Pomegranate juice, total pomegranate ellagitannins, and punicalagin suppress inflammatory cell signaling in colon cancer cells." *Journal of Agricultural and Food Chemistry*, 54(3), 980-985.

66. Ahad, Amjid; Ganai, Ajaz Ahmad; Mujeeb, Mohd; Siddiqui, Waseem Ahmad. (2014) "Ellagic acid, an NF- κ B inhibitor, ameliorates renal function in experimental diabetic nephropathy." *Chemico-Biological Interactions*, Volume 219, 2014, Pages 64-75, ISSN 0009-2797, <<https://doi.org/10.1016/j.cbi.2014.05.011>> <<http://www.sciencedirect.com/science/article/pii/S0009279714001690>>
67. Ahn, Putt, Kresty, Stoner, Fromm, & Hollenberg. (1996). "The effects of dietary ellagic acid on rat hepatic and esophageal mucosal cytochromes P450 and phase II enzymes." *Carcinogenesis*, 17(4), 821-8.
68. Akileswar, Chandrasekhar; Raghua, Ganugula; Muthenna, Puppala; Muellerb, Niklaus H.; Suryanaryana, Palla; Petrashb, J. Mark; Reddy, G. Bhanuprakash. (2014.) "Bioflavonoid ellagic acid inhibits aldose reductase: Implications for prevention of diabetic complications." *Journal of Functional Foods* 6, 374-383.
69. Ateşşahin, A., Çeribaşı, A., Yuçe, A., Bulmus, &, & Çikim, G. (2007). "Role of Ellagic Acid against Cisplatin-Induced Nephrotoxicity and Oxidative Stress in Rats." *Basic & Clinical Pharmacology & Toxicology*, 100(2), 121-126.
70. Bae, Ji-Young, Choi, Jung-Suk, Kang, Sang-Wook, Lee, Yong-Jin, Park, Jinseu, & Kang, Young-Hee. (2010). "Dietary compound ellagic acid alleviates skin wrinkle and inflammation induced by UV-B irradiation." *Experimental Dermatology*, 19(8), E182-E190.
71. Cerdá, B., Periago, P., Espín, J., & Tomás-Barberán, F. (2005). Identification of urolithin a as a metabolite produced by human colon microflora from ellagic acid and related compounds. *Journal of Agricultural and Food Chemistry*, 53(14), 5571-6.
72. Cerdá, B., Tomás-Barberán, F., & Espín, J. (2005). Metabolism of antioxidant and chemopreventive ellagitannins from strawberries, raspberries, walnuts, and oak-aged wine in humans: Identification of biomarkers and individual variability. *Journal of Agricultural and Food Chemistry*, 53(2), 227-35.
73. Chao CY, Mong MC, Chan KC, Yin MC. (2010) "Anti-glycative and anti-inflammatory effects of caffeic acid and ellagic acid in kidney of diabetic mice." *Mol Nutr Food Res* 2010;54:388–95.
74. Damas, J., Remeclé-Volon, G.,)1987). 'The thrombopenic effects of ellagic acid in the rat. Another model of platelet stimulation "in vivo".' *Thrombosis Research* 45, 153-163.
75. Devipriya, N.; Srinivasan, M.; Sudheer, A. R.; Menon, V. P. (2007) "Effect of ellagic

acid, a natural polyphenol, on alcohol-induced prooxidant and antioxidant imbalance: a drug dose dependent study." *Singapore Med. J.* 2007, 48, 311–318.

76. Ding, Zhang, Zhou, Chen, Wang, Jia, . . . Wen. (2014). "Dietary ellagic acid improves oxidant-induced endothelial dysfunction and atherosclerosis: Role of Nrf2 activation." *International Journal of Cardiology*, 175(3), 508-514.

77. Dolatshahi, M., Farbood, Y., Sarkaki, A., Mansouri, S., & Khodadadi, A. (2015). Ellagic acid improves hyperalgesia and cognitive deficiency in 6-hydroxidopamine induced rat model of Parkinson's disease. *Iranian Journal of Basic Medical Sciences*, 18(1), 38-46.

78. Doyle, B.; Griffiths, L.A. (1980) "The metabolism of ellagic acid in the rat," *Xenobiotica*, 10:4, 247-256, DOI: 10.3109/00498258009033752

79. Daniel, E.M.; Stoner, G.D. The effects of ellagic acid and 13-cis-retinoic acid on N-nitrosobenzylmethylamine-induced esophageal tumorigenesis in rats. *Cancer Lett.* 1991, 56, 117–124.

80. Espín, J., Larrosa, M., García-Conesa, M., & Tomás-Barberán, F. (2013). Biological Significance of Urolithins, the Gut Microbial Ellagic Acid-Derived Metabolites: The Evidence So Far. *Evidence-Based Complementary and Alternative Medicine*, 2013, 15.

81. Espín, Juan Carlos; González-Barrio, Rocío; Cerdá, Begoña; López-Bote, Clemente; Rey, Ana I.; Tomás-Barberán, Francisco A.. (2007) "Iberian Pig as a Model To Clarify Obscure Points in the Bioavailability and Metabolism of Ellagitannins in Humans." *Journal of Agricultural and Food Chemistry* 2007 55 (25), 10476-10485 DOI: 10.1021/jf0723864

82. Falsaperla, Morgia, Tartarone, Ardito, & Romano. (2005). "Support Ellagic Acid Therapy in Patients with Hormone Refractory Prostate Cancer (HRPC) on Standard Chemotherapy Using Vinorelbine and Estramustine Phosphate." *European Urology*, 47(4), 449-455.

83. Farbood, Sarkaki, Dianat, Khodadadi, Haddad, & Mashhadizadeh. (2015). Ellagic acid prevents cognitive and hippocampal long-term potentiation deficits and brain inflammation in rat with traumatic brain injury. *Life Sciences*, 124(C), 120-127.

84. García-Villalba, R., Beltrán, D., Espín, J., Selma, M., & Tomás-Barberán, F. (2013). Time course production of urolithins from ellagic acid by human gut microbiota. *Journal of Agricultural and Food Chemistry*, 61(37), 8797-806.

85. Ghorbanzadeh, Mansouri, Hemmati, Naghizadeh, Mard, & Rezaie. (2014).

- “Involvement of L-arginine/NO/cGMP/KATP channel pathway in the peripheral antinociceptive actions of ellagic acid in the rat formalin test.” *Pharmacology, Biochemistry and Behavior*, 126(C), 116-121.
86. Girish, Chandrashekar; Raj, Vishnu; Arya, Jayasree; Balakrishnan, Sadasivam. “Evidence for the involvement of the monoaminergic system, but not the opioid system in the antidepressant-like activity of ellagic acid in mice.” *European Journal of Pharmacology*, Volume 682, Issues 1–3, 2012, Pages 118-125, ISSN 0014-2999, <<https://doi.org/10.1016/j.ejphar.2012.02.034>> (<http://www.sciencedirect.com/science/article/pii/S0014299912001707>)
87. González-Barrio, R. A., Crozier, A., & Edwards, C. (2011). Colonic catabolism of ellagitannins, ellagic acid, and raspberry anthocyanins: In vivo and in vitro studies. *Drug Metabolism and Disposition*, 39(9), 1680-1688.
88. González-Sarrías, Antonio & Villalba, Rocío & Nuñez Sanchez, Maria Angeles & Carneiro, Joao & Zafrilla, Pilar & Mulero, Juana & Tomás-Barberán, Francisco & Espín, Juan Carlos. (2015). Identifying the limits for ellagic acid bioavailability: A crossover pharmacokinetic study in healthy volunteers after consumption of pomegranate extracts. *Journal of Functional Foods*. 19. 225-235. 10.1016/j.jff.2015.09.019
89. Gümüş, M., Yüksel, H., Evliyaoğlu, O., Kapan, M., Böyük, A., Önder, A., & Aldemir, M. (2011). Effects of Ellagic Acid on Copper, Zinc, and Biochemical Values in Serum and Liver of Experimental Cholestatic Rats. *Biological Trace Element Research*, 143(1), 386-393.
90. Häkkinen SH, Kärenlampi SO, Mykkänen HM, Heinonen IM, Törrönen AR. (2000) “Ellagic acid content in berries: Influence of domestic processing and storage.” *Eur Food Res Technol* 2000; 212: 75-80.
91. Hamad, R.; Abdul-Wahab & al momani, Waleed & Janakat, Sana & Oran, Sawsan. (2009). “Bioavailability of Ellagic Acid After Single Dose Administration Using HPLC.” *Pakistan Journal of Nutrition*. 8. 10.3923/pjn.2009.1661.1664.
92. Hassoun, E., Vodhanel, J., & Abushaban, A. (2004). The modulatory effects of ellagic acid and vitamin E succinate on TCDD-induced oxidative stress in different brain regions of rats after subchronic exposure. *Journal of Biochemical and Molecular Toxicology*, 18(4), 196-203.
93. Heber, D., Seeram, N.P., Wyatt, H., Henning, S.M., Zhang, Y., Ogden, L.G., Dreher, M., Hill, J.O.. (2007) “Safety and antioxidant activity of a pomegranate ellagitannin-enriched polyphenol dietary supplement in overweight individuals with increased waist size.” *Journal of Agricultural and Food Chemistry* 55, 10050–10054

94. Huang, S.T.; Wang, C.Y.; Yang, R.C.; Wu, H.T.; Yang, S.H.; Cheng, Y.C.; Pang, J.H.S.. (2011) "Ellagic acid, the active compound of *Phyllanthus urinaria*, exerts in vivo anti-angiogenic effect and inhibits MMP-2 activity." *Evid. Based Complement, Altern. Med.* 2011. doi:10.1093/ecam/nep207.
95. Iino, Ogawa, Tashima, Kato, & Takeuchi. (2001). "Less damaging effect of whisky in rat stomachs in comparison with pure ethanol: Role of ellagic acid, the nonalcoholic ingredient." *Gastroenterology*, 120(5), A150.
96. Kang, Eun Hwa, Kwon, Tae Young, Oh, Goo Taeg, Park, Weung Feel, Park, Sung-II, Park, Sung Kyu, & Lee, Young Ik. (2006). "The flavonoid ellagic acid from a medicinal herb inhibits host immune tolerance induced by the hepatitis B virus-e antigen." *Antiviral Research*, 72(2), 100-106.
97. Kannan, M. Mari et al. (2013) "Ellagic acid inhibits cardiac arrhythmias, hypertrophy and hyperlipidaemia during myocardial infarction in rats." *Metabolism - Clinical and Experimental*, Volume 62, Issue 1, 52 - 61
98. Khanduja KL, Gandhi RK, Pathania V, Syal N: (1997) "Prevention of N-nitrosodiethylamine-induced lung tumorigenesis by ellagic acid and quercetin in mice." *Food Chem Toxicol*, 1999, 37, 313–318.
99. Kim, S., Nishimoto, S., Bumgardner, J., Haggard, W., Gaber, M., & Yang, Y. (2010). "A chitosan/beta-glycerophosphate thermo-sensitive gel for the delivery of ellagic acid for the treatment of brain cancer." *Biomaterials*, 31(14), 4157-66.
100. Kim S, Gaber MW, Zawaski JA, Zhang F, Richardson M, Zhang XA, et al. (2009) "The inhibition of glioma growth in vitro and in vivo by a chitosan/ellagic acid composite biomaterial." *Biomaterials* 2009;30(27):4743–51.
101. Kowshik, J.; Giri, H.; Kranthi KiranKishore, T.; Kesavan, R.; Naik Vankudavath, R.; Bhanuprakash Reddy, G.; Dixit, M.; Nagini, S. (2014) "Ellagic acid inhibits VEGF/VEGFR2, PI3K/Akt and MAPK signaling cascades in the hamster cheek pouch carcinogenesis model." *Anti-Cancer Agents Med. Chem.* 2014, 14, 1249–1260.
102. Larrosa, M.; García-Conesa, M.T.; Espín, J.C.; Tomás-Barberán, F.A. "Bioavailability and Metabolism of Ellagic Acid and Ellagitannins." CRC Press: Boca Raton, FL, USA, 2012.
103. Larrosa M, García-Conesa MT, Espín JC, Tomás-Barberán FA. (2010) "Ellagitannins, ellagic acid and vascular health." *Mol Aspects Med* 2010; 31: 513-39.

104. Lei, Xing, Xiang, Zhao, Wang, Zhang, & Du. (2003). "Pharmacokinetic study of ellagic acid in rat after oral administration of pomegranate leaf extract." *Journal of Chromatography B*, 796(1), 189-194.
105. Lin SS, Hung CF, Ho CC, Liu YH, Ho HC, Chung JG. "Effects of ellagic acid by oral administration on N-acetylation and metabolism of 2-aminofluorene in rat brain tissues." *Neurochem Res* 2000; 25: 1503-8.
106. Liu, N.; Liu, J. T.; Zhang, Q. Z.. (2010) "Ellagic acid-induced hypercoagulable state in animals: a potentially useful animal hypercoagulable model for evaluation of anticoagulants," *Chinese Medical Sciences Journal*, vol. 25, no. 4, pp. 237–242, 2010.
107. Malini, P; Kanchana, G; Rajadurai, M. (2011) "Antidiabetic Efficacy of Ellagic Acid in Streptozotocin induced Diabetes Mellitus in Albino Wistar Rats." *Asian J Pharm Clin Res* 2011; 4(3):124-128.
108. Mansouri, M.T., Naghizadeh B, Ghorbanzadeh B, Farbood Y. (2013) "Central and peripheral antinociceptive effects of ellagic acid in different animal models of pain." *Eur J Pharmacol* 2013; 707: 46-53.
109. Mansouri MT, Naghizadeh B, Ghorbanzadeh B. (2014) "Involvement of opioid receptors in the systemic and peripheral antinociceptive actions of ellagic acid in the rat formalin test." *Pharmacol Biochem Behav* 2014; 120: 43-9.
110. Mertens-Talcott, S., & Percival, S. (2005). Ellagic acid and quercetin interact synergistically with resveratrol in the induction of apoptosis and cause transient cell cycle arrest in human leukemia cells. *Cancer Letters*, 218(2), 141-51.
111. Mishra, S., & Vinayak, M. (2011). "Anti-carcinogenic action of ellagic acid mediated via modulation of oxidative stress regulated genes in Dalton lymphoma bearing mice." *Leukemia & Lymphoma*, 2011, Vol.52(11), P.2155-2161, 52(11), 2155-2161.
112. Mishra, S., & Vinayak, M. (2013). "Ellagic acid checks lymphoma promotion via regulation of PKC signaling pathway." *Molecular Biology Reports*, 40(2), 1417-1428.
113. Murugan, Venkatesh, Mukherjee, Kakali, Maiti, Kuntal, & Mukherjee, Pulok K. (2009). Enhanced oral bioavailability and antioxidant profile of ellagic acid by phospholipids. *Journal of Agricultural and Food Chemistry*, 57(11), 4559-65.
114. Nejad, Khojasteh Hoseiny et al. (2018) "Ellagic Acid Improves Electrocardiogram Waves and Blood Pressure against Global Cerebral Ischemia Rat Experimental Models." *Electronic Physician* 7.4 (2015): 1153–1162. PMC. Web. 16 July 2018.

115. Ogawa, Y.; Kanatsu, K.; Iino, T.; Kato, S.; Jeong, Y. I.; Shibata, N.; Takada, K.; Takeuchi, K. (2002) "Protection against dextran sulphate sodium-induced colitis by microspheres of ellagic acid in rats." *Life Sci.* 2002, 71, 827–839.
116. Panchal, S., Ward, K., & Brown, L. (2013). "Ellagic acid attenuates high-carbohydrate, high-fat diet-induced metabolic syndrome in rats." *European Journal of Nutrition*, 52(2), 559-568.
117. Paivarinta, E., Pajari, A., Torronen, R., & Mutanen, M. (2006). "Ellagic acid and natural sources of ellagitannins as possible chemopreventive agents against intestinal tumorigenesis in the Min mouse." *Nutrition And Cancer-An International Journal*, 54(1), 79-83.
118. Perchellet JP, Gali HU, Perchellet EM, Klish DS, Armbrust AD. (1992) "Antitumor-promoting activities of tannic acid, ellagic acid, and several gallic acid derivatives in mouse skin." *Basic Life Sci.* 1992;59:783-801
119. Priyadarsini, R.V.; Kumar, N.; Khan, I.; Thiagarajan, P.; Kondaiah, P.; Nagini, S. (2012) "Gene expression signature of DMBA-induced hamster buccal pouch carcinomas: Modulation by chlorophyllin and ellagic acid." *PLoS ONE* 2012, 7, e34628.
120. Rani P., Kesavan, Ganugula, T., Kumar P., Reddy, & Dixit. (2013). "Ellagic acid inhibits PDGF-BB-induced vascular smooth muscle cell proliferation and prevents atheroma formation in streptozotocin-induced diabetic rats." *The Journal of Nutritional Biochemistry*, 24(11), 1830-1839.
121. Rao, Chinthalapally V.; Tokumo, Kenji; Rigotty, Jeff; Zang, Edith; Kelloff, Gary; and Reddy, Bandaru S. (1991) "Chemoprevention of Colon Carcinogenesis by Dietary Administration of Piroxicam, a-Difluoromethylornithine, 16a-Fluoro-5-androsten-17-one, and Ellagic Acid Individually and in Combination." *Cancer Research* 51, 4528-4534, September 1, 1991.
122. Ratnam DV; Chandraiah G; Meena AK; Ramarao P; Kumar MN. (2009) "The co-encapsulated antioxidant nanoparticles of ellagic acid and coenzyme Q10 ameliorates hyperlipidemia in high fat diet fed rats." *J Nanosci Nanotechnol.* 2009 Nov;9(11):6741-6.
123. Rehman, Tahir, Ali, Qamar, Lateef, Khan, . . . Sultana. (2012). "Cyclophosphamide-induced nephrotoxicity, genotoxicity, and damage in kidney genomic DNA of Swiss albino mice: The protective effect of Ellagic acid." *Molecular and Cellular Biochemistry*, 365(1), 119-127.
124. Rogerio AP, Fontanari C, Borducchi E, Keller AC, Rusco M, Soares EG,

Albuquerque DA, Faccioli LH. (2008) "Anti-inflammatory effects of *Lafoensia pacari* and ellagic acid in a murine model of asthma." *Eur J Pharmacol* 580:262–270

125. Rogerio, A., Fontanari, C., Melo, M., Ambrosio, S., Souza, G., Pereira, P., . . . Faccioli, L. (2006). "Anti-inflammatory, analgesic and anti-oedematous effects of *Lafoensia pacari* extract and ellagic acid." *Journal of Pharmacy and Pharmacology*, 58(9), 1265-1273.

126. Romo Vaquero, María; Villalba, Rocío; González-Sarrías, Antonio; Beltrán, David; Tomás-Barberán, Francisco; Espín, Juan Carlos; Selma, María. (2015). "Interindividual variability in the human metabolism of ellagic acid: Contribution of *Gordonibacter* to urolithin production. *Journal of Functional Foods*." 17. 785-791. 10.1016/j.jff.2015.06.040.

127. Saiko, Steinmann, Schuster, Graser, Bressler, Giessrigl, . . . Szekeres. (2015). "Epigallocatechin gallate, ellagic acid, and rosmarinic acid perturb dNTP pools and inhibit de novo DNA synthesis and proliferation of human HL-60 promyelocytic leukemia cells: Synergism with arabinofuranosylcytosine." *Phytomedicine*, 22(1), 213-222.

128. Scalbert, A., & Williamson, G. (2000). "Dietary intake and bioavailability of polyphenols." *Journal Of Nutrition*, 130(8), 2073S-2085S

129. Seeram, Lee, & Heber. (2004). "Bioavailability of ellagic acid in human plasma after consumption of ellagitannins from pomegranate (*Punica granatum L.*) juice." *Clinica Chimica Acta*, 348(1), 63-68.

130. Seeram, N., Adams, L., Henning, S., Niu, Y., Zhang, Y., Nair, M., & Heber, D. (2005). "In vitro antiproliferative, apoptotic and antioxidant activities of punicalagin, ellagic acid and a total pomegranate tannin extract are enhanced in combination with other polyphenols as found in pomegranate juice." *Journal of Nutritional Biochemistry*, 16(6) 360 - 367. 2005. ISSN: 0955-2863

131. Seeram NP, Aronson WJ, Zhang Y, Henning SM, Moro A, Lee RP, Sartippour M, Harris DM, Rettig M, Suchard MA, Pantuck AJ, Belldegrün A, Heber D. (2007) "Pomegranate ellagitannin-derived metabolites inhibit prostate cancer growth and localize to the mouse prostate gland." *J Agric Food Chem*. 2007; 55(19):7732–7. [PubMed: 17722872]

132. Sepúlveda L, Ascacio A, Rodríguez-Herrera R, Aguilera-Carbó A, Aguilar CN. (2013) "Ellagic acid: Biological properties and biotechnological development for production processes." *Afr J Biotechnol* 2013; 10: 4518-23.

133. Shimogaki, Tanaka, Tamai, & Masuda. (2000). "In vitro and in vivo evaluation of

ellagic acid on melanogenesis inhibition." *International Journal of Cosmetic Science*, 22(4), 291-303.

134. Singh K, Khanna AK, and Chander R. (1999) "Hepatoprotective effect of ellagic acid against carbon tetrachloride induced hepatotoxicity in rats." *Indian J Exp Biol* 37, 1025–1026, 1999.

135. Singletary K, Liao CH (1989) "Ellagic acid effects on the carcinogenicity, DNA-binding and metabolism of 7, 12-dimethyl-benz(a)anthracene (DMBA)." *In Vivo* 3:173–175

136. Soh, Patrice Njomnang, Witkowski, Benoit, Olagnier, David, Nicolau, Marie-Laure, Garcia-Alvarez, Maria-Concepcion, Berry, Antoine, & Benoit-Vical, Françoise. (2009). "In Vitro and In Vivo Properties of Ellagic Acid in Malaria Treatment." *Antimicrobial Agents and Chemotherapy*, 53(3), 1100-1106.

137. Suzuki, Noriaki, Masamune, Atsushi, Kikuta, Kazuhiro, Watanabe, Takashi, Satoh, Kennichi, & Shimosegawa, Tooru. (2009). "Ellagic Acid Inhibits Pancreatic Fibrosis in Male Wistar Bonn/Kobori Rats." *Digestive Diseases and Sciences*, 54(4), 802-810.

138. Takagi, A; Sai, K; Umemura, T; Hasegawa, R; Kurokawa, Y. (1995) "Inhibitory effects of vitamin E and ellagic acid on 8-hydroxyguanosine formation in liver nuclear DNA of rats treated with 2-nitropropane." *Cancer Lett* 1995;91:139–144.

139. Tang B, Chen GX, Liang MY, Yao JP, Wu ZK. (2015) "Ellagic acid prevents monocrotaline-induced pulmonary artery hypertension via inhibiting NLRP3 inflammasome activation in rats." *Int J Cardiol* 2015; 180: 134-41.

140. Teel, Robert W. (1987) "Distribution and metabolism of ellagic acid in the mouse following intraperitoneal administration." *Cancer Letters* , Volume 34 , Issue 2 , 165 - 171

141. Thresiamma, K. C. & Kuttan, R. (1996) "Inhibition of liver fibrosis by ellagic acid." *Indian J. Physiol. Pharmacol.* 40, 363–366. 1996.

142. Thresiamma, K. C.; George, J.; Kuttan, R. (1998) "Protective effect of curcumin, ellagic acid and bixin on radiation induced genotoxicity." *Journal of Experimental and Clinical Cancer Research*, vol. 17, no. 4, pp. 431–434, 1998.

143. Tomás-Barberán, F., García-Villalba, R., González-Sarrías, A., Selma, M., & Espín, J. (2014). "Ellagic acid metabolism by human gut microbiota: Consistent observation of three urolithin phenotypes in intervention trials, independent of food source, age, and health status." *Journal of Agricultural and Food Chemistry*, 62(28), 6535-8.

144. Türk, Ateşşahin, Sönmez, Çeribaşı, & Yüce. (2008). "Improvement of cisplatin-induced injuries to sperm quality, the oxidant-antioxidant system, and the histologic structure of the rat testis by ellagic acid." *Fertility and Sterility*, 89(5), 1474-1481.
145. Ueda H, Kawanishi K, Moriyasu M. (2004) "Effects of ellagic acid and 2-(2,3,6-trihydroxy-4-carboxyphenyl)ellagic acid on sorbitol accumulation in vitro and in vivo." *Biol Pharm Bull* 2004; 27: 1584-7.
146. Umesalma, S., & Sudhandiran, G. (2010). "Differential inhibitory effects of the polyphenol ellagic acid on inflammatory mediators NF-kappaB, iNOS, COX-2, TNF-alpha, and IL-6 in 1,2-dimethylhydrazine-induced rat colon carcinogenesis." *Basic & Clinical Pharmacology & Toxicology*, 107(2), 650-5.
147. Uzar, E., Alp, H., Cevik, M., Fırat, U., Evliyaoglu, O., Tufek, A., & Altun, Y. (2012). Ellagic acid attenuates oxidative stress on brain and sciatic nerve and improves histopathology of brain in streptozotocin-induced diabetic rats. *Neurological Sciences*, 33(3), 567-574.
148. Vatter, D., & Shetty, K. (2005). "Biological Functionality Of Ellagic Acid: A Review". *Journal of Food Biochemistry*, 29(3), 234-266.
149. Whitley, Stoner, Darby, & Walle. (2003). "Intestinal epithelial cell accumulation of the cancer preventive polyphenol ellagic acid—extensive binding to protein and DNA." *Biochemical Pharmacology*, 66(6), 907-915.
150. Woo MS, Choi HS, Seo MJ, Jeon HJ, Lee BY. (2015) "Ellagic acid suppresses lipid accumulation by suppressing early adipogenic events and cell cycle arrest." *Phytother Res* 2015; 29: 398-406.
151. Yang CS, Tzou BC, Liu YP, Tsai MJ, Shyue SK, Tzeng SF. (2008) "Inhibition of cadmium-induced oxidative injury in rat primary astrocytes by the addition of antioxidants and the reduction of intracellular calcium." *J Cell Biochem* 2008; 103: 825-34.
152. Yoshimura, Yukihiro; Nishii, Saori; Zaima, Nobuhiro; Moriyama, Tatsuya; Kawamura, Yukio; (2013) "Ellagic acid improves hepatic steatosis and serum lipid composition through reduction of serum resistin levels and transcriptional activation of hepatic ppara in obese, diabetic KK-Ay mice." *Biochemical and Biophysical Research Communications*, Volume 434, Issue 3, 2013, Pages 486-491, ISSN 0006-291X, <<https://doi.org/10.1016/j.bbrc.2013.03.100>> <<http://www.sciencedirect.com/science/article/pii/S0006291X13005792>>
153. Yu, Chang, Wu, & Chiang. (2005). "Reduction of oxidative stress and apoptosis in

hyperlipidemic rabbits by ellagic acid.” *The Journal of Nutritional Biochemistry*, 16(11), 675-681.

Other References (Not Studies Included in Meta-Analysis)

154. Dickerson, Cynthia; Ensor, Mark; Lodder, Robert A. “Establishing EDI for a Clinical Trial of a Treatment for Chikungunya.” (2018). In: Shi Y. et al. (eds) *Computational Science – ICCS 2018*. ICCS 2018. Lecture Notes in Computer Science, vol 10861. Springer, Cham

155. Nair, Anroop B., and Shery Jacob. (2016). “A Simple Practice Guide for Dose Conversion between Animals and Human.” *Journal of Basic and Clinical Pharmacy* 7.2 (2016): 27–31. PMC. Web. 10 Oct. 2018.

Chapter 5 References:

1. Drennen, James K. et al. “Nondestructive Near-Infrared Analysis of Intact Tablets for Determination of Degradation Products.” *Journal of Pharmaceutical Sciences* , Volume 79 , Issue 7 , 622 - 627
2. Drennen, James K.; Lodder, R. A.. 1988. “Qualitative Analysis Using Near Infrared Spectroscopy: A Comparison of Discriminant Methods in Dissolution Testing.” *Spectroscopy*, 1991, October, 6(8), 34-39.
3. Lodder, RA. “Can You Find Intelligent Communications in Ultrahigh -Dimensional Big Data from Near-Infrared Optical SETI?” SoCIA 2018 (Reno, NV) April 2018
4. Lodder, R. A. Method for detecting subpopulations in spectral analysis. 6-9-1992. Google Patents.
5. Lodder, R. A.; Hieftje, G. M. Detection of Capsule Tampering by Near-Infrared Reflectance Analysis. *Analytical Chemistry* 1987, 59 (15), 1921-1930.
6. Lodder, R. A.; Hieftje, G. M. Detection of Subpopulations in Near-Infrared Reflectance Analysis. *Applied Spectroscopy* 1988, 42 (8), 1500-1512.
7. Lodder, R. A.; Hieftje, G. M. Quantile BEAST Attacks the False-Sample Problem in Near Infrared Reflectance Analysis. *Applied Spectroscopy* 1988, 42 (8), 1353-1365.
8. MultiPolyRegress program by Ahmet Cecen.
<<http://www.ahmetcecen.tech/MultiPolyRegress-MatlabCentral/>>

Conclusion References:

1. Anderson, L.V. “Workplace Wellness Programs are a Sham.” *Slate.com*. September 1 2016 5:45 AM.
<http://www.slate.com/articles/health_and_science/the_ladder/2016/09/workplace_wellness_programs_are_a_sham.html>

2. "Harvesting the data exhaust stream: Changing the way the insurance game is played." Accenture. Accessed Oct. 1, 2018.
<<https://ins.accenture.com/rs/897-EWH-515/images/Harnessing-the-Data-Exhaust-Stream-POV.pdf>>
3. Hastie, T., & Stuetzle, W. (1989). Principal curves. *Journal of the American Statistical Association*, 84(406), 502-516.
<<https://www.tandfonline.com/doi/abs/10.1080/01621459.1989.10478797>>
4. Hill, Kashmir. "How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did." *Forbes.com*. Feb 16, 2012, 11:02am.
<<https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#4a2556b56668>>
5. "Humana Acquires Anvita Health, a Leading Health Care Analytics Company." *Humana.com*. Wednesday, December 7, 2011 9:00 am EST.
<<https://press.humana.com/press-release/current-releases/humana-acquires-anvita-health-leading-health-care-analytics-company>>
6. "Humana Takes a Bold Step Forward in Population Health with the Formation of Transcend and Transcend." *Humana.com*. Tuesday, March 24, 2015 9:45 am EDT. Insights
<<https://press.humana.com/press-release/current-releases/humana-takes-bold-step-forward-population-health-formation-transcend->>
7. Kramer, Adam D. I.; Guillory, Jamie E.; Hancock, Jeffrey T. "Emotional contagion through social networks." *Proceedings of the National Academy of Sciences* Jun 2014, 111 (24) 8788-8790; DOI: 10.1073/pnas.1320040111
8. Meyer, Robinson. "Everything We Know About Facebook's Secret Mood Manipulation Experiment." *The Atlantic*. June 28, 2014.
<<https://www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebook-secret-mood-manipulation-experiment/373648/>>
9. Sullivan, Gail. "Sheryl Sandberg Not Sorry for Facebook Mood Manipulation Study." *The Washington Post*. July 3, 2014
<https://www.washingtonpost.com/news/morning-mix/wp/2014/07/03/sheryl-sandberg-not-sorry-for-facebook-mood-manipulation-study/?noredirect=on&utm_term=.432c68acb5eb>
10. Williams, Alexander. Personal Communication. 21 Aug. 2018.

VITA:

Cynthia Dickerson

From: Louisville, KY

Education:

University of Louisville – B.S. in Biology (Subcellular Biology & Physiology)

Professional Positions Held:

Teacher - Montessori High School of Lexington, Kentucky

Scholastic and Professional Honors:

1. PhRMA Foundation Predoctoral Informatics Fellowship Recipient: "A New Method for Estimating NOAELs"
2. Graduate Student Assistantship - University of Kentucky College of Pharmacy:
3. Graduate Student Incentive Program Recipient - University of Kentucky College of Pharmacy
4. 2017 Vitality Medical Scholarship Recipient
5. Ashland Inc. Distinguished Lectures and Symposium on Drug Discovery & Development 2016 Elevator Talk Competition First Place Winner: "*Multivariate Analysis: Making Drugs Safe in the Future*"
6. Guaranteed Entrance to Medical School Scholar - University of Louisville
7. Presidential Scholarship - University of Louisville
8. 2011 National Merit Scholar

Publications:

Dickerson, Cynthia; Ensor, Mark; Lodder, Robert A. "*Establishing EDI for a Clinical Trial of a Treatment for Chikungunya.*" In: Shi Y. et al. (eds) Computational Science – ICCS 2018. ICCS 2018. Lecture Notes in Computer Science, vol 10861. Springer, Cham

Cynthia Dickerson, Robert A. Lodder. "*A Novel Statistical Approach to NOAEL: QBEST Applied to Dosing of Ellagic Acid.*" WebmedCentral TOXICOLOGY 2018,WMC005451.

Dickerson, Cynthia; Lodder, Robert A. "*The QNOAEL vs. BMD for Point of Departure.*" BIORXIV/2018/329763.

Dickerson, Cynthia; Lodder, Robert A. "*Calculating the Acceptable Daily Intake or NOAEL with QBEST (QNOAEL).*" Publication in progress.

Dickerson, Cynthia; Lodder, Robert A. "*Validating the Accuracy of QBEST on Synthetic Data: Round the Compass Rose.*" Publication in progress.

Cynthia Dickerson