A Statistical Clinical Decision Support Tool

For Determining Thresholds in Remote Monitoring

Using Predictive Analytics

by

Celeste Fralick

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2013 by the
Graduate Supervisory Committee:

Jitendran Muthuswamy – Co-Chair
Terrance O'Shea – Co-Chair
Jeffrey LaBelle
Vincent Pizziconi
Kimberly Shea

ARIZONA STATE UNIVERSITY

May 2013

ABSTRACT

Statistical process control (SPC) and predictive analytics have been used in industrial manufacturing and design, but up until now have not been applied to threshold data of vital sign monitoring in remote care settings. In this study of 20 elders with COPD and/or CHF, extended months of peak flow monitoring (FEV1) using telemedicine are examined to determine when an earlier or later clinical intervention may have been advised. This study demonstrated that SPC may bring less than a 2.0% increase in clinician workload while providing more robust statistically-derived thresholds than clinician-derived thresholds. Using a random K-fold model, FEV1 output was predictably validated to .80 Generalized R-square, demonstrating the adequate learning of a threshold classifier. Disease severity also impacted the model. Forecasting future FEV1 data points is possible with a complex ARIMA (45, 0, 49), but variation and sources of error require tight control. Validation was above average and encouraging for clinician acceptance. These statistical algorithms provide for the patient's own data to drive reduction in variability and, potentially increase clinician efficiency, improve patient outcome, and cost burden to the health care ecosystem.

DEDICATION

This dissertation is, first and foremost, dedicated to my children, Cameron Null and Hunter Null. Without their unending love, support, and flexibility, I would not have been able to realize my dream. They never let the last nine years be uneventful! I hope I have been a good role model for my boys and my grandchildren.

This dissertation is also dedicated to my husband, Dr. Joe Fralick, who never stopped believing in me, even thirty-five years later.

ACKNOWLEDGMENTS

My sincere thanks to Dr. Jitendran Muthuswamy, last in nine years of advising. I definitely saved the best chair until last and have learned from his wisdom and kind prodding. My goal is to emulate his sound and scientific curiosity in my future endeavors.

My deep appreciation is given to my excellent committee: Dr. Terrance O'Shea, one of Intel's finest brains and action-oriented researchers; Dr. Vince Pizziconi, ASU's very own Einstein; Dr. Kimberly Shea, who brought clinical reality to my work; and Dr. Jeffrey LaBelle, an impressive scholar and all-around good guy whose support, kindness, and even keel are most appreciated.

Thanks as well to many other folks who spent the time to coach me in analytics, critical thinking, and the ways of life throughout nine long years – Doug Busch, John and Kathryn McMahon, Chris and Pam Biendet, Walter Flom, Dr. Mariano Phielipp, Dr. Ron Borg, and Dr. Michael Goul. A special thanks to Dr. Reid Norman for his excellent networking skills and assistance.

A special thanks to Drs. Metin and Yasemin Akay and Dr. Peter Steinmetz. Thank you for saving my life and my sanity. I wouldn't be here without your good hearts and brains.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Chapter 1: Introduction

The transformation of healthcare delivery with information technology is critical to the care of an aging population. The extension of life coupled with the aging populous is a challenging economic, social, governmental, educational, and medical problem (Mechanic 1999) (Dishongh 2010) (Dishman 2010). The global population of those 60+ years old is expected to increase from the 10.0% it was in 2000 to 21.8% in 2050 and to 32.2% in 2100 (Lutz 2008) (Merritt 2007) with chronic care expenditures accounting for 78% of healthcare costs (Anderson 2004). With shortages of hospital beds (Evans 2004), assisted living housing (Zimmerman 2001), doctors, and other health care professionals (PricewaterhouseCoopers Health Research Institute 2007) it is clearly time to improve the efficiency of the delivery of healthcare. One such solution involves telemedicine tools such as remote healthcare monitoring.

In its infancy telemedicine was not embraced by clinicians due to lack of reimbursement from government or private insurance, lack of strong evidence-based data tied to medical outcome, technology challenges, and lack of usability. Forty years later, telemedicine is being identified as a key to solving the impending onslaught of elders (Lee 2000) (United States Government 2012) (Latifi 2000). Ease of technology connectivity, infrastructure and device interoperability has hastened managed care and government agencies to invest in remote healthcare monitoring studies in a variety of clinical settings with promising results. By connecting vital sign monitor “clients” (e.g., blood pressure monitors, glucose monitors, pulse oximeters, weight scales, peak flow meters) to aggregate form factor “managers” (e.g., computers, tablets, USB devices, phones, purpose-built devices) patients can be remotely monitored to reduce office visits and unnecessary hospitalization, increase clinician efficiency, and provide healthier

functioning of the chronically ill (Oeff, Monitoring Multiple Cardiovascular Parameters Using Telemedicine in Patients with Chronic Heard Failure 2005) (Benatar 2003) (Montgomery 1994).

The Intel Health Guide System® is one such personal health system combining an in-home patient device with an online interface, allowing clinicians to monitor and remotely manage care (Intel Corporation 2010). Targeted disease states are chronic obstructive pulmonary disease (COPD), congestive heart failure (CHF), and diabetes. Clinicians develop protocols to select which vital sign to monitor and how often, and are asked to enter a low and high threshold for each vital sign based on clinical knowledge and experience. The physiological output is based on two levels: threshold violations in red and normal in green (see Figure 1). Yellow indicates that the patient has not taken their vital sign measurement per clinician determined schedule. When the patient's captured and transferred data is interpreted by the clinician, he makes evidence-based decisions such as adjusting the patient's monitoring frequency, type and frequency of prescription-based drugs, in-office visits, and/or hospitalization.

While evidence-based decisions are commonplace, information technology can now provide the means to analyze remote patient data for trends and patterns, even predicting future measurements, leading to more robust decisions by the clinician and supplementing the inherent benefits of telemedicine. The common green, yellow, and red classification is overly simplistic, barely qualitative, and not individually quantified. Clinicians can become "alert-weary" and ignore the threshold violation, leading to potentially deadly results. Instead, applying statistical process control theory combined with predictive analytics to individualized patient data may provide the clinician with

additional resources to hasten earlier intervention, predict an episodic event, and eventually improve patient outcomes. This, in turn, may lower overall healthcare cost long-term.

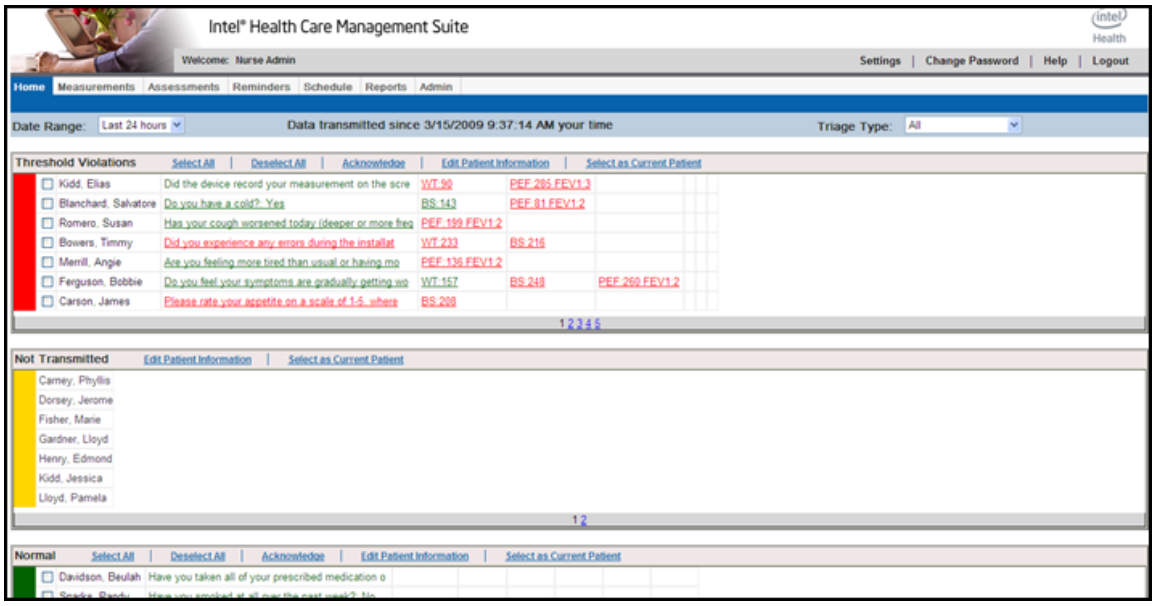


Figure 1. Threshold Output. Threshold output for the Intel Health Care Management Suite ® used in conjunction with the Intel Health Guide System®.

In this study, two thresholds (statistically-derived and clinician-derived) are applied to each FEV1 data point for each individual patient (FEV1 is Forced Expiratory Volume in 1 second, derived from a peak flow meter). This will lead to a four-schema classification: both thresholds were violated (failed), both were not violated (passed), or only one was violated (pass-fail or fail-pass). This four-schema classification corresponds respectively to True Positive, True Negative, False Positive, and False Negative.

The pre-predictive algorithm “statistical process control chart” of each patient’s FEV1 output provides a foundation to compare post-predictive algorithm results, to wit: can an

individual's FEV1 measurement be predicted prior to its occurrence, and is that predicted data point in violation of a statistically- or clinician-derived threshold as determined prior to the application of the predictive algorithm? A summary of the hypotheses are presented In Table 1.

Table 1. Hypotheses.

Hypothesis	H_0	H_1
A	There will not be a statistically valid increase of clinical interventions when statistical thresholds are applied (False Negatives) prior to application of predictive algorithms. FN=FP	There will be a statistically valid (p<.005) increase of clinical interventions when statistical thresholds are applied (False Negatives) prior to application of predictive algorithms. FN≠FP
B	An FEV1 classification cannot be predicted, demonstrating less than an .80 Generalized R-square.	An FEV1 classification can be predicted demonstrating ≥ .80 Generalized R-square.
C	The pre- and post-predicted data point will not demonstrate a statistical difference in zone designation.	The pre- and post-predicted data point will be statistically different (p<.005) in zone designation.
D	FEV1 cannot be forecasted, demonstrating less than .60 R-square	FEV1 can be forecasted, demonstrating ≥.60 R-square

It is anticipated that this study will support the postulate that, with more robust and personalized thresholds, clinicians may intercede in a timelier manner and subsequently make improved decisions about the care of elderly patients with COPD and/or CHF. The clinician can respond to statistically-derived thresholds (i.e., violations) rather than respond to random spikes that may not be a reflection of a true change in the physiological condition of the individual but merely a result of normal individual variation. Prevention is intended, as opposed to crisis management post-measurement, and personalized medicine is enabled by statistical process control for each patient.

The Shewart-Western Electric analyses will be applied in this study to determine the statistically-derived threshold. Statistical process control (e.g., "SPC") is rooted in a statistical test, developed by Shewart (Pyzdek 2003), which makes use of 3 sigma control limits as a criterion for applying error of the first kind (looking for assignable causes when no such causes exist, or Type I errors), and error of the second kind (not looking for assignable causes when such causes do exist or Type II errors). Combined with Western Electric Rules (Western Electric Company 1956), which tests eight Rules[1] whether a statistically improbable measurement has occurred or not, they will determine if an individual measurement has violated the statistically-derived threshold (also known as "out of control"). These decision Rules will also indicate why the threshold has been violated, based on the probability of that individual measurement occurring.

The first specific aim of this study is to perform a Shewart-Western Electric analyses to examine statistically-derived threshold violations on a single vital sign and compare to a clinician-derived threshold violations for individual COPD and/or CHF elder patients. This will be accomplished by analyzing FEV1 from a peak flow device captured remotely from an Intel Health Guide System®. Analyses will include determination of each patient's individual FEV1 data point as to whether it passed and/or failed a statistically- or clinician-derived threshold (i.e., test the four schema classification), artifact contamination, and a confusion matrix. Patient data will be de-identified, and sources of error will be reported. This will test the first hypothesis (hypothesis A) of Table 1 that there will be a statistically significant (p<.005) increase of clinical interventions when statistical thresholds are applied (False Negatives), prior to applying predictive algorithms.

[1] The term "Rules" used throughout this document shall refer to the Western Electric Rules.

The second specific aim of this study is to apply predictive algorithms to individual FEV1 measurements and compare the statistically- and clinician-derived threshold violations for the same population as described above. Again, analyses will determine whether each patient's individual FEV1 data points passed or failed a statistically- and/or clinician-derived threshold (i.e., the four schema classification), and subsequently compared to pre-predictive algorithm data point classification. Algorithms will be explored that can best detect (≥.80 Generalized R-square) subtle changes in signal level to demonstrate learning of the threshold classifications. Vital sign captured remotely from an Intel Health Guide System® is as above. Data acquisition, data preprocessing, feature selection, feature extraction, classifier design, and data training will be employed and optimized for the vital sign. Patient data will be de-identified and sources of error will be reported. Table 2 summarizes the four-schema classification. This second aim will test the second hypothesis (hypothesis B) in Table 1 that the FEV1 classification can be predicted demonstrating ≥ .80 Generalized R-square. It will also test the third hypothesis (hypothesis C) in Table 1 that the pre- and post-predicted data point will be statistically different ($p<.005$).

Table 2. Threshold Violations and Interventions. “TP”=True Positive; “FN”=False Negative; “FP”= False Positive; “TN”=True Negative.

Intervention	Test	State	Statistically Derived	State	Clinician Derived	Combined State
Intervened	TP	1	Threshold violation	1	Threshold violation	1,1
Did not intervene but should have (statistically)	FN	1	Threshold violation	0	No threshold violation	1,0
Intervened but shouldn't have (statistically)	FP	0	No threshold violation	1	Threshold violation	0,1
No intervention necessary	TN	0	No threshold violation	0	No threshold violation	0,0

The third specific aim of this study is to explore the ability to forecast an FEV1 value in the future that has not yet been actually performed by the patient. This will test the fourth hypothesis (hypothesis D) that FEV1 can be forecasted with ≥ .60 R-square.

The term “analytics” has been increasing with frequency in many domains, due in part to the increase of hand-held devices, compute power, and the ability to be connected 24/7 (Addicam 2012). With this burgeoning field, the intermingling of domain-specific terms can be a source of confusion. For this study, the term “prediction” will be applied to predicting (or “estimating”) by a classifier using learning algorithms, while the term “forecast” will be applied to identifying a future value(s) in time.

Lastly, the general approach of this study is 1) Alpha: Proof of Concept; 2) Beta: Developing and Testing with One Patient; and 3) Final Construct. The purpose of the

Alpha stage is to determine if statistical process control can be applied in a field setting and the student to become comfortable with applying SPC to a vital sign. The purpose of the Beta stage is to empirically explore various analytics and the analytics process with one patient. Neither the Alpha nor the Beta stage will be specifically considered in addressing the specific aims and their respective hypotheses, and will be used solely for exploring potential data algorithms and nuances of the data itself. The purpose of the Final Construct stage is to utilize all patients to address the specific aims and their respective hypotheses. Methods and Results sections will include Alpha, Beta, and Final Construct.

Chapter 2: Background Literature

This study is unique in that it strives to bring together telemedicine, a remote monitoring device, statistical process control, and predictive analytics. The area of study is innovative and significant as the use of remote monitoring grows and learning from other fields can be introduced to quickly evolve this industry. Increasing amounts of data that are readily available speaks to the urgent need for clinicians to derive useful information from that data to improve their patients' well-being. A review of the various fields that are brought together for this study are presented below, in Section 1) Telemedicine, 2) The Intel Health Guide System®, 3) Statistical Process Control, and 4) Predictive Analytics. For completeness, a brief review of COPD and CHF is also provided in Section 5.

Section 1. Telemedicine

Telemedicine is generally defined as the use of medical information from one site to another via electronic communication to improve one's health status, and the addition of medical devices to monitor typical vital signs is defined as remote care or remote patient monitoring (American Telemedicine Association 2012) (Kyriacou 2003). It is estimated that the use of telemedicine and remote care monitoring will double in the next decade as those born between 1946-1964 turn 65 and require increased health care to deal with problems associated with aging, such as COPD and CHF (Zarocostas 2010) (Myers 2006). Health Data Management reported in 2009 that within 5 years the U.S. telemedicine market will have reached $3.6B in revenue as health care costs increase, wireless technology advances, and data transmission speeds improve (Health Data Management 2009). WinterGreen Research reported that the telemedicine dedicated device and software markets will grow from the $736M in 2011 to $2.5B in 2014 (WinterGreen Research 2012), with the ubiquitous use of computer tablets, Additional data indicates that telemedicine will have a global 19% compounded annual growth rate (CAGR) between 2010 and 2014, due in part to the growing demands in Europe, increasing online healthcare services and telesurgery (Healthcare IT News 2011).

Telemedicine has often proved to be beneficial to health outcomes. The remuneration of improved outcome, reduced costs, improved efficiency, and patient care are emerging as telemedicine becomes more widely used and supported by payers, patients, and clinicians. In a 2005 study of 24 patients with CHF, the use of remote care monitoring with vital sign instrumentation reduced hospital admissions by 62% (Oeff 2005). In another study, 216 randomized CHF patients were provided either home nurse or remote care monitoring of their weight, blood pressure, heart rate, and SpO2 levels.

After 3 months, the remote care patients had fewer CHF readmissions, less hospital charges (even up to 12 months later), and improved perceived quality of life (Benatar 2003). A recent four month telemonitoring study of 57 subjects with advanced (i.e., severe or very severe) COPD found no improvement in FEV1 or oxygen saturation, but significant lowering of blood pressure, the number of prescribed antibiotics and steroids, and clinical consultations (Jensen 2012).

Many other individual studies of telemedicine have proven beneficial and cost-effective. Pare (Pare 2006) demonstrated a $355 savings per COPD patient over a six month period within a group of 29 patients, including a control group, and indicated the savings would be more if the technology wouldn't have been so costly. Dale (Dale 2003) reported a 50% decrease in rates of hospital admission for 55 patients with COPD. Haesum (Haesum 2012) reported a TELEKAT (Telehomecare, Chronic Patients, and the Integrated Healthcare System, Denmark) study of 111 severe and very severe COPD elder patients where a cost-utilization study was performed. While the study does show an incremental improvement in cost-effectiveness, the conclusion is even better at stating sources of error in deriving such information and the common sources of variability.

This inherent variability in patients, disease, and process is reflective of a "review of reviews" of the effectiveness of telemedicine (Ekeland 2010). In a comprehensive examination of telemedicine reviews since 2005 and following rigorous boundary conditions, Ekeland concluded that a third of the reviews determined that "telemedicine is promising", about a third that "evidence is limited and inconsistent", and about a final third that "telemedicine is effective". Ekeland further concluded that the telemedicine

field was dynamic, complex, and evolving, and that the knowledge and understanding about telemedicine costs were inconsistent or lacking. Indeed, Ekeland highlights that the intricacies of the social and organizational costs of telemedicine are not well developed, and many studies still need to be done (e.g., gender studies, service delivery issues).

Eventually, acceptance of telemedicine may occur after widespread acknowledgement (and payment) by insurance companies. Telemedicine may reap the benefits of the October 2013 implementation of ICD-10-CM/PCS (International Classification of Diseases, 10th Edition, Clinical Modification/Procedure Coding System). The new coding system is significantly more detailed and consistent with today's medical practice than the 30-year old ICD-9. It includes the code, S9110, for telemonitoring of the patient in their home (AMA 2012), which may enable further adoption.

Vital sign measurements currently afforded to remote care patients typically include regulated medical devices such as blood pressure monitors, weight scales, peak flows, glucose monitors, and pulse oximeters. Wireless (e.g., 3G, 4G, WiFi, Bluetooth®) or tethered (e.g., USB) to a manager (e.g., computer, netbook, smart phone), the results of the vital sign monitoring are typically stored and forwarded as data packets routed through a hosting service or website to a clinician's computer via Broadband, mobile service, or telephone cable. Some can provide continuously streaming data for ECGs or SpO2, but data transfer rates have been limiting, and clinicians fall into data overload (Kyriacou 2003) (Khoor 2003).

Section 2. The Intel Health Guide System®

The Intel Health Guide System® (HGS, Model PHS 6000) is a tabletop telemedicine unit that stores, collects, and transmits data from compatible wireless or tethered (e.g., USB) physiological monitoring devices selected by the healthcare professional (Intel Corporation 2010). Information collected is stored, displayed on the screen, and transmitted via Broadband through a remote-site server to a clinician software and hardware interface, thus allowing the patient to be remotely located from the healthcare professional. The screen displays instructions, reminders, and education selected and entered by the clinician for the patient to read. Calendaring and managing tasks are available on the personal health system for the patient to use. The system can also receive email and establish video conference links with caregivers and health care providers. A software and hardware interface enables technicians to remotely complete routine maintenance and troubleshooting via the Broadband connection.

The system is contained in a small plastic enclosure with a touch screen, video camera with privacy screen, microphones, and a reminder light mounted into the top of the case. On the back of the unit is a power socket, two medical device sockets for connection to specific physiological monitors, a headphone socket, a Broadband internet socket for connection to a Broadband cable, and a phone socket for connection to a standard phone line. It is an FDA Class 2 regulated medical product, CE marked, and available in at least ten different geographies globally.

The Intel Health Care Management Suite® (HCMS) is a web-based patient monitoring software application for clinicians to be used in conjunction with their patient's Health Guide System®. Clinicians have the ability to video conference with patients, present

targeted multi-media educational content, develop personal care plans to manage chronic disease, access specific patient data, and prioritize patients based on "management by exception". Clinicians can select specific threshold values (high and/or low) for a particular vital sign for a specific patient, and outliers are automatically flagged per Table 3. Patient measurement data can be reviewed in a graphic format to highlight trending results, as shown previously in Figure 1.

Table 3. The Intel HCMS® Patient Prioritization Scheme.

Color	Indication
Red	Patient's data has been received and the vital sign is above the threshold set by the clinician
Yellow	Patient's data has not been received
Green	Patient's data has been received the vital sign is below the threshold set by the clinician

While adequate, the clinician's selection of threshold values can be improved from simply a clinical judgment criteria to a more robust and statistically-derived criteria. Numerous studies have demonstrated that the application of statistics can reveal previously hidden information that may evolve the field of patient care. Heart rate variability has been well studied statistically to quantitatively assess cardiovascular regulation (Barbieri 2005) (Akselrod 1981) (Malic 2006) (Kleiger 1987). The biosignal decomposition of pulse oximetry, blood pressure, galvanic skin response, and skin temperature have also been studied using various statistical methods, such as independent component analysis, Kalman filter, multivariate autoregressive analysis, and K-nearest neighbor (Mower 2007) (Li 2008) (McNames 2008) (Stetson 2004) (McNames 2006). In telemedicine, however, the clinician relies on the output of a regulated medical device as he would in an office setting, and does not have an

opportunity to elegantly decompose the signals to ascertain an improved course of action in patient care. Until such time that medical device manufacturers, regulators, statisticians, clinicians, and payers are accepting of the biosignal decomposition, the clinician is left with the current clinically approved and regulated output from the medical devices (Table 4).

Table 4. Medical Devices. Example of medical devices with units, biosignal, and normal range. ("Normal range" is dependent on standard used, weight, age, gender, height, race, fitness, underlying genetics, and/or illness.)

Medical Device	Units	Biosignal	Normal Range
Pulse oximeter	bpm SpO_2 in %	Heart rate Estimation of the saturation of oxygen; the relationship of oxygen-bound hemoglobin compared to oxygen-unbound hemoglobin	60-100 bpm 95 – 100 %
Blood pressure	mmHg	Force per unit area that the blood exerts on the walls of blood vessels. Varies both in time and distance along the circulatory system.	120/80 mmHg
Peak flow	L L/m	FEV1: Forced expiratory volume in 1 second - this is the volume of air which can be forcibly exhaled from the lungs in the first second of a forced expiratory maneuver. PEF: Peak expiratory flow is the fastest speed air can be blown out of the lungs after inhalation, measured in liters per minute	Normal is 75-80% of predicted value. Typical absolute value 1.1 – 6.3 L 390-740 L/m
Weight scale	Lbs or kg	Weight	Age/gender/height dependent

Thus, in an effort to drive more robust information that the clinician can draw from monitoring simple vital signs, the application of statistical process control and predictive analytics from the industrial sector is considered. Coupled with clinical judgment, these tools can provide a translational conduit from combining key medical and industrial tools to an improved decision on patient care, and perhaps improved patient outcomes.

Section 3. Statistical Process Control

Statistical process control (SPC) is defined as the use of valid analytical methods to identify the existence of special causes of variation in a process (Pyzdek 2003). Ubiquitously used in industry since the 1950s, improvements are made by identifying process variation, identifying its cause, selecting improvement methods and removing the causes. Variation naturally occurs due to common causes, but key to measurable process improvement is to identify when special causes of variation occur. In this manner, reaction is based on outliers caused by special causes and the natural variation of the process (common causes) can be explored "off-line". To achieve this, control charts can be used with control limits (e.g., thresholds) to predict the natural variation of the process due to common and special causes. The term "control" was best described by Shewart in 1931 (Pyzdek 2003):

> *"*A phenomenon will be said to be controlled when, through the use of past experience, we can predict, at least within limits, how the phenomenon may be expected to vary in the future. Here it is understood that prediction within limits means that we can state, at least approximately, the probability that the observed phenomenon will fall within the given limits."

Testing the data on a control chart for unnatural patterns due to special causes can be done statistically based on probability distributions. Examination of an "x-bar" or mean chart[2] in a normal distribution yields the probabilities of a process in control as seen in Table 5 (Western Electric Company 1956).

Table 5. Probabilities and Zones. Used in tests for unnatural patterns (special causes) (Western Electric Company 1956).

Area of Chart	Probability	Chart Designation	Zone
Above outer third	.00135	Above 3σ	
Outer third	.02135	Above 2σ	Zone A
Middle third	.1360	Above 1σ	Zone B
Inner third	.3413	Above centerline, x-bar	Zone C
Inner third	.3413	Below centerline, x-bar	
Middle third	.1360	Below 1σ	Zone B
Outer third	.02135	Below 2σ	Zone A
Below outer third	.00135	Below 3σ	

In applying these probabilities, three characteristics of a natural pattern emerge: 1) most points are near the centerline, 2) a few points are near the control limits, and 3) no points (or only a rare point) are beyond the control limits. Unnatural patterns, or those caused by special causes, always involve the absence of one of these characteristics. The "Western Electric Rules" (Western Electric Company 1956) utilize these probabilities to test for instability of a process and whether the system is actually changing or not (Table 6). These Rules can easily be applied to a trend chart automatically so the reader is notified when a statistical threshold has been breached, as shown in Figure 2 by a red circle and the annotation of the Rule that failed (i.e., "1" and "5", see further discussion below).

[2] Other control charts exist for asymmetrical distributions, such as p (percentage - Poisson, Bionomial), c (defects – Poisson), and u charts (average number of defects – Poisson).

Table 6. Probability of Point. Example of the probability of the point occurring in a specific zone.

Point Location that Violates Statistical Thresholds	Probability of the Point Occurring in this Zone
Single point outside of the 3σ limit	.0013
2 out of 3 successive points fall in Zone A or beyond	.0015
4 out of 5 points fall in Zone B or beyond	.0027
8 in a row fall in Zone C or beyond on one side	.0039

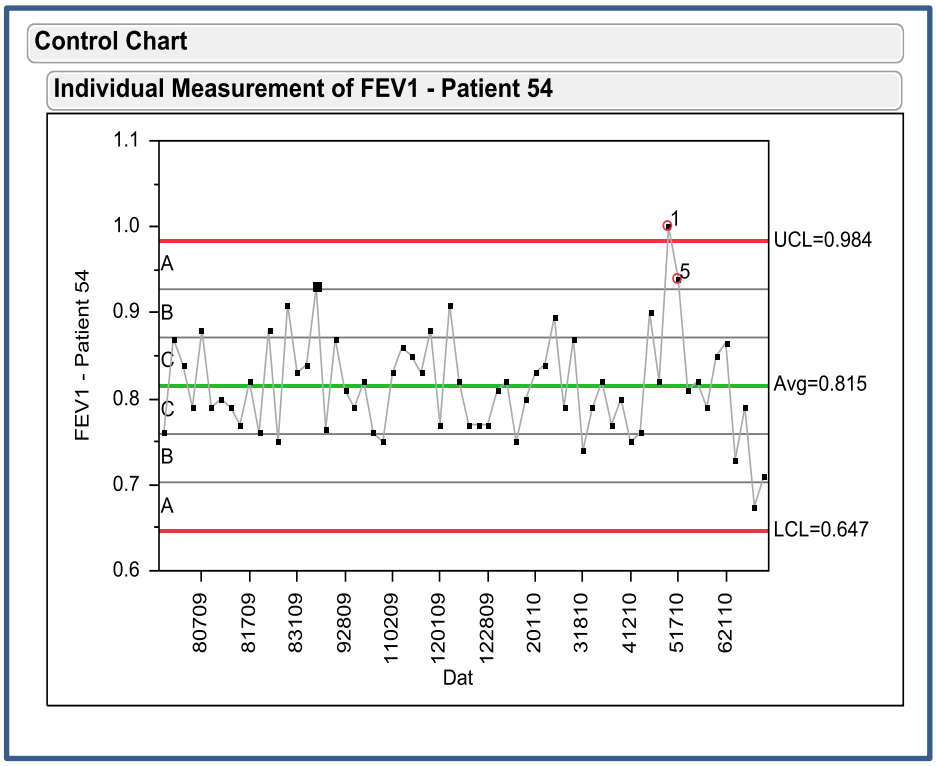


Figure 2. SPC Chart. Example of Statistical Process Control chart (x-bar) with Western Electric Rules applied on a patient. The two points circled in red labeled “1” and “5” have failed those specific Rules (per JMP® Pro 9.0.3, see Table 7). UCL = Upper Control Limit = +3 sigma; LCL = Lower Control Limit = -3 sigma standard deviation.

Over the years, the Western Electric Rules have been supplemented with additional Rules for trending, with equally improbability of occurrence as the original Western Electric. While many statistical experts have been cited for additional Rules (Quinn-Curtis, Inc 2012), the supplemental Rules are typically attributed to Dr. Douglas Montgomery (Montgomery 2004). Typical SPC software (e.g., JMP®, MiniTab®) includes all those listed in Table 7 as the full complement of the Western Electric Rules. (Note: in some software, the Rules may have different numbers, but their point location and detection remain the same.)

Table 7. Western Electric Rules (as used by JMP® Pro 9.0.3 statistical software).

Rule	Point Location	Detection
Rule 1	One point beyond Zone A	Detects a shift in the mean, an increase in the standard deviation, or a single aberration in the process.
Rule 2	Nine points in a row in a single (upper or lower) side of Zone C or beyond	Detects a shift in the process mean.
Rule 3	Six points in a row steadily increasing or decreasing	Detects a trend or drift in the process mean. Small trends will be signaled by this test before Test 1.
Rule 4	Fourteen points in a row alternating up and down	Detects systematic effects such as two alternately used machines, vendors, or operators.
Rule 5	Two out of three points in a row in Zone A or beyond and the point itself is in Zone A or beyond.	Detects a shift in the process average or increase in the standard deviation. Any two out of three points provide a positive test.
Rule 6	Four out of five points in a row in Zone B or beyond and the point itself is in Zone B or beyond.	Detects a shift in the process mean. Any four out of five points provide a positive test.
Rule 7	Fifteen points in a row in Zone C, above and below the center line	Detects stratification of subgroups when the observations in a single subgroup come from various sources with different means.
Rule 8	Eight points in a row on both sides of the center line with none in Zones C	Detects stratification of subgroups when the observations in one subgroup come from a single source, but subgroups come from different sources with different means.

Statistical process control, and other quality improvement tools such as Six Sigma®, theory of constraints, SERVQUAL (Service Quality Framework) and Lean® have routinely been applied in healthcare settings to improve efficiencies and patient care, but

thus far have been limited to in-house processes such as throughput, cycle times, percent work completed, hospital trauma mortality, and service management (Kotagal 2009) (Fralick 2006) (Clark 1998) (Young 2004). The application of Statistical Process Control to monitoring vital signs has been inadequate with only cursory investigation in the literature. Bansal (Bansal 2008) reported variances of two thermometers with no statistical inferences for the patients themselves. Yang ((Yang 2006) (Yang 2008) utilized cusum (cumulative sum control) charts to detect change in blood pressure with excellent results, although the population was anesthetized surgical patients in controlled situations rather than remote settings.

The emergence of smartphones and tablets has buoyed the personal medical applications that can be downloaded and trended. HeartWise (Ollapp, SwEng LLC 2012), for example, trends blood pressure, pulse, and weight, but the statistics are limited to mean, trending over time, and box charts. Real-time rhythm and beat classification of electrocardiogram (EKG) was reported by Rodriguez (Rodriguez 2005) (J. Rodriguez 2005), and a number of EKG monitors are available that can have results uploaded into a personal computer or tablet, or have specific analyzer software available for the clinician such as atrial fibrillation, myocardial infarction, or QT interval. None, thus far, have attempted to apply statistical process control to cardiac or other vital sign measurements.

A recent study by Hokan proposed a low-cost adaptive tracking system for COPD patients in a home environment (Hokan 2013). Using a two-layer five-parameter back propagation artificial neural network learning cycle, Hokan was able to successfully demonstrate the classification of COPD patterns into normal, obstructive (including

asthma) and restrictive behavior. It did not provide prediction or utilize SPC. It was unclear why a poor user interface on a mobile device was adequate, and no data supporting self-management or low-cost was evident in the study. However, the study did demonstrate the evolution of tracking vital signs at home via a mobile device, as well as the use of a learning algorithm of a neural network using tansig in the activation nodes; other transformations are available, as will be demonstrated in this study.

Lastly, the increasing use of mobile phones for managing health provides a platform to develop applications that are growing at a rapid pace. Total 2011 revenues generated by mobile apps, health devices, and related services grew 7-fold to $718M. While the market is young, mobile health apps are expected to continue to reach billions by 2016 (Dolan 2012). Health applications for the Apple® iPhone® alone were estimated at 13,000 for 2012, with over 20% targeted for chronic or cardio states (Dolan 2011). Most mobile-phone apps are educational while some are trending/tracking or depicting patterns (e.g., iBGStar glucose meter). It is inevitable that with the growing “internet of things” and petabytes of cloud data available, increased personalization will follow. The use of statistical process control, therefore, is a novel solution for health personalization in telemedicine or mobile devices of the future.

Section 4. Predictive Analytics

Predictive analytics examines *a priori* knowledge of data and extracts variables to predict data points or trends in the future. Utilizing regression methods, machine learning, or a host of different statistical tools, predictive analytics have become an elegant method with which to predict the future with a defined accuracy based on the models applied. Pattern recognition, for example, is a statistical tool often used in neural network

applications in machine learning. It is defined as "the act of taking in raw data and taking an action based on the category of the pattern" (Duda 2001). Using *a priori* pattern knowledge of the vital sign or the intrinsic pattern regularity itself, models can be created that may augment the clinician's decision to better assess the patient's trending. The oft-used term "dynamic" implies that the pattern is being re-assessed with each new data point entered by the patient, and counteracts intrinsic data decay of a predictive model.

The emergence of predictive analytic tools in new applications has been aided by the increase in compute processing. From microprocessor initial clock speeds of 108 KHz with 2300 transistors in 1971 to 2.9 GHz with 1.4B transistors in 2012 (Intel Corporation 2012), the ability to quickly compute challenging algorithms is becoming mainstay in literature as well as in consumer goods. On-board predictive analytics can be seen in automotive, appliances, building maintenance, industrial controls, speech recognition, gaming consoles, and communication. The use of decision support tools that use pattern recognition, such as Computer Aided Diagnosis (CADx) for radiological images, have become commonplace in hospitals and clinics. CADx assists clinician's interpretation of disease detection (such as breast cancer). Telemedicine, however, has been slow to adopt these tools, likely due to the nascent technology of remote monitoring; interoperability, integrity, privacy/security and market penetration of the technology are the primary focus while decision support tools tend to be lagging in implementation (Helal 2009). Requirements for evidence-based medicine, payer reimbursement models, and a challenging regulatory environment have also impeded the adoption of predictive analytics in remote monitoring.

However, with the congruence of increasing compute power, aging populaces, rising health care costs, remote monitoring for chronic disease, interoperability and connectivity demands by users, predictive analytics are beginning to be pulled from other applications and investigated for use in vital sign monitoring.

As mentioned above in Background, Section 2, The Intel Health Guide System®, few investigators have up-leveled pattern recognition in predictive analytics from the decomposition of the biosignal to the actual output that a clinician sees. This may be due to Yang's (Yang 2008) observation that artifact contamination and the clinician's intuitive observations are not addressed; Garg (Garg 2010) echoed similar sentiments regarding noisy data with varying space and time attributes. Helal (Helal 2009) investigated diabetes patients unobtrusively in their homes ("Smart Homes") while collecting behavioral, blood pressure, and glucose data. He applied Hidden Markov Model classification, achieving 98% recognition accuracy of activities and chewing. A review of "Smart" wearables for remote health monitoring by Lymberis (Lymberis 2003) indicated that embedded medical decisions have relied thus far on fuzzy logic or neural network models. Lisetti (Lisetti 2004) performed an extensive literature survey in the study of physiological signals recognizing human emotions: a number of statistical tools were applied in studies such as Analysis of Variance (ANOVA) and Multivariate Analysis of Variance (MANOVA). In other studies Lisetti reported, Hidden Markov Models, Sequential Floating Forward Search (SFFS), Fisher Projection (FP), Principal Component Analysis (PCA), and Discriminate Function Analysis (DFA) were applied. No common analytic tool was used, however, suggesting that there is a wide range of applications and ramifications for applying a particular statistic. Accuracy and validation of these tools were not reported in his review.

As Pantelopoulos (Pantelopoulos 2010) indicates, a large amount of multidimensional data will have to be interpreted by clinicians in order to detect changes and trends. Another area of focus has been the use and theory of decision support systems (Western Electric Company 1956) (Falas 2008) (Fodor 2010) (Eren 2008) which are typically computerized expert systems that utilize artificial intelligence and machine learning to supplement a clinician's decisions. At its most basic level, decision support systems can be suggestions based on simple algorithms. Eren (Eren 2008) reports that there are two types of decision support systems: rules-based or expert systems (including probabilistic and cognitive models). While this study is not intended to be an "elegant" decision support system, it will approach the data output with predictive analytic tools to challenge a clinician's threshold value selection. Further integration into a more complex system is out of scope for this study.

The foundation of any predictive analytic process typically includes 1) data acquisition from the sensor 2) feature generation and pre-processing, 2) feature extraction and selection, 4) classifier design, and 5) system evaluation using a training set (Theodoridis 2009). As shown in Figure 3, this end-to-end process is often iterative and each step of the process requires complex trial and error of the algorithms.

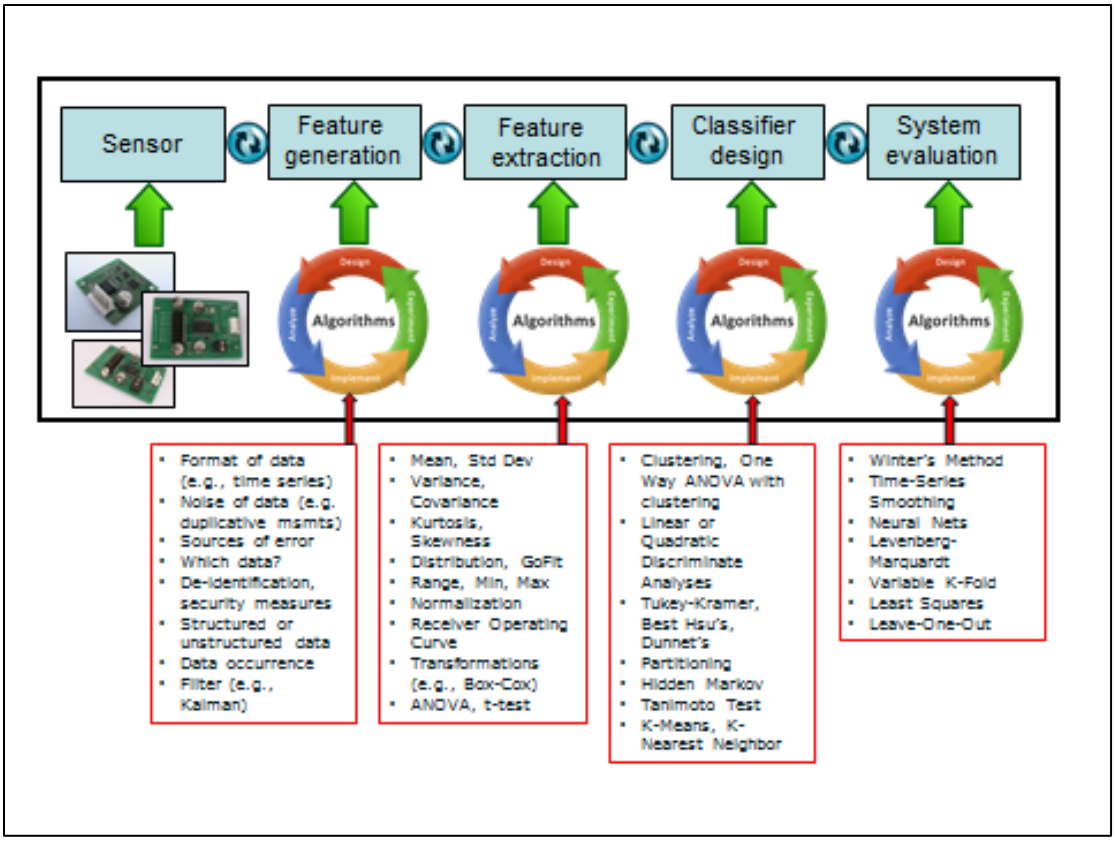


Figure 3. Predictive Analytics Process. Typical process flow (Theodoridis 2009) for predictive analytics, with algorithm examples and iterative symbols added by student.

Section 5. COPD and CHF

Chronic obstructive pulmonary disease (COPD) is a chronic lung disease that is progressive and irreversible. The damaged airways, which can no longer efficiently process oxygen, can result in wheezing, coughing, and shortness of breath. Chronic bronchitis and emphysema are forms of COPD. Congestive heart failure (CHF) is progressively debilitating and often leads to death. The heart can no longer pump blood efficiently and effectively, and the lungs and lower extremities fill with fluid as circulation is challenged. These two diseases can occur together, particularly in late stage, and are

often co-morbid with other diseases such as pneumonia, arrhythmia, and hypertension (Clinical Cases and Images 2012) (Growth House, Inc. 2012).

Clinical management of COPD includes removal of lung irritants, such as smoke, pollutants, and inadvertent gas exposure. Treatment includes bronchodilators (anticholinergic and beta-adrenergic agonist drugs), anti-inflammatory agents including steroids, and supplemental oxygen, if needed (Net Wellness 2008) (The Merck Manual Home Health Handbook 2008). Management of fluids, exercise, and irritants help prevent flare-ups, and antibiotics are typically prescribed if bacterial infection is suspected.

Clinical management of CHF includes diuretics, anticoagulants, beta blockers, and ACE (angiotensin converting enzyme) inhibitors. Finding the specific cause of the heart failure (such as hypertension or valve defect) can guide treatment (American Heart Association 2012).

AdvaMed's report (Stachura 2007) on remote monitoring of COPD and CHF indicates that demonstrated health benefits occurred in terms of reduced hospitalization days, reduced clinic visits, enhanced quality of life, and satisfaction with technology. In addition, cost benefits were demonstrated for patients, home care agencies, and the health care system.

Remote monitoring of vital signs for patients with COPD and CHF are currently limited by peripheral devices approved by regulatory agencies for telemedicine as well as the ecosystem to transfer the reading to the clinician without error. With a weight scale,

blood pressure monitor, peak flow device, and pulse oximeter, however, the clinician can monitor key physiological parameters for change (Table 8), and treat the patient accordingly.

Table 8. Physiological Parameters. Changes in key physiological parameters for COPD and CHF. See Table 4 for descriptions of each parameter.

Disease	Weight	Blood pressure	FEV1	PEF	SpO2	Pulse
CHF	Increases	Increases	Decreases	Decreases	Decreases	Increases/ Arrhythmia
COPD	Increases in severe cases	Increases	Decreases	Decreases	Decreases	Increases/ Arrhythmia

Chapter 3. Methodology

Section 1. Alpha Test: Proof of Concept (PoC)

The first opportunity to apply Shewart-Western Electric analyses will be to examine four field failures of a peak flow device (brand and model not specified due to corporate respect), attached as a peripheral to the HGS. It is used for remote monitoring of elders (>70 yrs) with a clinical diagnosis of COPD and/or CHF (the HCMS and HGS requires a clinician's prescription). Patients will be de-identified, and no patient contact occurred. Each patient will have an FEV1 (Forced Expiratory Volume in 1 second from 0.01 to 9.99 L) and PEF (Peak Expiratory Flow rate from 50 to 900 L/min.)[3]. Data will be analyzed (using JMP® 9.0.3) within patient only, and sorted by FEV1 and PEF.

[3] FEV1 and PEF are indirect measurements of airway capability. While they can be correlated, that correlation can be offset by about 30%. No correlation between FEV1 and PEF was attempted herein.

This PoC will determine if there is abnormal variability of normal peak flow readings in patients who reported unusually and often physiological improbable high peak flow readings. By applying statistical threshold violations using the Shewart-Western Electric analyses, this initial investigation will demonstrate the field efficacy of applying SPC. Results will assist in determining whether the device will be withdrawn from patients and stop usage, or allow the patients to continue using the device until long-term failure is seen. This preliminary PoC will serve as an alpha test of SPC for remote monitoring and aid in subsequent research design, source of errors, and data analysis in this overall study. It will not address specific aims or hypotheses.

Section 2. Beta Test: Developing and Testing with One Patient

Patient #1, from the population described above, will be selected to develop and test the Predictive Analytics Process described in Figure 3. This patient will be selected 1) to maintain continuity from the alpha test, and 2) because the patient's data detected a change in mean as denoted by the rule violations.

Various statistics and algorithms throughout the Predictive Analytics Process (Figure 3) will be explored in order to optimize solutions for testing the hypotheses described in Chapter 1. This will provide the basis for testing additional patients. The beta test will not address specific aims or hypotheses.

Section 3. Final Construct: Sensor, Feature Generation and Pre-processing

Following the alpha and beta testing, additional patients' data will be investigated. The total population data set will be captured from the HGS and HGMS and consist of twenty unknown elder patients (>70 yrs, specific age not given) with a history of COPD and/or

CHF at various states of disease. (Four of the twenty patients come from the Alpha Test, above). Patients will have been monitored by clinicians and provided feed-forward sensor data of FEV1 from peak flow devices (brand and model not specified due to corporate request) at a frequency pre-determined by the clinician. This frequency is typically on a daily basis, but the data may be captured intermittently (i.e., occasionally skip a day or two) depending on the patient's compliance level; monitoring frequency is different for every patient. Patient data will be de-identified, limited to that provided, and created as a time series. No disease outcome data, interventions, or subsequent monitoring will be available. (Patients' weight, blood pressure, SpO2, and heart rate were captured but led to too small of sample for analysis and correlation.)

Additional pre-processing will consist of filtering for double entry; some device usability nuances can cause double entry in HCMS, causing integrity issues in analysis (consistent with Yang's observation (Yang 2008)). Data will be filtered for sensor entries made within ten minutes. These data will be averaged, and/or obvious outliers removed. If threshold data is not provided by the clinician, GOLD (Global Initiative for Chronic Obstructive Lung Disease 2013) standards will be sought and applied, and/or an independent clinician will determine a theoretical threshold. Consideration of a malfunctioning device may be taken into consideration if physiological implications are nonsensical; obvious outliers will be purged.

Section 4. Final Construct: Feature Extraction and Selection

The feature extraction and selection will consist of fundamental statistical analyses within each patient's FEV1 output: mean, standard deviation, 3-sigma control limits,

skewness, kurtosis, variance, covariance, and distribution (should any subsequent analytic tool be highly dependent on a specific distribution).

Data will then be subjected to Shewart-Western Electric analyses, utilizing the Rules in Table 7. These statistically-derived thresholds will be compared to the clinician-derived thresholds. Per Table 2 above, the four schemas (TP, FN, FP and TN) will be classified for each data point within each patient. This classification of each data point is the basis for future analyses and provides data for the classifier design, below.

The Shewart-Western Electric Analysis and subsequent comparison between [TP, TN] and [FN, FP] utilizing cumulative *z* statistic will satisfy the first specific aim and Hypothesis A, and provide foundation for the second specific aim. All tests at $p<0.005$ will be considered statistically significant.

Section 5. Final Construct: Classifier Design

Many options exist for classifier designs within predictive analytics. However, due to the Shewart-Western Electric analyses and subsequent categorization of each data point, the data is, by design, already classified into four distinct schemas as noted above.

Section 6. Final Construct: System Evaluation

Models will be explored to find the best fit for predicted responses to satisfy the second specific aim. Since data is limited to that provided, later data points of the sensor will attempted to be removed and used for comparison between predicted and actual, based on statistically- or clinician-derived thresholds. Because of the inherent time-series nature of the data, regression models, additive models (e.g., Winters Method), or neural

network algorithms may be applicable based on model fit. The predicted data point based on the model will be compared to the actual data point and its inherent four schema classifier (TP, TN, FP, FN) utilizing statistical tools. All tests with a Generalized R-squared >.80 will be considered significant. (R-squared reflects the amount of variation in a model, and ranges from 0-1.0. It reflects how well the regression line fits the data, with 1.0 indicating the regression line fitting the data very well. It is used typically in predictive models to assess how much variation there will be in a future point, and how well the model fits the data.) As noted in the Introduction above, the term "predictive", while used similarly in the nascent industry of analytics, is actually learning (inclusive of training and validation) the four schema threshold classifications.

Training and validation of the data will initially be performed using the Levenberg-Marquardt algorithm, an iterative back propagation technique that locates the minimum of a function that is expressed as the sum of squares of nonlinear functions. Other training and validation algorithms may be considered depending on results of the R-squared. Training and validation are expected successful at .80 Generalized R-squared. This will address the second specific aim and its hypotheses B and C.

An attempt to forecast a future point in time that has not yet been performed by the patient will be made using various analytic tools. It will address the third specific aim and its hypothesis D.

Section 7. Validation with Clinicians: Ground Truth

Five clinicians will be interviewed (face-to-face) to ascertain the ground truth of selecting a statistically-derived threshold and/or the clinician-derived threshold. Assessing their

professional calibration regarding this study may provide insight into practical usability and clinical interpretation. The interview, slated for a 15-30 minutes, will be carefully worded to minimize bias. The design of the scripted interview will be vetted by a clinician (Dr. Ronald Borg, Phoenix, AZ) and a member(s) of the student's Graduate Supervisory Committee. Lastly, the interview will apply good Usability and User Experience practice as described by the STC Community (Society of Technical Communication 2013). The scripted interview can be found in Appendix A.

Chapter 4. Data Analysis and Results

Section 1. Alpha Test: Proof of Concept (PoC)

Four patients' peak flow sensor data were provided by the field. These four showed unusually high or physiologically improbable results for COPD and/or CHF conditions. Preprocessing was performed as described above for each patient's data point (approximately 200 original measurements each for four patients). Shewart-Western Electric analyses were applied as statistical thresholds. No clinician applied thresholds (high or low) were provided. Time span generally ranged from 12 to 18 months for each patient, but the frequency of measurement was different for each patient, depending on clinician protocol and/or patient compliancy. This frequency of measurement difference within each patient may play a significant part in predictive modeling.

FEV1 and PEF data from patients #1, 10, 11, and 12 were subjected to Shewart-Western Electric analyses. Patients #1 (Figures 4 and 5) and #10 (Figures 6 and 7) have zones A, B, and C identified, corresponding with +/- 1σ, 2σ, and 3σ. UCL (Upper Control Limit) and LCL (Lower Control Limit) are +/- 3σ, respectively. Each point for FEV1 and PEF was tested to the Western Electric Rules of Table 7. A number of

measurements have violated the statistical threshold (noted in red with the corresponding rule violated); most notably, a number of statistical threshold violations are above the mean where a clinician may typically only identify the lower threshold. Low FEV1 and PEF measurements are clinically significant in a COPD and/or CHF patient (indicative of breathing difficulties); however, without the statistical analyses, the clinician has eliminated the opportunity to immediately identify a potential “common cause” (such as a failing peak flow device). The clinician has also eliminated the opportunity to engage in improved personalized telemedicine, even though a measurement may appear close to the mean but violated threshold. In Patient #1, for example, the FEV1 measurements post 11/30/09 indicate that a likely improvement is occurring and prescription medicine may be altered. Conversely, the clinician may miss the opportunity to find a drifting peak flow meter that is not reflective of the patient’s actual health.

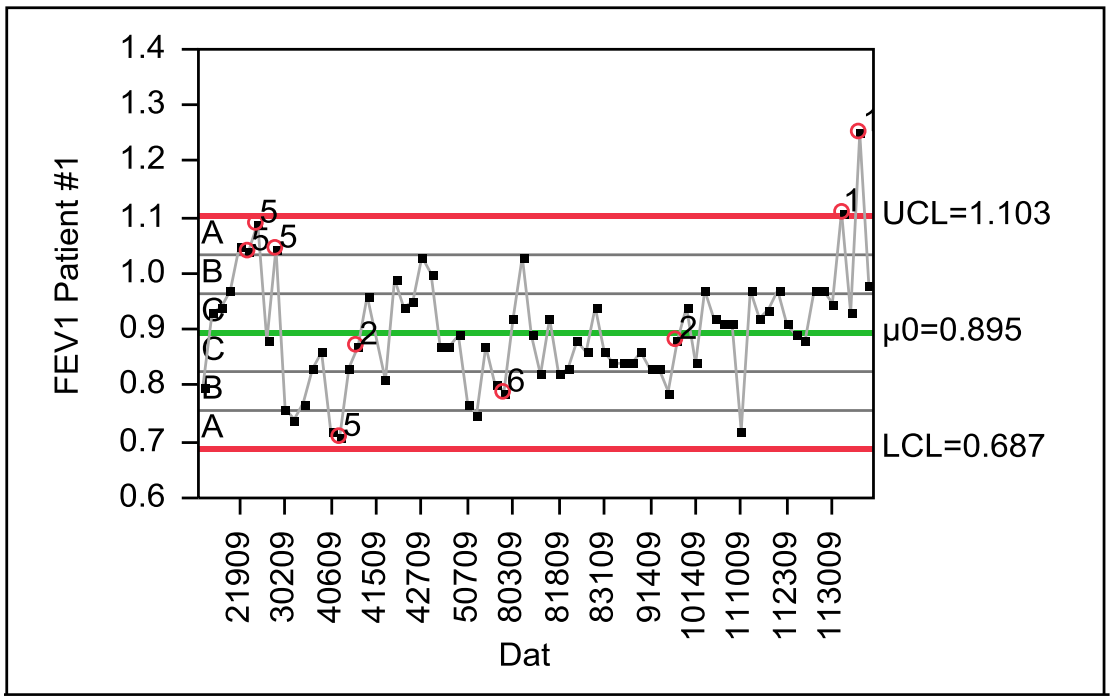


Figure 4. Patient #1 FEV1 Control Chart. A, B, C zones correspond to +/- 1σ, 2σ, and 3σ, respectively. UCL = Upper Control Limit or +3σ. LCL = Lower Control Limit or -3σ. Avg = statistical mean. Red circles indicate statistical threshold violations per Table 7, with the specific rule violated.

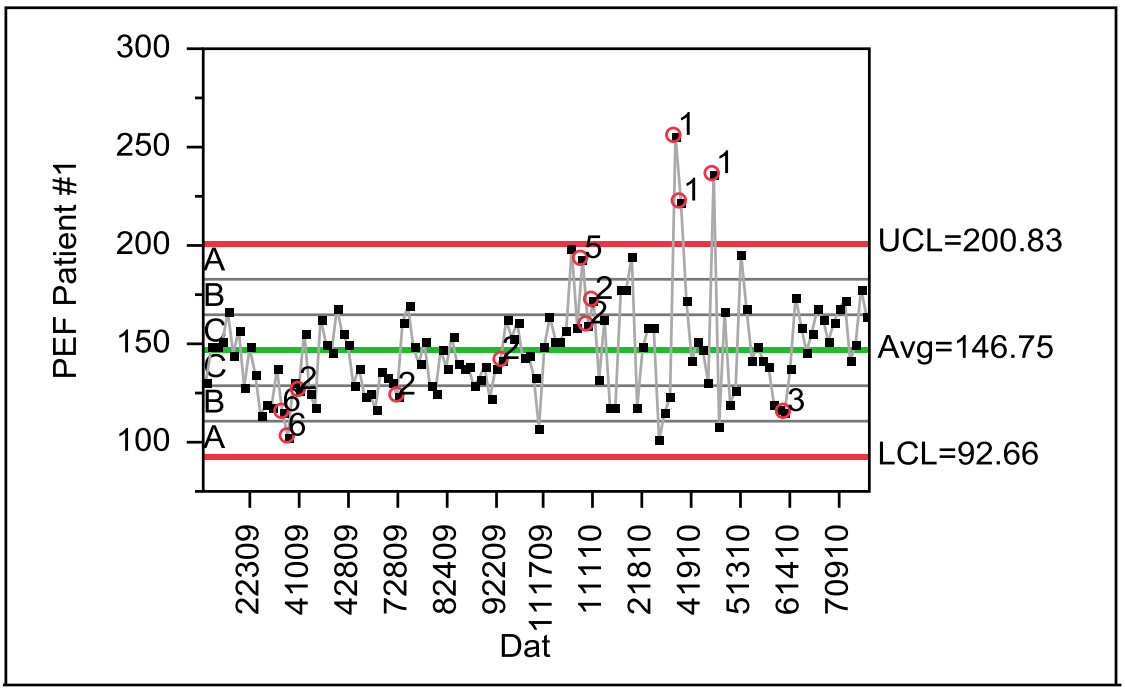


Figure 5. Patient #1 PEF Control Chart. A, B, C zones correspond to +/- 1σ, 2σ, and 3σ, respectively. UCL = Upper Control Limit or +3σ. LCL = Lower Control Limit or -3σ. Avg = statistical mean. Red circles indicate statistical threshold violations per Table 7, with the specific Rule violated.

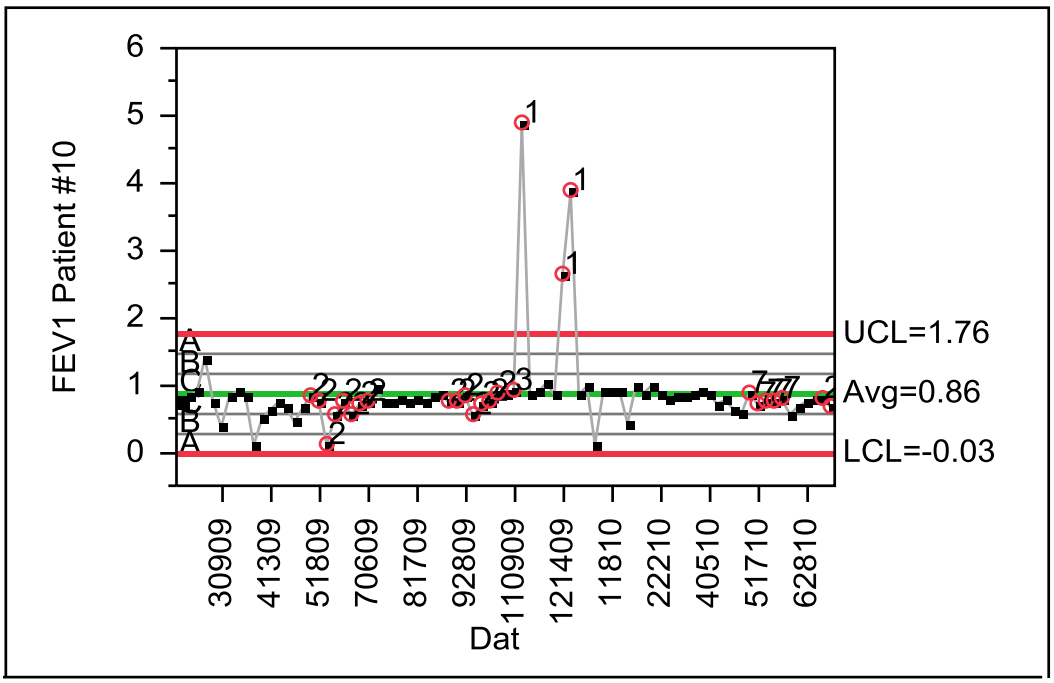


Figure 6. Patient #10 FEV1 Control Chart. A, B, C zones correspond to +/- 1σ, 2σ, and 3σ, respectively. UCL = Upper Control Limit or +3σ. LCL = Lower Control Limit or -3σ. Avg = statistical mean. Red circles indicate statistical threshold violations per Table 7, with the specific Rule violated.

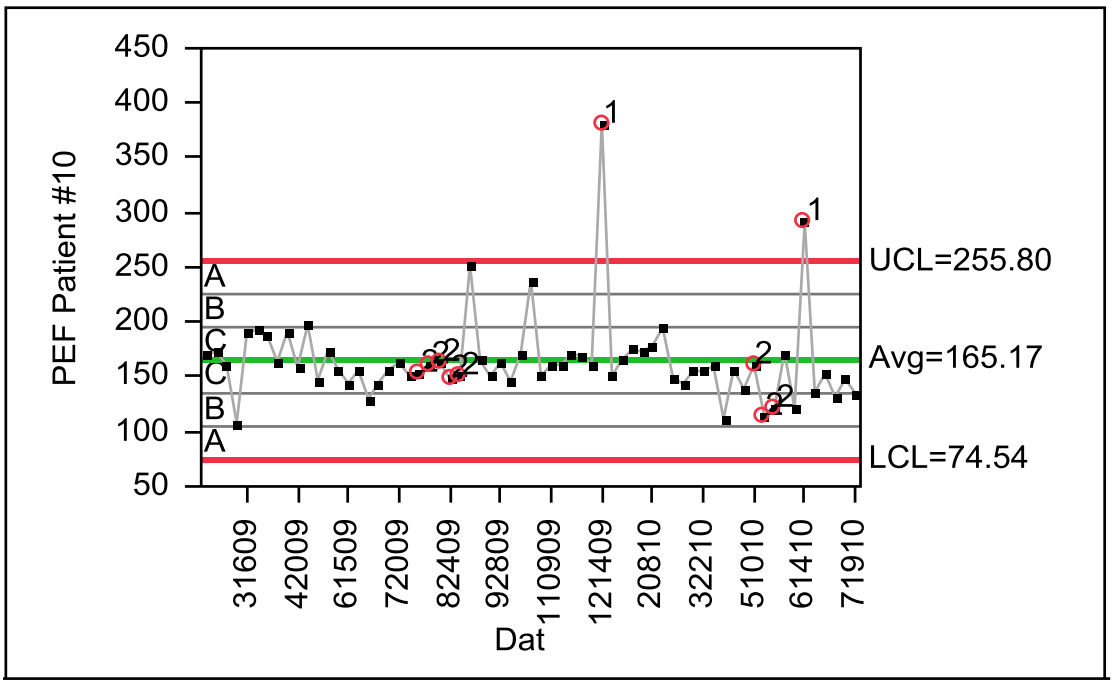


Figure 7. Patient #10 PEF Control Chart. A, B, C zones correspond to +/- 1σ, 2σ, and 3σ, respectively. UCL = Upper Control Limit or +3σ. LCL = Lower Control Limit or -3σ. Avg = statistical mean. Red circles indicate statistical threshold violations per Table 7, with the specific Rule violated.

Patient #11 was subjected to the same preprocessing and analyses as Patients #1 and #10, above. Figures 8 and 9 depict the FEV1 and PEF, respectively, of Patient #11, with statistical threshold violations indicated in red. Once the common cause data (in this case, a failing peak flow meter) were removed at 9/01/09, Shewart-Western Electric analyses was re-applied. As seen in Figures 10 and 11, only “special causes” are now identified by red and specific rule violated is noted. These examples exemplify the importance of identifying and characterizing “the process” (in this case, the patient), prior to implementing Shewart-Western Electric analyses.

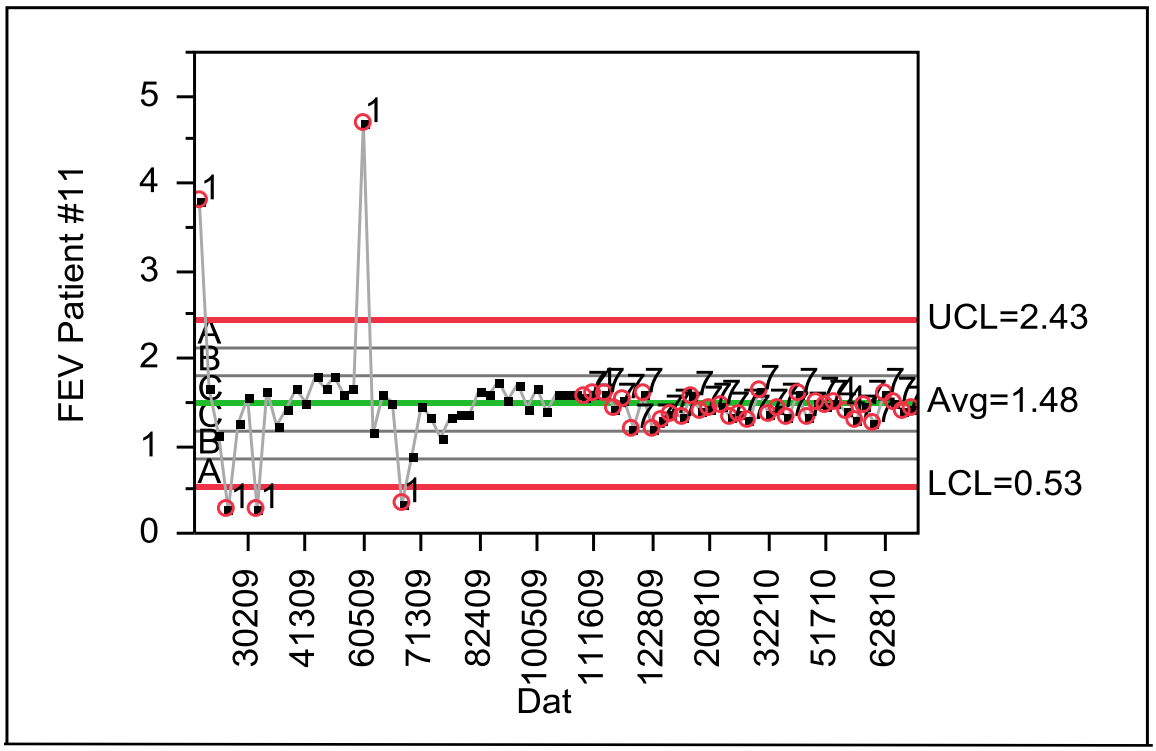


Figure 8. Patient #11 FEV1 Control Chart. A, B, C zones correspond to +/- 1σ, 2σ, and 3σ, respectively. UCL = Upper Control Limit or +3σ. LCL = Lower Control Limit or -3σ. Avg = statistical mean. Red circles indicate statistical threshold violations per Table 7, with the specific rule violated.

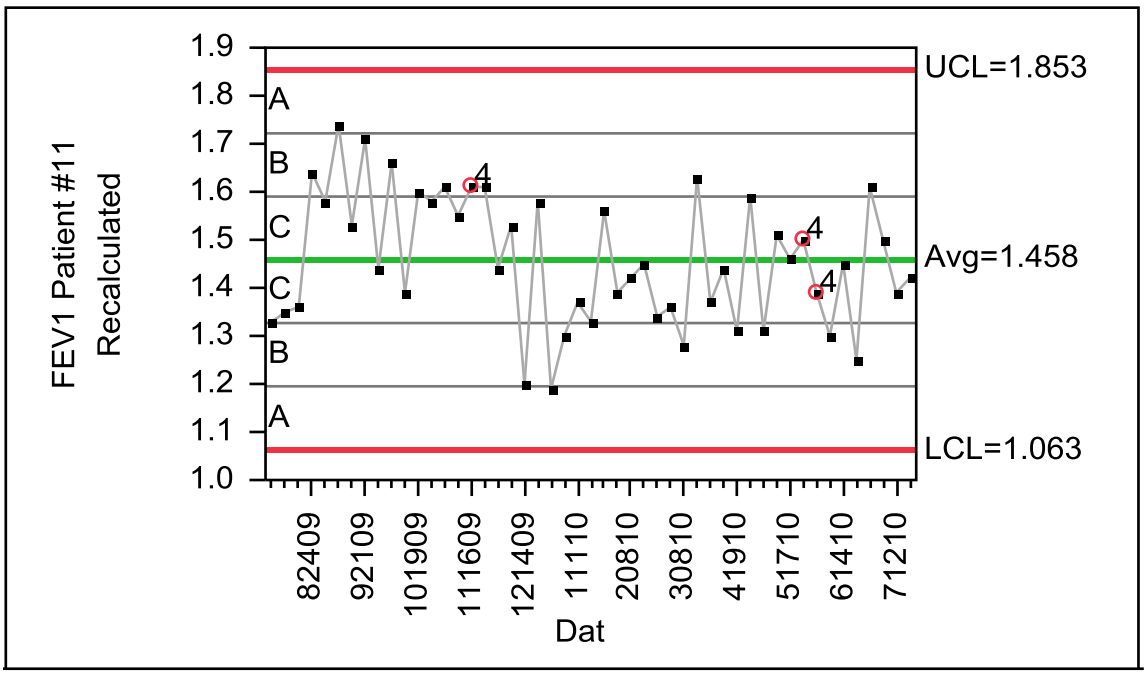


Figure 9. Patient #11 Recalculated FEV1 Control Chart. Obvious outliers have been removed. A, B, C zones correspond to +/- 1σ, 2σ, and 3σ, respectively. UCL = Upper Control Limit or +3σ. LCL = Lower Control Limit or -3σ. Avg = statistical mean. Red circles indicate statistical threshold violations per Table 7, with the specific rule violated.

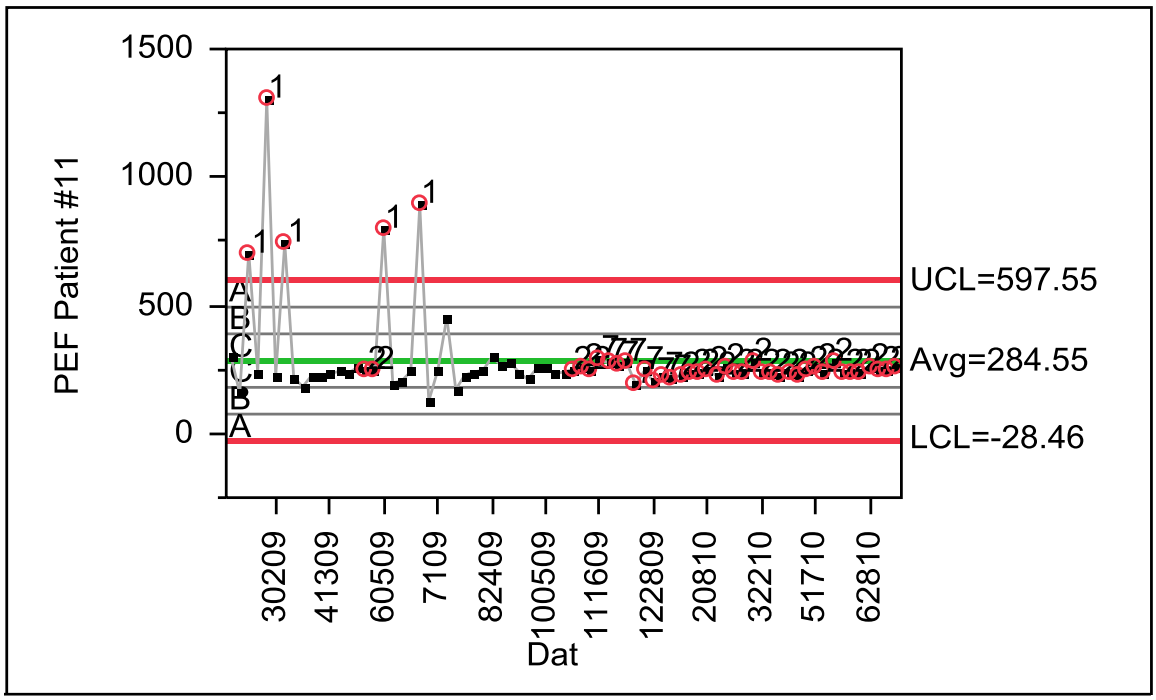


Figure 10. Patient #11 PEF Control Chart. A, B, C zones correspond to +/- 1σ, 2σ, and 3σ, respectively. UCL = Upper Control Limit or +3σ. LCL = Lower Control Limit or -3σ. Avg = statistical mean. Red circles indicate statistical threshold violations per Table 7, with the specific Rule violated.

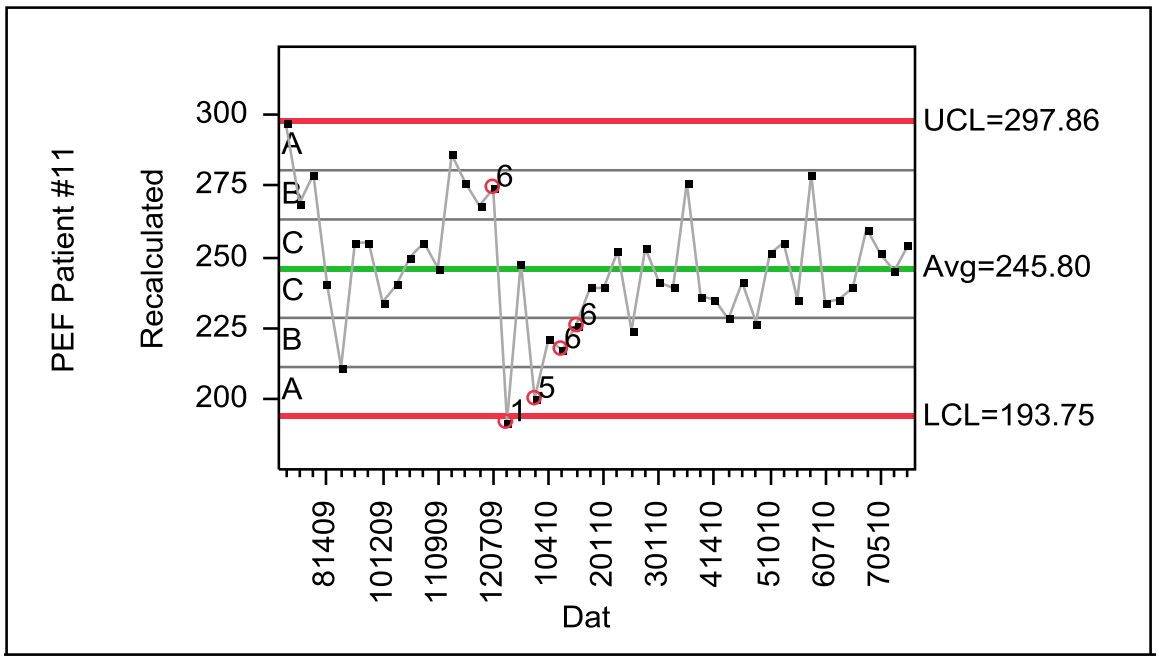


Figure 11. Patient #11 Recalculated PEF Control Chart. Obvious outliers have been removed. A, B, C zones correspond to +/- 1σ, 2σ, and 3σ, respectively. UCL = Upper Control Limit or +3σ. LCL = Lower Control Limit or -3σ. Avg = statistical mean. Red circles indicate statistical threshold violations per Table 7, with the specific Rule violated. Note the reduction of violations compared to Figure 10, prior to outliers being removed. Once the common cause of a failing peak flow device was removed at 09/01/09, Shewart-Western Electric analyses were re-applied, showing fewer as well as different statistical threshold violations.

Patient #12, preprocessed and analyzed per above, depicts a similar common cause as Patient #11 (a suspected failing peak flow meter), but in this case the outliers occur throughout the time span in Figure 12. By identifying the common cause (e.g., characterizing the process) and removing data point outliers from the population, new statistical thresholds are applied in Figure 13. It is clear that, should a clinician only apply upper and lower thresholds, unique opportunities to understand the patient and his clinical manifestations evaporate due to the lack of statistical analyses. The following questions may be posed by the clinician; without the statistical analyses the clinician may be copasetic, unaware that probabilities exist that something may be going very right or very wrong with the patient's health:

- Is the patient using the device correctly?
- Is the peak flow device operating appropriately and accurately?
- Is the patient in distress?
- Is the patient improving?
- Are the patient's activity and/or environment influencing the peak flow reading?
- Has the patient and device's security and privacy been violated?

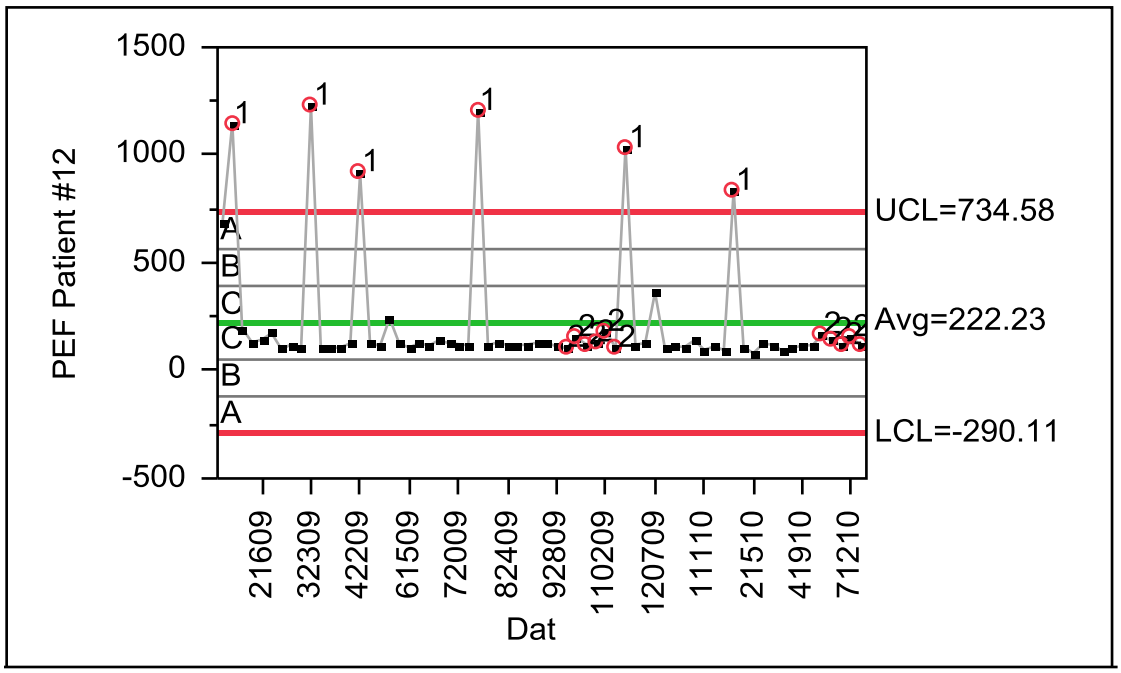


Figure 12. Patient #12 PEF Control Chart. A, B, C zones correspond to +/- 1σ, 2σ, and 3σ, respectively. UCL = Upper Control Limit or +3σ. LCL = Lower Control Limit or -3σ. Avg = statistical mean. Red circles indicate statistical threshold violations per Table 7, with the specific rule violated.

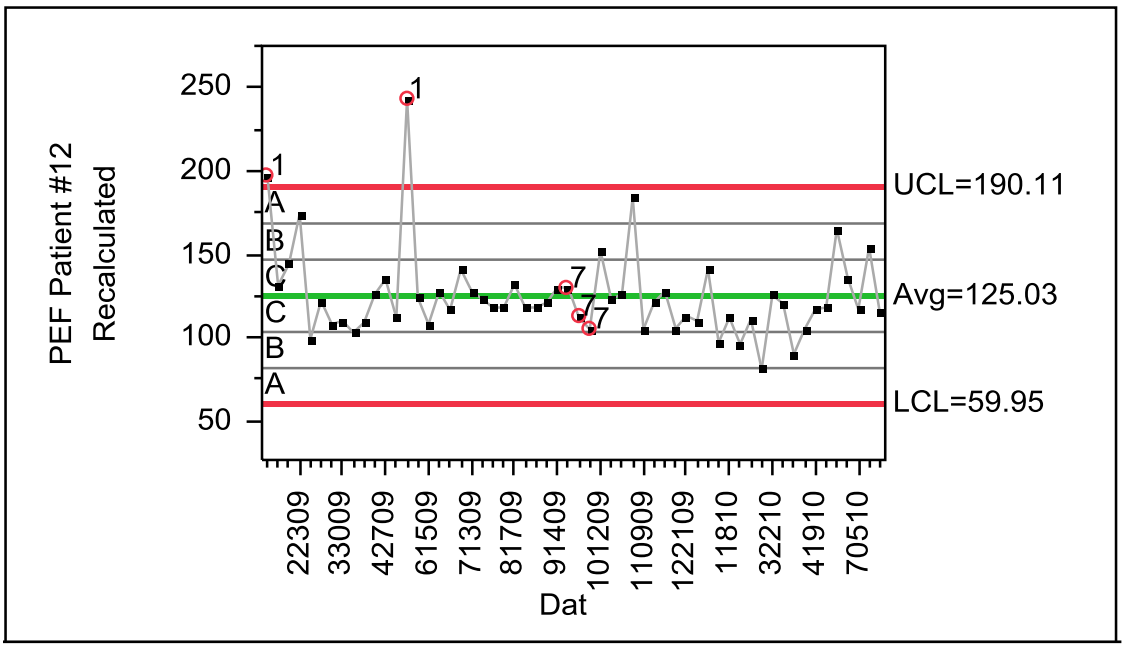


Figure 13. Patient #12 Recalculated PEF Control Chart. Obvious outliers have been removed. A, B, C zones correspond to +/- 1σ, 2σ, and 3σ, respectively. UCL = Upper Control Limit or +3σ. LCL = Lower Control Limit or -3σ. Avg = statistical mean. Red circles indicate statistical threshold violations per Table 7, with the specific rule violated. Note the reduction of violations compared to Figure 12, prior to outliers being removed. Once the common cause of a failing peak flow device was removed at 09/01/09, Shewart-Western Electric analyses were re-applied, showing fewer as well as different statistical threshold violations.

It was determined through visual and electrical validation that the peak flow device high points were due to an optical sensor failure, internal to the peak flow device. The sensor was outdated, and it was determined that shelf-life and reliability issues (causing increased failure rate towards end of life or “bathtub” curve effects) impacted the longevity of the peak flow device and caused the intermittent failures, i.e., a “common cause”. An automated Shewart-Western Electric analyses would have highlighted the failure immediately and reduced customer dissatisfaction due to field failure issues. The

data suggested that, should a clinician observe erratic and improbable results, the device should be returned and replaced.

This alpha test demonstrated the field efficacy of applying SPC in that applying clinician high and low thresholds alone are inadequate for seeing the unusual drift of readings, including a statistical shift in mean.

Section 2. Beta Test: Developing and Testing the Process with One Patient

Results for the Predictive Analytics Process (Figure 3) follow for Patient #1, and include sensor, feature generation and pre-processing, feature extraction and selection, classifier design, and system evaluation. Patient #1 was randomly selected from the patient population.

Sensor, Feature Generation, and Pre-processing

Raw, de-identified FEV1 data from Patient #1 peak flow device was collected and filtered for double and too-frequent entries as described above. Thresholds were not provided by the clinician; therefore, a threshold of .8 L/s was identified based on the GOLD criteria and patient's mean value, indicative of severe COPD (only low thresholds are typically used in COPD and/or CHF telemonitoring). This was validated as acceptable by a local physician, given the lack of information regarding the patient.

Feature Extraction and Selection

Basic statistical analyses were completed for Patient #1. The distribution, mean, standard distribution, variance, skewness, and kurtosis were produced and evaluated (Figure 14). The Shapiro-Wilk test for normality indicates at, an alpha level of 0.05, the

p-value (stated below as “Prob<W”) is slightly significant for normality. Skewness (lack of symmetry) of .64 is likely reflective of the longer upper tail. Kurtosis (data that are peaked or flat relative to the normal distribution) is 1.3 reflective of the heaviness of the tail. Normalization and/or transformation may be necessary to better fit the data into normal distribution if required. This distribution was verified to be Log Normal.

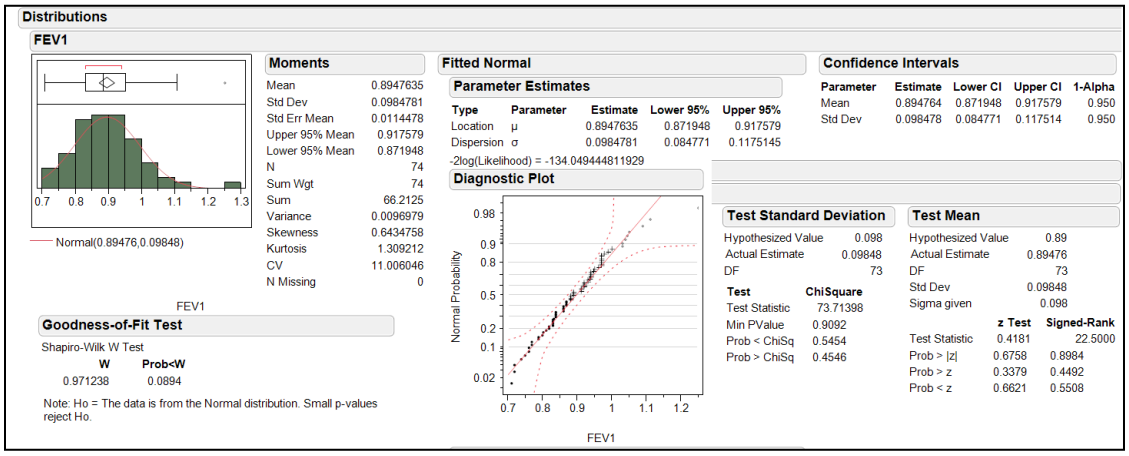


Figure 14. Patient #1 FEV1 Test for Normality and Moments.

Next, a Shewart-Western Electric analysis was performed on Patient #1 data, indicating Lower Control Limit (-3σ) of 0.687, Upper Control Limit (+3σ) of 1.103, and mean of 0.896 (Figure 15). Nine points that violated the statistically-derived thresholds (i.e., out of control) was marked with the rule number violated. The clinician-derived threshold was marked at .8 L/s FEV1, and, therefore, all points below .8 are classified as violating the clinician-derived threshold, a total of 12. Two of the failed clinician-derived points also failed the Western Electric Rules. Each data point was classified into TP, FP, FN, and FP per Table 2. Summary of the classification for Patient #1 is in Table 9.

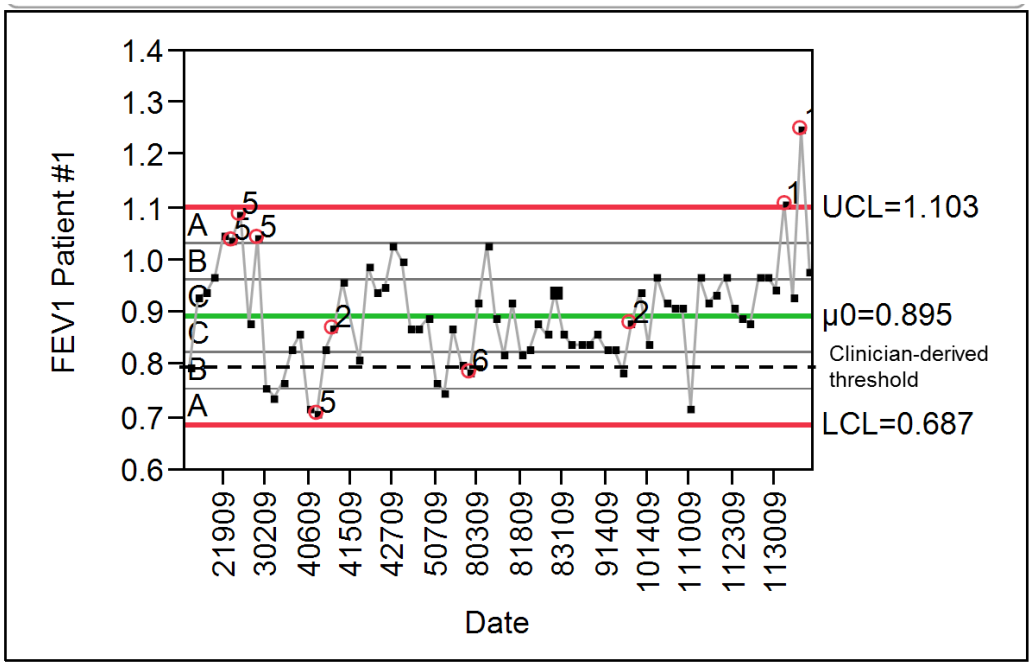


Figure 15. Patient #1 FEV1 Control Chart with Thresholds. Clinician-derived threshold of .8 L and statistically-derived thresholds are marked with rule violated. UCL = Upper Control Limit. LCL = Lower Control Limit.

Table 9. FEV1 Classification of Breached Thresholds. Each point in the time series was assigned either TP, FN, FP, or TN depending on whether the point passed or failed a statistically- or clinician-derived threshold.

Intervention	Test	Number breached violations for Patient #1	State	Statistically Derived	State	Clinician Derived	Combined State
Intervened	TP	2	1	Threshold violation	1	Threshold violation	1,1
Did not intervene but should have (statistically)	FN	7	1	Threshold violation	0	No threshold violation	1,0
Intervened but shouldn't have (statistically)	FP	10	0	No threshold violation	1	Threshold violation	0,1
No intervention necessary	TN	55	0	No threshold violation	0	No threshold violation	0,0

As an alternative, CUSUM (cumulative sum) and V-mask were also calculated based on Montgomery (Montgomery 1994) and Lucas (Lucas 1982). It provides a different schema (sequential) for identifying out-of-control points (i.e., failed thresholds) by calculating the cumulative sum of differences between the values and the mean, and applying a V-mask (for both min and max)[4]. The V-mask provides visual detectability of the failed threshold easier, and can detect small changes of about 1.5σ or less.

CUSUM considers several points prior, while the Western Electric Rules consider the point before, and, therefore, CUSUM is able to detect smaller shifts in the data. If the

[4] A V-mask is calculated via K, the rise in the arm corresponding to one sampling unit, and h, the rise in the arm to the distance d from origin to vertex.

process remains in control, the points will hover around zero, while if they are drifting upwards or downwards, the process is shifting and an assignable cause should be investigated. In the case of the COPD patient, this may be an alternate method for clinicians to assess their patients' health.

Data from Patient #1 were subjected to CUSUM analysis (Figure 16), indicating that the patient is downward trending, and exceeds the threshold towards the latter dates. While this approach minimizes false positives, it does not allow each point to be easily classified into the four-schema classification (TP, FN, FP, TN) put forth in this study.

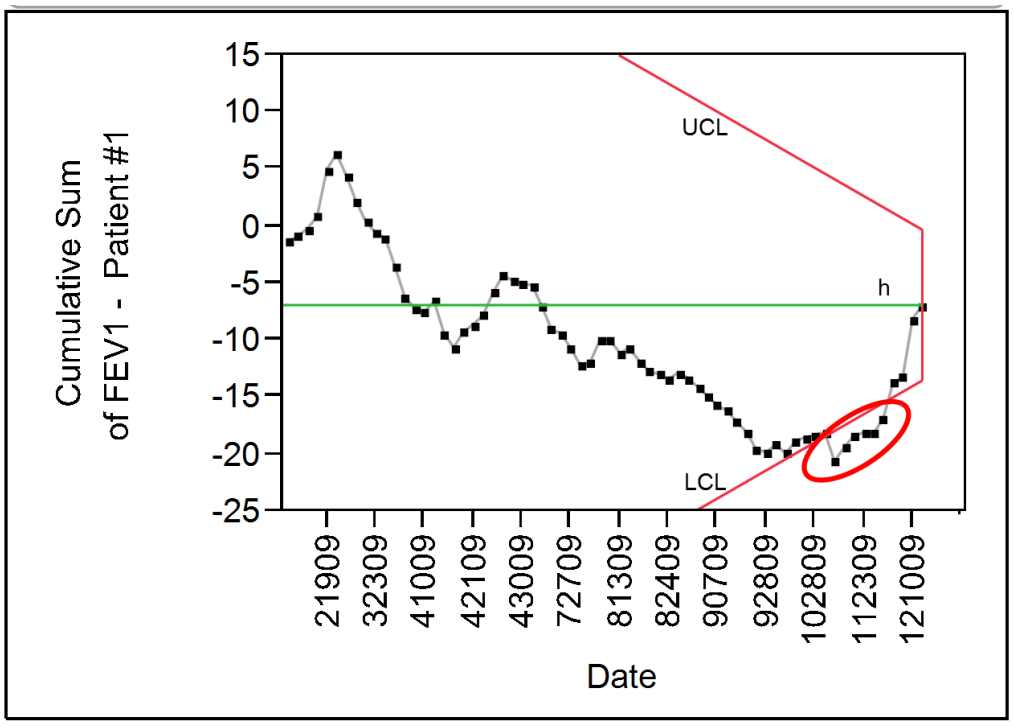


Figure 16. Patient #1 FEV1 CUSUM Results. Circled area indicates out-of-control points, but patient also exhibits downward trending. V-masks (i.e., UCL and LCL) should be applied after every measurement to insure faster feedback on "threshold" violation. "h" is the vertical distance between the upper and lower arms of the V-mask.

Additional statistics were calculated for FEV1 for Patient #1. A Student's t-test to compare means between [TP, TN] and [FP. FN] with means of 0.9008 and 0.8744, respectively, resulted in t = 0.9705 (assuming equal variances) and t = -0.64011 (assuming unequal variances) with $p<.05$. This indicated that the means were not significantly different from one another. Other means comparison tests also verified that the means were not significantly different (Tukey-Kramer, Best Hsu's MCB, Dunnett's). An ANOVA indicated variances are equal (F=0.9419); however, other tests (O'Brien (F=21.1728), Levene (F=21.5023), and Bartlett (F=21.6117)) reject that hypothesis ($p<0.05$). Power is a poor 0.1598. Obviously, these statistics are disappointing and likely reflective of the low sample size; in classifier design (discussed below), further classification was attempted using clustering techniques.

Classifier Design

Further investigations were performed on the [TP, TN] and [FN, FP] classification applied to Patient #1 FEV1 values. Figure 17 represents the Receiver Operating Characteristic (ROC) curve, a calculation where the highest true positive rate and lowest false positive rate is considered perfect classification (at point 0,1). The ROC curve presents sensitivity, the probability that a given x value correctly predicts [FN, FP] and specificity, the probability of incorrectly predicting [FN, FP] (shown in Figure 18 as 1-specificity). Specificity is also known as recall. Given [TP, TN] and [FN, FP]. A ROC curve for Patient #1 yielded the matrix in Figure 18 and resulting graph in Figure 19. The resulting graph indicates that, between these two groups, the probabilistic difference in predicting [FN, FP] interventions is above a random guess for Patient #1 given the area under curve = 0.651.

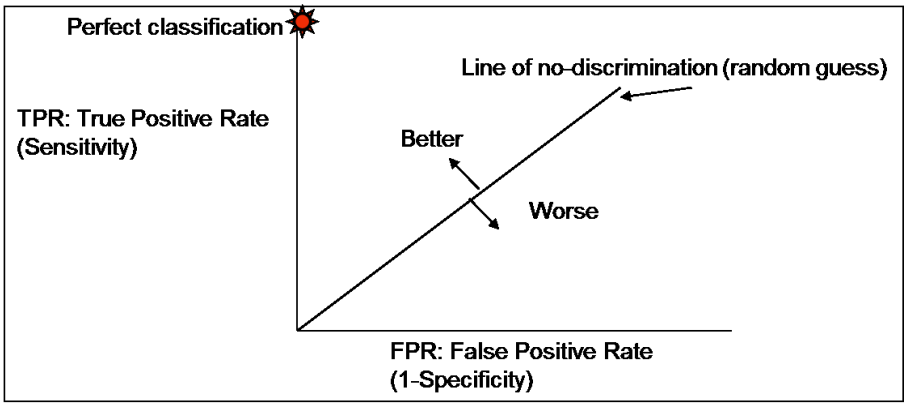


Figure 17. ROC. General diagram of a Receiver Operating Curve (ROC)

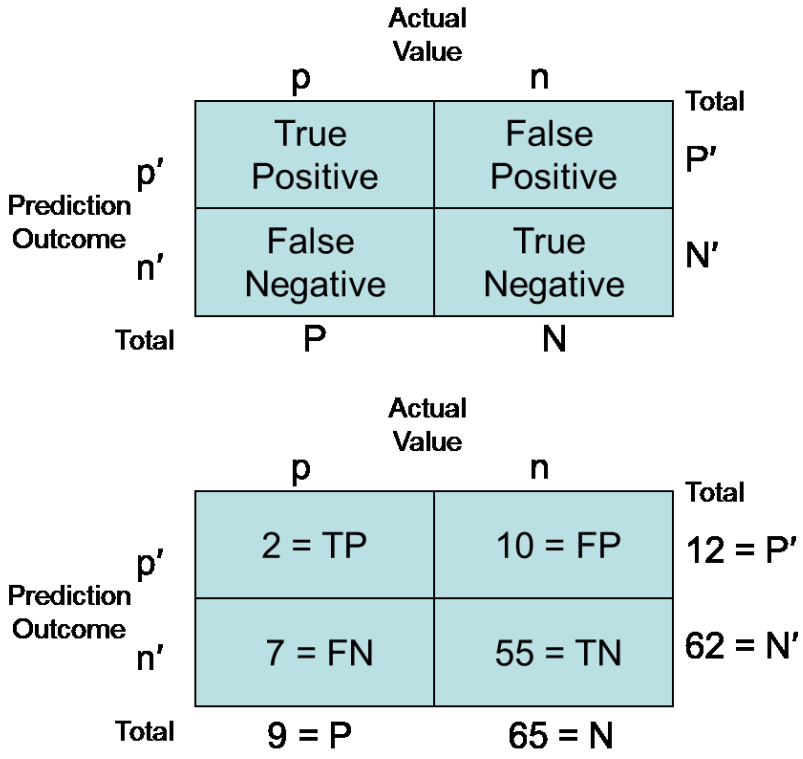


Figure 18. Patient #1 Confusion Matrix. Patient #1 FEV1 data sorted by TP, FN, FP, TN for ROC preparation.

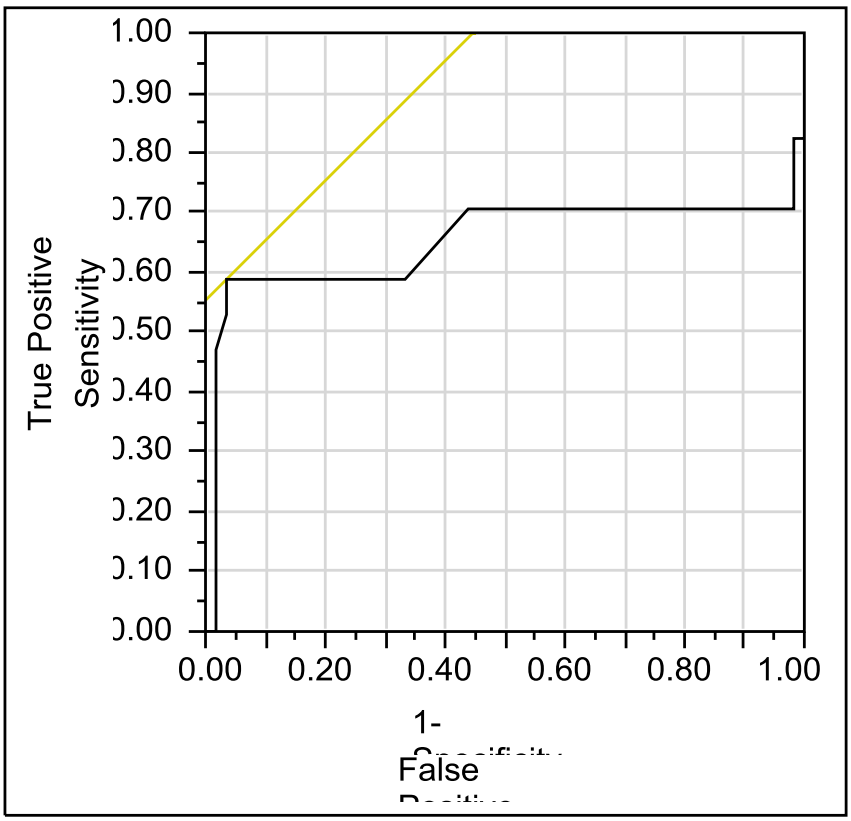


Figure 19. Patient #1 FEV1 ROC. Patient #1 FEV1 Receiver Operating Curve using [FN, FP] to be positive level. Area under curve = 0.651. The line drawn at a 45 degree angle tangent to the curve segregates the true positive rate (TPR = TP(TP+FN)) and the false positive rate (FPR = FP(FP+TN); this line indicates where the cost of true positive and the false negative are the same. The curve represents a logistic probability of each value x correctly or incorrectly predicting [FN, FP]. This graph indicates that the probabilistic difference in predicting [FN, FP] interventions is above a random guess.

While the results of the Shewart – Western Electric analysis essentially provide feature classification, additional analysis between the classifications was performed. Clusters were assigned: cluster 1 = FP; cluster 2 =; TN; cluster 3 = FN; and cluster 4 = TP. Analysis of means of FEV1 by cluster (Figure 20) shows visually that the means of

clusters 2 and 3 (uppermost circle and smallest circle, respectively) are significantly different from means of clusters 1 and 4 – expected since the clinician threshold was set at .8 for PEV1. Clusters 2 and 3 represent no clinical threshold violations while cluster 1 and 4 represent statistical threshold violation and no statistical threshold violation (per Table 9, above).

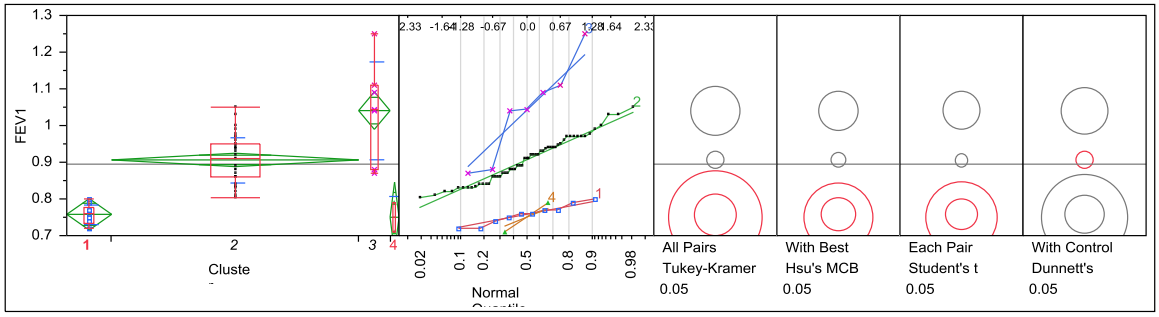


Figure 20. Patient #1 FEV1 Data by Cluster.

An ANOVA was completed (Figure 21), with an improved R-squared of .54, and statistically significant F Ratio of 27.88. Variances of each cluster were compared using O'Brien's method and found to be unequal (F ratio = 11.6, prob>F = <0.0001). Following these analyses, a number of other classifier algorithms were tested. Linear and Quadratic Discriminate analyses were completed: the Quadratic indicated that of 74 readings and their respective four clusters, 15 points (20.3%) were "misclassified". In examination of these misclassifications, 13/15 would have failed (i.e., had a threshold violation either statistically or by clinician) suggesting an increase in false negatives.

Oneway Anova
Summary of Fit

Rsquare	0.544404
Adj Rsquare	0.524878
Root Mean Square Error	0.06788
Mean of Response	0.894764
Observations (or Sum Wgts)	74

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Cluster	3	0.38541000	0.128470	27.8816	<.0001*
Error	70	0.32253909	0.004608		
C. Total	73	0.70794909			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
1	10	0.75800	0.02147	0.71519	0.8008
2	55	0.90632	0.00915	0.88806	0.9246
3	7	1.04071	0.02566	0.98954	1.0919
4	2	0.75000	0.04800	0.65427	0.8457

Figure 21. Patient #1 FEV1 ANOVA Summary of Fit by Cluster. The fit is unexceptional, noted by the "Rsquare" = 0.544404.

System Evaluation

Following Figure 22, a number of models were explored to find the best fit for predicted responses. Since new data is unavailable for each patient, the approach was to utilize as few as FEV1 readings as possible to predict whether the next reading would fall into a specific cluster.

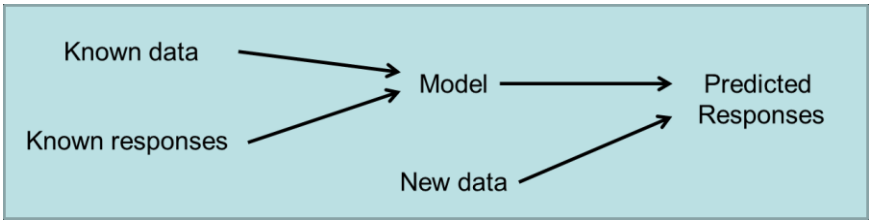


Figure 22. General Process for Achieving Predicted Responses.

Winter's Method (equal to a seasonal ARIMA, or a triple exponential smoothing method to forecast within a time series) performed better than other time-series smoothing models, though it is still a relatively poor fit (Figure 23). The resulting graph was still not adequate to validate a cluster, as observed by the unremarkable "RSquare".

Model: Winters Method (Additive)
Model Summary

DF	54
Sum of Squared Errors	0.59508385
Variance Estimate	0.01102007
Standard Deviation	0.10497653
Akaike's 'A' Information Criterion	-84.338562
Schwarz's Bayesian Criterion	-78.209408
RSquare	-0.4407742
RSquare Adj	-0.4913277
MAPE	8.60995228
MAE	0.07722535
-2LogLikelihood	-90.338562
Stable	Yes
Invertible	No

Hessian is not positive definite

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Level Smoothing Weight	0.6687704	0.121079	5.52	<.0001*
Trend Smoothing Weight	5.5511e-17	0.000000	.	<.0001*
Seasonal Smoothing Weight	1.0000000	5.199606	0.19	0.8482

Figure 23. Patient #1 FEV1 Winter's Method (Smoothing Model).

Continuing to search for an appropriate model, a neural network algorithm was attempted. The network was trained using the Levenberg-Marquardt back-propagation tool (MatLab® R2011b (7.13.0.564)). Using Patient #1's FEV1 measurements as inputs, and clusters as outputs, Figure 24 and 25 reveal that this model, while improved ($R > .40$), is inadequate for predicting.

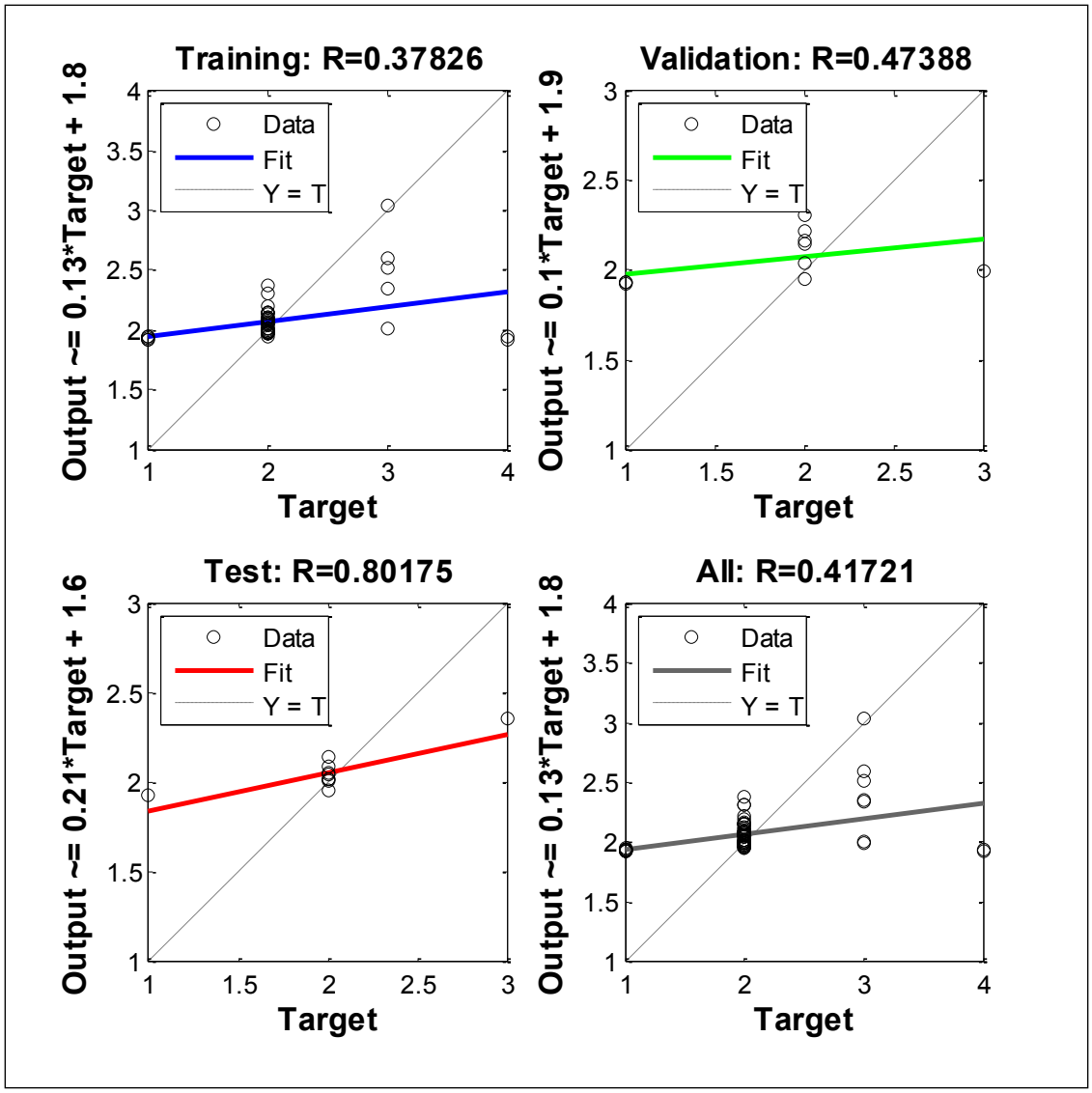


Figure 24. Patient #1 LM Back-propagation Model. Patient #1 FEV1 predictive results based on using a Levenberg-Marquardt back-propagation neural network.

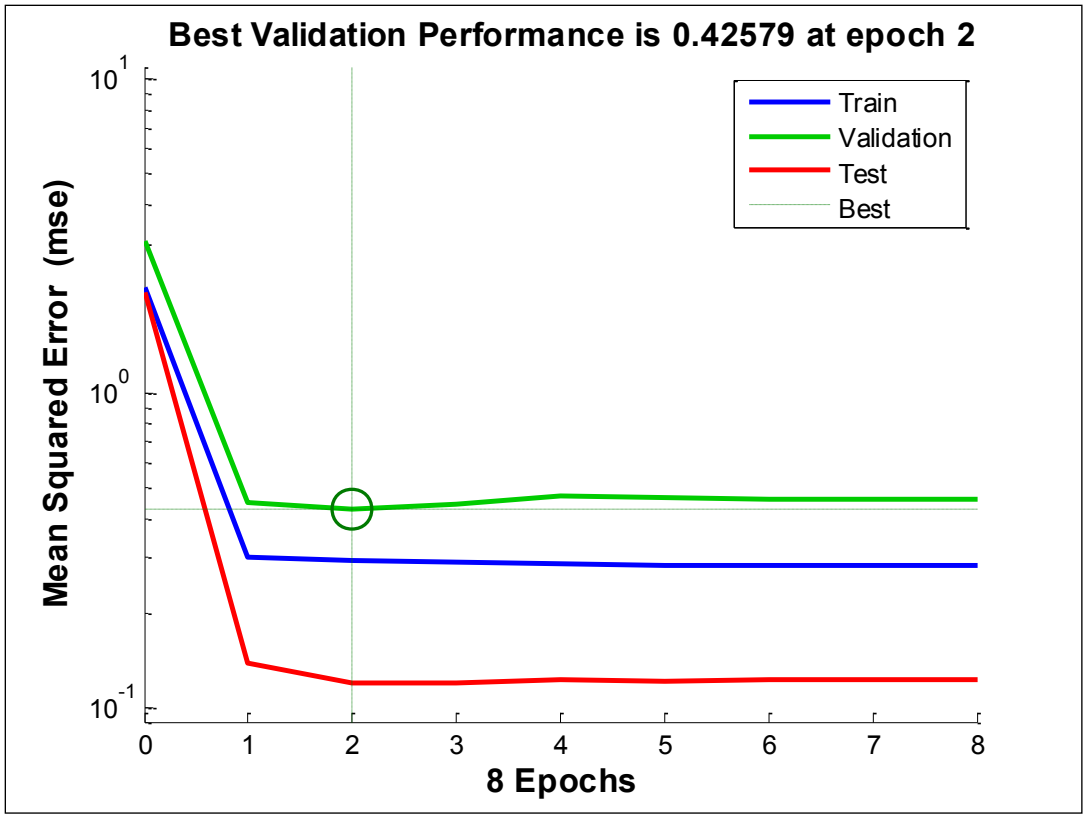


Figure 25. Patient #1 FEV1 Mean Squared Error of LM Validation. Patient #1 FEV1 results of Levenberg-Marquardt validation of the network using mean squared error. The best validation performance yielded 0.42579.

Completing the Beta Test, further exploration using neural networks and/or other time series algorithms for the Final Construct using all twenty patients was launched. The term and classification of "cluster" was found unnecessary for future analyses. Prediction of the classification of TP, FN, FP, or TN, predicted points (i.e., values) in time, and control chart zones (A, B, or C signifying +/- 3σ, 2σ, or 1σ, respectively) will be investigated based on the Beta Test learnings. Additional FEV1 data for Patient #1 was discovered; therefore, Patient #1's analyses will be redone.

Section 3. Final Construct: Sensor, Feature Generation, and Pre-processing Results

Raw, de-identified FEV1 data from twenty patients was secured from the server supporting the telemonitoring unit. No patient contact occurred, and data was de-identified prior to receipt by the student. Each patient's data was filtered for double and too-frequent entries as described above. Obvious outliers were removed; some had been missed during Beta Testing. Approximately 10-15% of the entries over the course of the remote monitoring (typically about twelve to eighteen months) were discarded due to these reasons. It was discovered during the course of the data analysis that none of the thresholds were provided by the clinicians; therefore, each patient was assigned their own lower threshold based on GOLD (Global Initiative for Chronic Obstructive Lung Disease 2013) criteria and patient's mean value. These thresholds were validated visually by a local clinician to be appropriate given the limited available information. All specific aims and their respective hypotheses were validated in the Final Construct.

Section 4. Final Construct: Feature Extraction and Selection

The upper and lower control limits, mean, minimum, maximum, range, standard distribution, variance, skewness, kurtosis and distribution were examined for each of the twenty patients (Tables 10 and 11). Shewart-Western Electric analyses were also performed, and each one of the patients' data points was classified into TP, FN, FP, or TN (Table 12).

Table 10. All Patients' Basic Statistics One. Patients' #1-20 FEV1 N, Upper Control Limit (UCL), Lower Control Limit (LCL), mean, standard deviation, minimum, maximum, and range. Summary statistics are noted below the value (SUM = summation; AVG = average).

Patient #	N	UCL	LCL	Mean	Std Dev	Min	Max	Range
1	126	1.181	0.667	0.924	0.125	0.710	1.300	0.590
2	67	0.950	0.543	0.747	0.105	0.485	1.000	0.515
3	81	1.093	0.601	0.847	0.880	0.620	1.020	0.400
4	70	0.998	0.400	0.699	0.143	0.420	1.150	0.730
5	77	0.852	0.319	0.586	0.103	0.230	0.890	0.660
6	70	0.954	0.513	0.733	0.076	0.540	0.870	0.330
7	60	0.812	0.221	0.517	0.113	0.160	0.660	0.500
8	57	1.230	0.350	0.790	0.187	0.250	1.220	0.970
9	66	3.816	1.817	2.817	0.348	1.880	3.880	2.000
10	78	1.646	0.142	0.752	0.198	0.100	1.400	1.300
11	49	1.853	1.063	1.458	0.135	1.190	1.740	0.550
12	62	1.968	0.340	1.154	0.308	0.130	1.610	1.480
13	382	1.732	0.798	1.265	0.182	0.670	1.800	1.130
14	50	1.040	0.596	0.818	0.818	0.126	0.625	1.220
15	59	0.9	0.374	0.637	0.090	0.450	0.780	0.330
16	92	2.041	0.995	1.518	1.518	0.227	0.820	2.090
17	60	0.725	0.435	0.58	0.580	0.058	0.400	0.680
18	59	1.343	0.425	0.884	0.884	0.186	0.580	1.410
19	64	0.984	0.647	0.815	0.059	0.675	1.000	0.325
20	43	1.068	0.566	0.817	0.100	0.620	1.040	0.420
Summary	1672	1.359	0.591	0.968	0.347	0.486	1.189	0.882
Statistics	SUM	AVG	AVG	AVG	AVG	AVG	AVG	AVG

Table 11. All Patients' Basic Statistics Two. Patients' #1 – 20 FEV1 N, variance, standard error, coefficient of variance, skewness, kurtosis, and distribution. Summary statistics are noted below the value (SUM = summation, AVG = average).

Patient #	N	Variance	Std Error	Coeff Var	Skewness	Kurtosis	Distribution
1	126	0.016	0.011	13.537	0.985	1.134	Normal 2 Mixture
2	67	0.011	0.013	14.018	-0.030	0.036	Normal
3	81	0.008	0.010	10.391	-0.293	-0.331	Weibull
4	70	0.020	0.017	20.463	0.736	0.351	Log Normal
5	77	0.011	0.012	17.582	-0.666	2.870	Johnson Su
6	70	0.006	0.009	10.355	-0.283	-0.497	Weibull
7	60	0.013	0.015	21.837	-1.482	2.024	Johnson SI
8	57	0.035	0.025	23.615	0.056	0.903	Normal
9	66	0.121	0.043	12.343	0.144	0.630	Normal
10	78	0.039	0.022	26.278	-1.067	4.096	Johnson SI
11	49	0.018	0.019	9.281	0.063	-0.828	Gamma/Normal
12	62	0.095	0.039	26.650	-1.936	4.102	Normal 2 Mixture
13	382	0.033	0.009	14.402	0.124	0.205	Normal
14	50	0.595	0.016	0.018	0.904	0.823	Johnson SI
15	59	0.008	0.012	14.075	-0.486	-0.750	Weibull
16	92	1.270	0.052	0.024	-0.236	0.355	Johnson SI
17	60	0.280	0.003	0.008	-0.898	0.950	Weibull
18	59	0.830	0.035	0.024	0.035	0.594	Normal 2 Mixture
19	64	0.003	0.007	7.238	0.536	0.698	Log Normal
20	43	0.010	0.015	12.178	0.134	-0.208	Johnson Su
Summary	1672	0.171	0.019	12.716	-0.183	0.858	
Statistics	SUM	AVG	AVG	AVG	AVG	AVG	

Table 12. All Patients' Four Schema Results, Pre-predictive Algorithm. Patients' #1-20 FEV1 N, Upper Control Limit (UCL), Lower Control Limit (LCL), mean, clinician-derived threshold, and the number of data points that are classified into TP, FN, FP, or TN. Summary statistics are noted below the values (SUM = summation; AVG = average).

Patient #	N	Mean	Clinician Threshold	# TP	# FN	# FP	# TN	Percent Increase from FP to FN
1	126	0.924	0.75	2	17	5	102	9.524
2	67	0.747	0.65	6	17	9	35	11.940
3	81	0.847	0.70	1	2	2	76	0.000
4	70	0.699	0.55	2	18	8	42	14.286
5	77	0.586	0.40	3	1	1	72	0.000
6	70	0.733	0.65	0	1	11	58	-14.286
7	60	0.517	0.35	3	0	2	55	-3.333
8	57	0.790	0.67	2	0	10	45	-17.544
9	66	2.817	2.40	2	10	6	48	6.061
10	78	0.752	0.30	0	28	3	47	32.051
11	49	1.458	1.30	0	3	6	40	-6.122
12	62	1.154	0.75	3	3	2	54	1.613
13	382	1.265	0.85	2	56	1	323	14.398
14	50	0.818	0.70	4	12	4	30	16.000
15	59	0.637	0.50	0	0	8	51	-13.559
16	92	1.518	1.25	3	11	6	72	5.435
17	60	0.58	0.52	5	3	6	46	-5.000
18	59	0.884	0.69	4	4	4	47	0.000
19	64	0.815	0.73	0	2	3	59	-1.563
20	43	0.817	0.70	0	1	6	36	-11.628
Summary	1672	0.968	0.771	2.1	9.5	5.2	66.9	1.914
Statistics	SUM	AVG	AVG	AVG	AVG	AVG	AVG	AVG
				42	189	103	1338	
				SUM	SUM	SUM	SUM	

Hypothesis A of the first specific aim can now be tested (prior to the application of the Predictive Analytics Process): H_0: FN=FP, and H_1: FN≠FP, where H_1 states that there will be a statistically significant (p<.005) difference in number of clinical interventions when statistically-derived thresholds are applied. The standard normal (z) table shows that the lower critical z value for p<.005 is approximately 1.04. The computed z value must be higher than the 1.04 in order to reject the null hypothesis. Because the computed z = 1.1062, the null hypothesis is rejected. It can be concluded that, at this level of significance, there is a difference between FN and FP (p<.005) when statistically-derived thresholds are applied. Therefore, H_0 of Hypothesis A is rejected, and H_1 is accepted. The clinician will have intervened an average of an additional 86 times (or 11.3% =189/1672)*100) than he would have normally, when he intervened but shouldn't have (or 6.2% = (103/1672)*100), based on statistically-derived thresholds. Figure 26 demonstrates the varied swings of the percent increase from FP to FN across all twenty patients. Average increase, however, was only 1.914%.

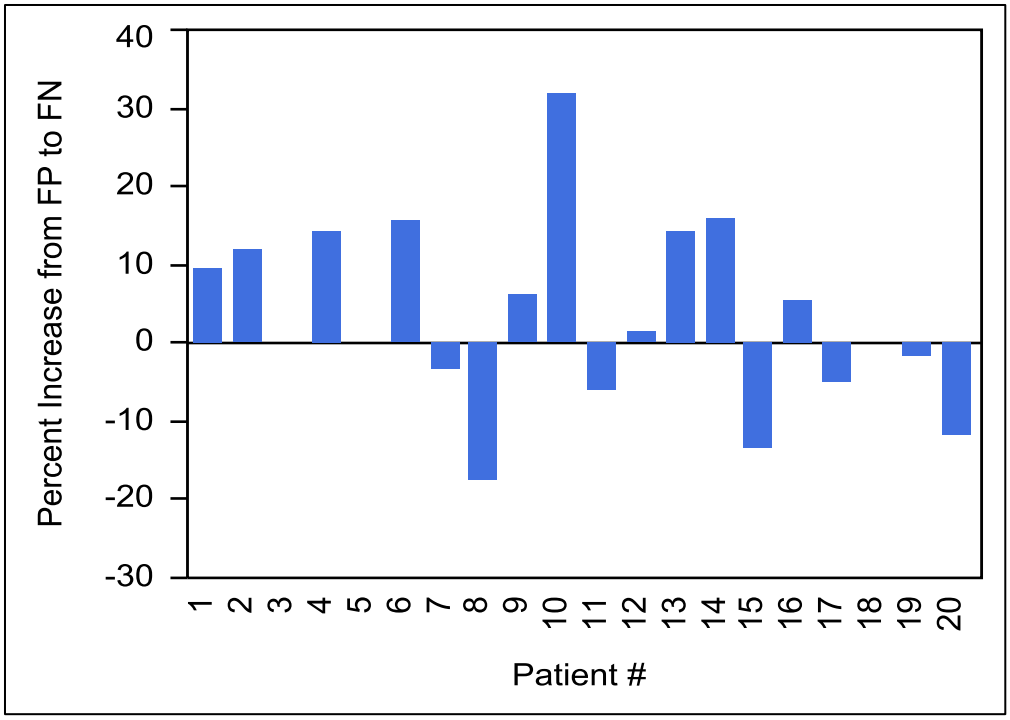


Figure 26. All Patients' FEV1 Percent Increase from FP to FN.

To understand if there was any stratification in the population, each patient's FEV1 values were evaluated for severity of the disease based primarily on GOLD standards. Though each patient's disease state was not provided (i.e., COPD, CHF, or COPD/CHF or other co-morbidity such as diabetes), FEV1 values fell into four different likely states of severity and were examined with the statistics of all patients (Figure 27 - 39 and Tables 13 – 15). Note small sample size per disease severity. The results of the non-parametric tests of Wilcoxson/Kruskal-Wallis indicating significant difference between the groups (mild/moderate and severe/very severe) are included in appropriate graphs via their *z* value and Chi-Square.

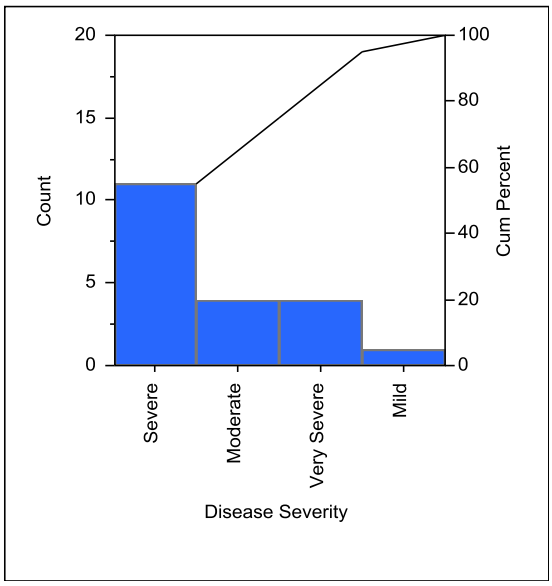


Figure 27. All Patients' FEV1 Pareto Plot of Disease Severity. Patients #1-20 FEV1 Pareto plot of likely disease severity (out of 20 total patients in study)

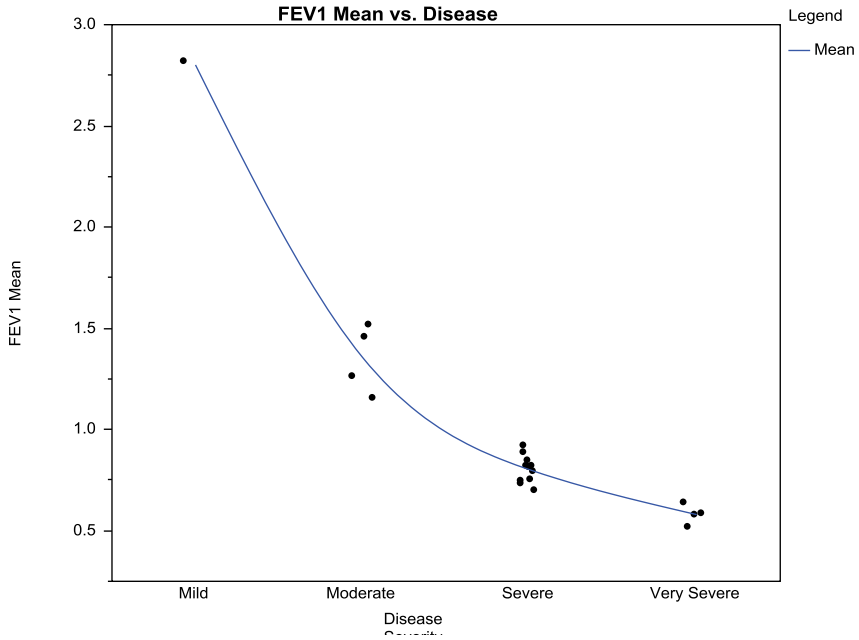


Figure 28. All Patients' FEV1 Mean vs. Disease Severity. Actual values plus smoother.

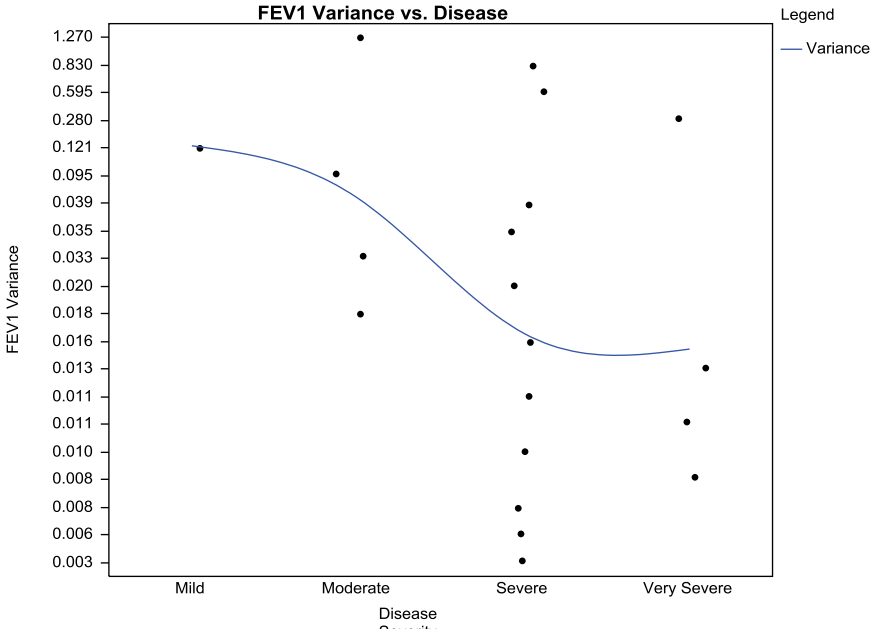


Figure 29. All Patients' FEV1 Variance vs. Disease Severity. Actual values plus smoother. *z* = 1.745, Chi-Square = 3.202 indicating significant difference.

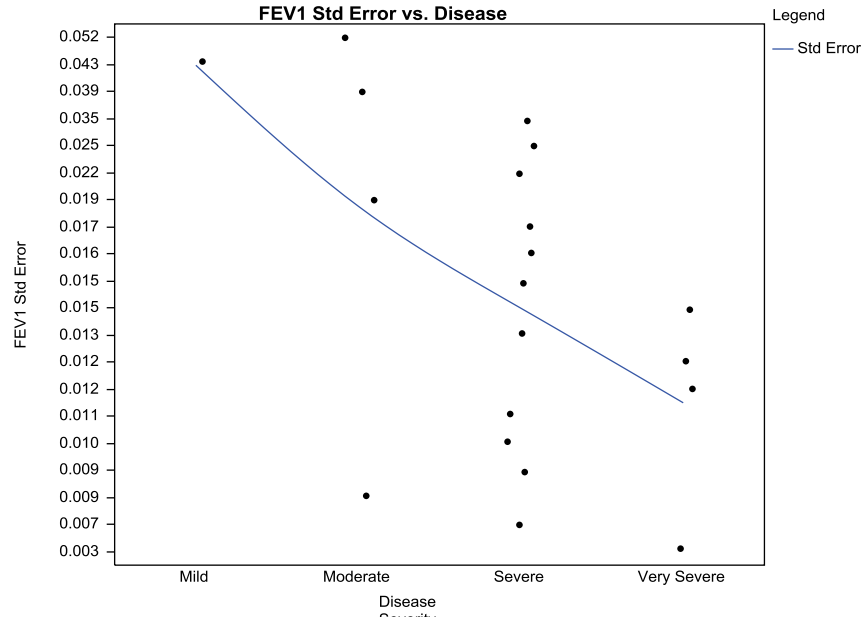
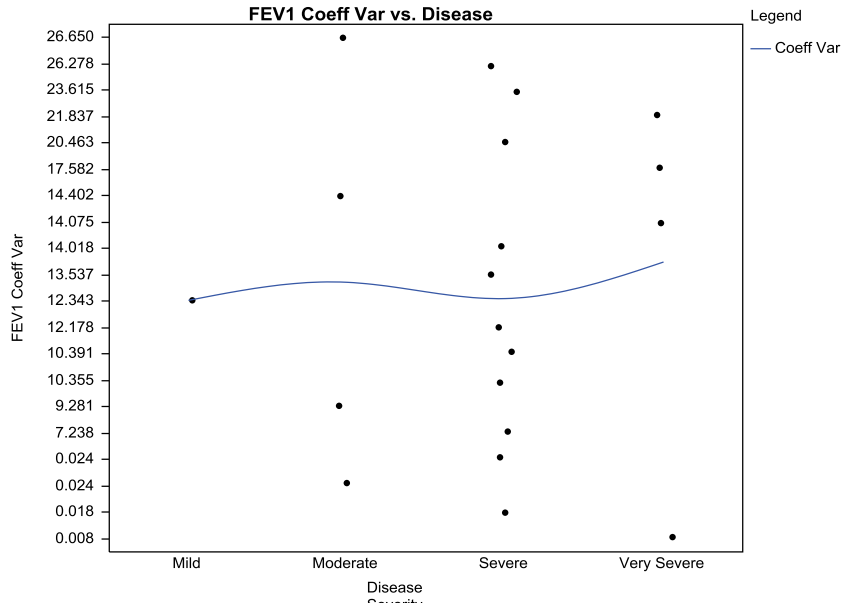


Figure 30. All Patients' FEV1 Standard Error vs. Disease Severity. Actual values plus smoother. *z* = 1.833, Chi-Square = 3.522 indicating significant difference.



.

Figure 31. All Patients' FEV1 Coefficient of Variation vs. disease severity. Actual values plus smoother. z = 0, Chi-Square = 0.002 indicating no significant difference.

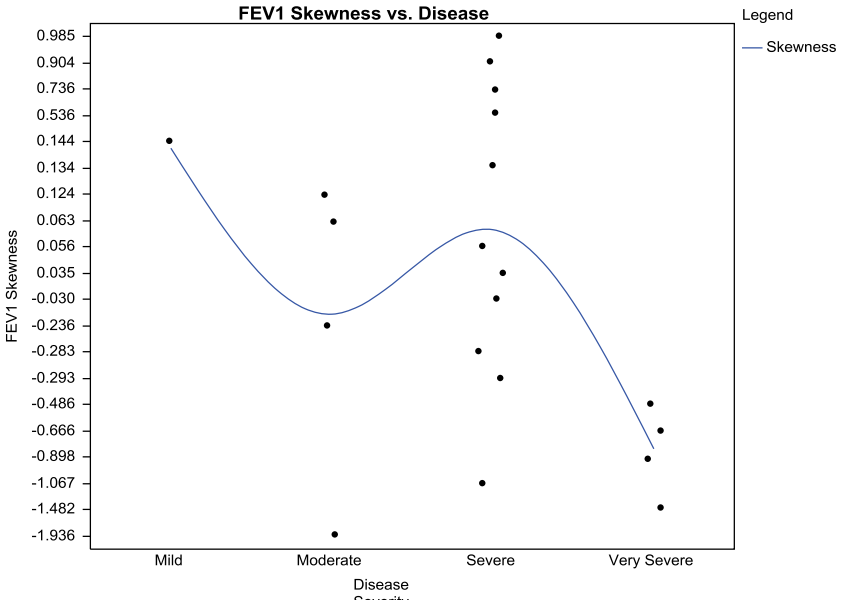


Figure 32. All Patients' FEV1 Skewness vs. Disease Severity. Actual values plus smoother. z = 0, Chi-Square = 0.002 indicating no significant difference.

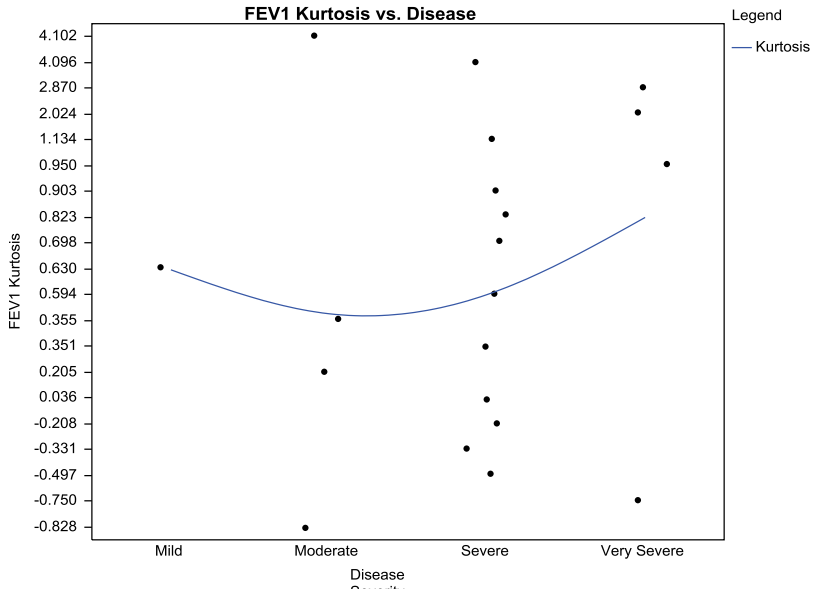


Figure 33. All Patients' FEV1 Kurtosis vs. Disease Severity. Actual values plus smoother. z = 0.348, Chi-Square = 0.154 indicating no significant difference.

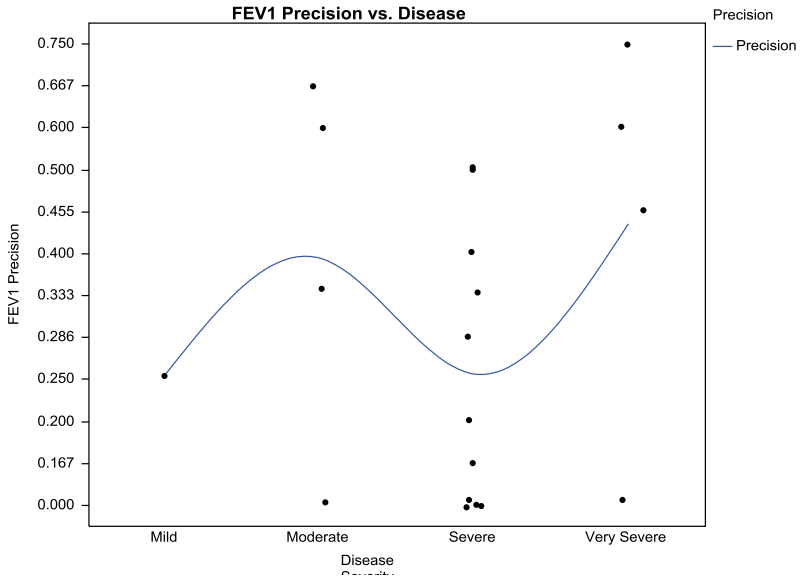


Figure 34. All patients' FEV1 Precision vs. Disease Severity. Actual values plus smoother. Precision = TP/(TP+FP), and represents positive predictive value (PPV) and quality of exactness. z = 0.664, Chi-Square = 0.502 indicating no significant difference.

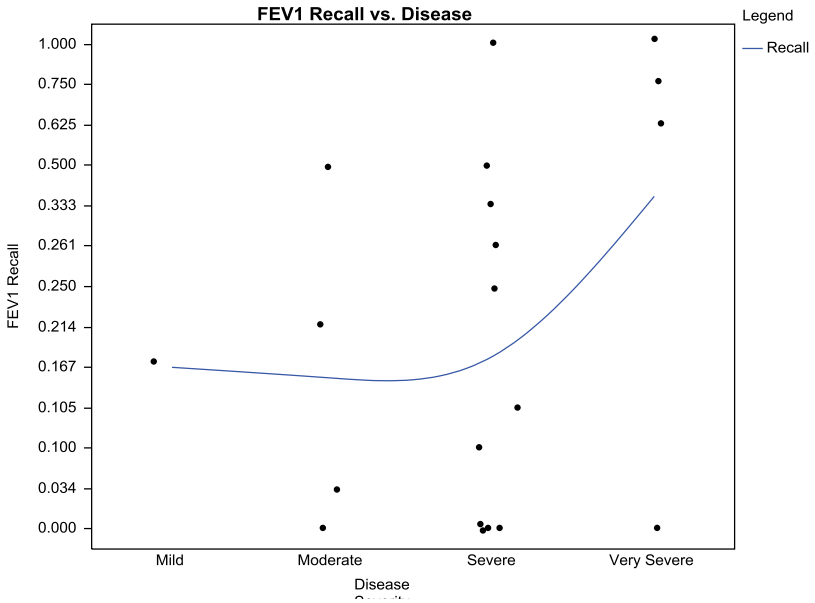


Figure 35. All Patients' FEV1 Recall vs. Disease Severity. Actual values plus smoother. Recall = TP/(TP+FN), and represents quality of completeness and sensitivity. *z* = -0.443, Chi-Square = 0.237 indicating no significant difference.

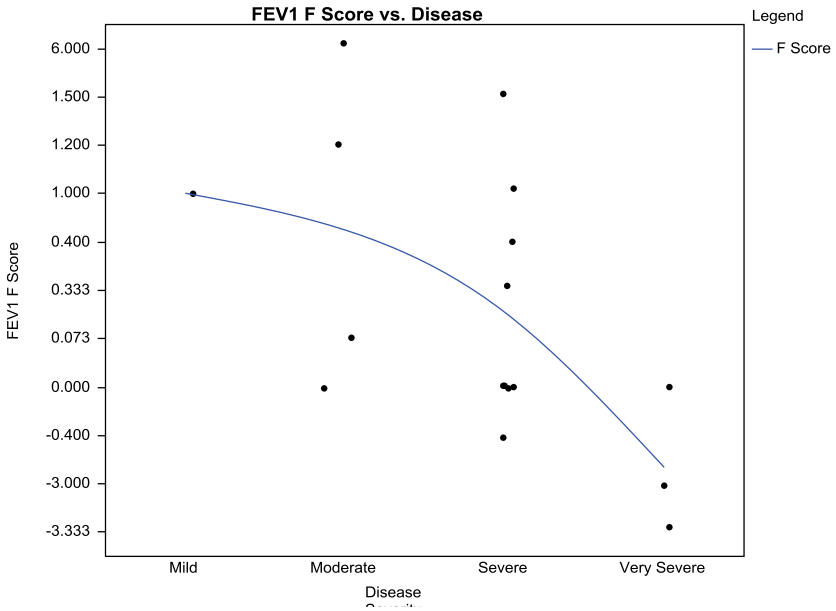


Figure 36. All Patients' FEV1 F-score vs. Disease severity. Actual values plus smoother. F-score = 2TP/(TP+FN)-(TP+FP). *z* = 1.779, Chi-Square = 3.359 indicating a significant difference; however, three patients (#3, 5, and 18) have zero increase and no F-score could be calculated resulting in inconclusive significance despite the *z* and Chi-Square values.

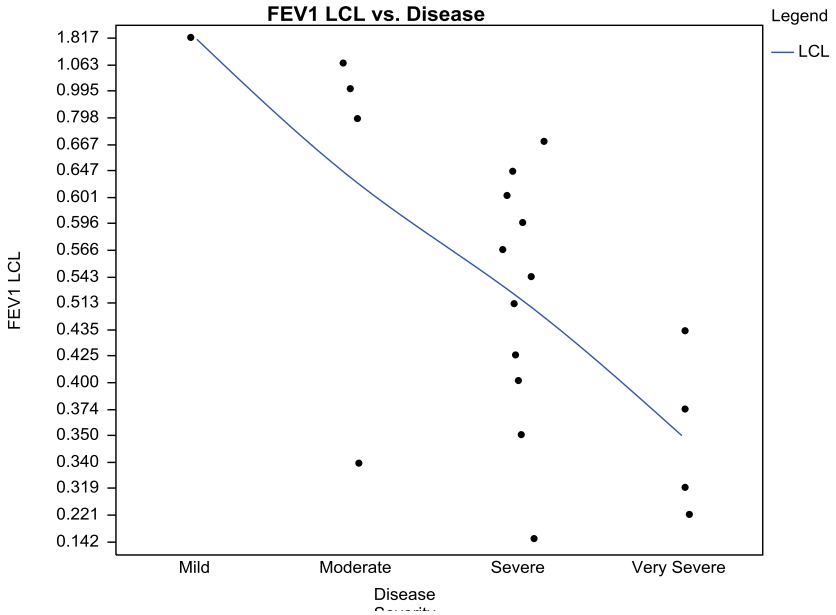


Figure 37. All Patients' FEV1 LCL vs. Disease Severity. LCL = lower control limit or -3σ. Actual values plus smoother.

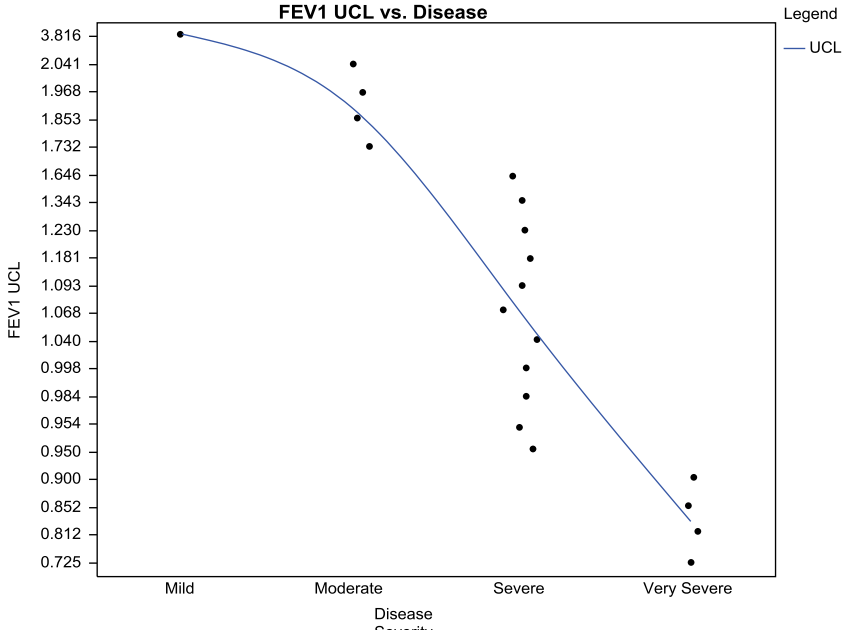


Figure 38. All Patients' FEV1 UCL vs. Disease Severity. UCL = Upper Control Limit, or +3σ. Actual values plus smoother.

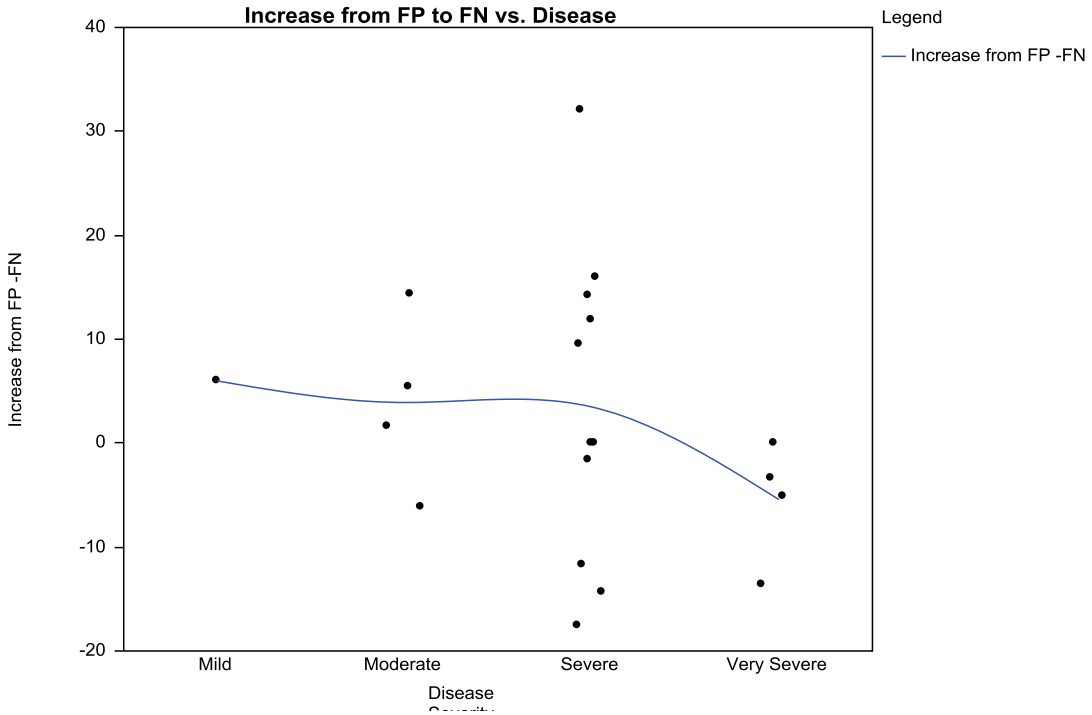


Figure 39. All Patients' FEV1 FP to FN Increase vs. Disease Severity. Mean % increase from FP to FN. Actual values plus smoother. *z* = .7867, Chi-Square = 0.690 indicating significant difference. Mean of mild/moderate = 4.277 and severe/very severe = 1.126. Standard deviation of mild/moderate = 7.452 and severe/very severe = 13.552.

Table 13. All Patients' FEV1 Disease Severity with Control Limits. Patient #, N, UCL (Upper Control Limit), LCL (Lower Control Limit), mean, and disease severity. Summary statistics are noted below the values (SUM = summation; AVG = average).

Patient #	N	UCL	LCL	Mean	Disease Severity
1	126	1.181	0.667	0.924	Severe
2	67	0.950	0.543	0.747	Severe
3	81	1.093	0.601	0.847	Severe
4	70	0.998	0.400	0.699	Severe
5	77	0.852	0.319	0.586	Very Severe
6	70	0.954	0.513	0.733	Severe
7	60	0.812	0.221	0.517	Very Severe
8	57	1.230	0.350	0.790	Severe
9	66	3.816	1.817	2.817	Mild
10	78	1.646	0.142	0.752	Severe
11	49	1.853	1.063	1.458	Moderate
12	62	1.968	0.340	1.154	Moderate
13	382	1.732	0.798	1.265	Moderate
14	50	1.040	0.596	0.818	Severe
15	59	0.9	0.374	0.637	Very Severe
16	92	2.041	0.995	1.518	Moderate
17	60	0.725	0.435	0.58	Very Severe
18	59	1.343	0.425	0.884	Severe
19	64	0.984	0.647	0.815	Severe
20	43	1.068	0.566	0.817	Severe
Summary	1672	1.359	0.591	0.968	
Statistics	SUM	AVG	AVG	AVG	

Table 14. All Patients' FEV1 Disease Severity with Basic Stats. Patient #, N, variance, standard error, coefficient of variation, skewness, kurtosis, and disease severity. Summary statistics are noted below the values (SUM = summation; AVG = average).

Patient #	N	Variance	Std Error	Coeff Var	Skewness	Kurtosis	Disease Severity
1	126	0.016	0.011	13.537	0.985	1.134	Severe
2	67	0.011	0.013	14.018	-0.030	0.036	Severe
3	81	0.008	0.010	10.391	-0.293	-0.331	Severe
4	70	0.020	0.017	20.463	0.736	0.351	Severe
5	77	0.011	0.012	17.582	-0.666	2.870	Very Severe
6	70	0.006	0.009	10.355	-0.283	-0.497	Severe
7	60	0.013	0.015	21.837	-1.482	2.024	Very Severe
8	57	0.035	0.025	23.615	0.056	0.903	Severe
9	66	0.121	0.043	12.343	0.144	0.630	Mild
10	78	0.039	0.022	26.278	-1.067	4.096	Severe
11	49	0.018	0.019	9.281	0.063	-0.828	Moderate
12	62	0.095	0.039	26.650	-1.936	4.102	Moderate
13	382	0.033	0.009	14.402	0.124	0.205	Moderate
14	50	0.595	0.016	0.018	0.904	0.823	Severe
15	59	0.008	0.012	14.075	-0.486	-0.750	Very Severe
16	92	1.270	0.052	0.024	-0.236	0.355	Moderate
17	60	0.280	0.003	0.008	-0.898	0.950	Very Severe
18	59	0.830	0.035	0.024	0.035	0.594	Severe
19	64	0.003	0.007	7.238	0.536	0.698	Severe
20	43	0.010	0.015	12.178	0.134	-0.208	Severe
Summary	1672	0.171	0.019	12.716	-0.183	0.858	
Statistics	SUM	AVG	AVG	AVG	AVG	AVG	

Table 15. All Patients' FEV1 Disease Severity with ROC Data. Patient #, N, increase from FP to FN, precision, recall, F-score, and disease severity. Summary statistics are noted below the values (SUM = summation; AVG = average).

Patient #	N	Increase from FP to FN	Precision (TP/(TP+FP)	Recall (TP/(TP+FN)	F-score	Disease Severity
1	126	9.524	0.286	0.105	0.333	Severe
2	67	11.940	0.400	0.261	1.500	Severe
3	81	0.000	0.333	0.333	#DIV/0!	Severe
4	70	14.286	0.200	0.100	0.400	Severe
5	77	0.000	0.750	0.750	#DIV/0!	Very Severe
6	70	-14.286	0.000	0.000	0.000	Severe
7	60	-3.333	0.600	1.000	-3.000	Very Severe
8	57	-17.544	0.167	1.000	-0.400	Severe
9	66	6.061	0.250	0.167	1.000	Mild
10	78	32.051	0.000	0.000	0.000	Severe
11	49	-6.122	0.000	0.000	0.000	Moderate
12	62	1.613	0.600	0.500	6.000	Moderate
13	382	14.398	0.667	0.034	0.073	Moderate
14	50	16.000	0.500	0.250	1.000	Severe
15	59	-13.559	0.000	0.000	0.000	Very Severe
16	92	5.435	0.333	0.214	1.200	Moderate
17	60	-5.000	0.455	0.625	-3.333	Very Severe
18	59	0.000	0.500	0.500	#DIV/0!	Severe
19	64	-1.563	0.000	0.000	0.000	Severe
20	43	-11.628	0.000	0.000	0.000	Severe
Summary	1672	1.914	0.302	0.292		
Statistics	SUM	AVG	AVG	AVG		

Each patients' FEV1 data was subjected to Shewart-Western Electric analysis, with A, B, C zones (+/- 3σ, 2σ, and 1σ, respectively) and rule violated annotated (Figures 40 -59).

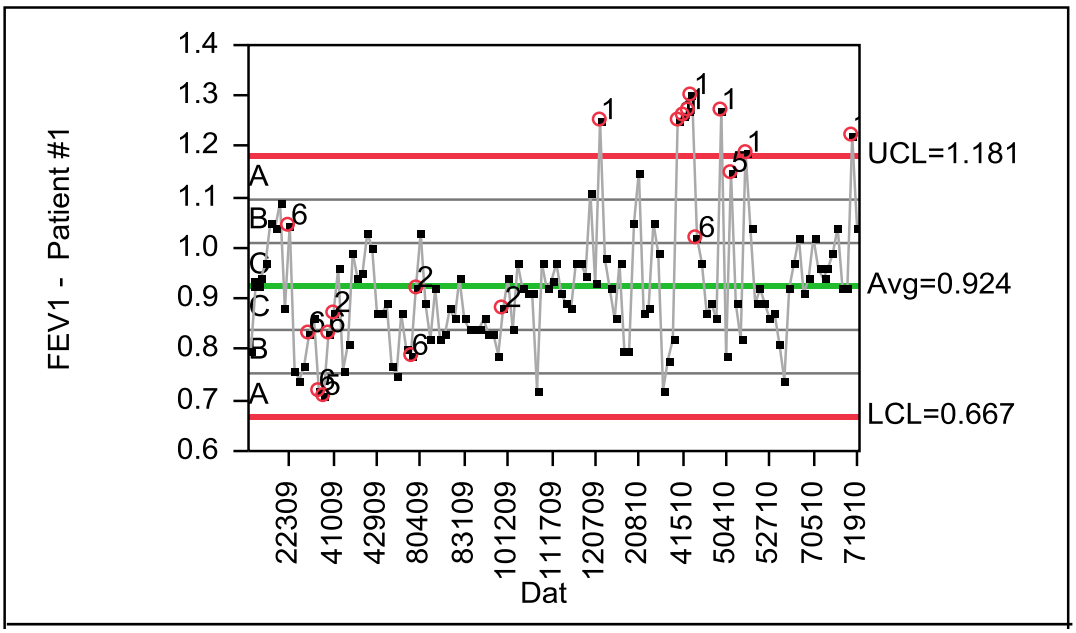


Figure 40. Patient #1 FEV1 Shewart-Western Electric Analysis. (Note additional data compared to Beta Test.)

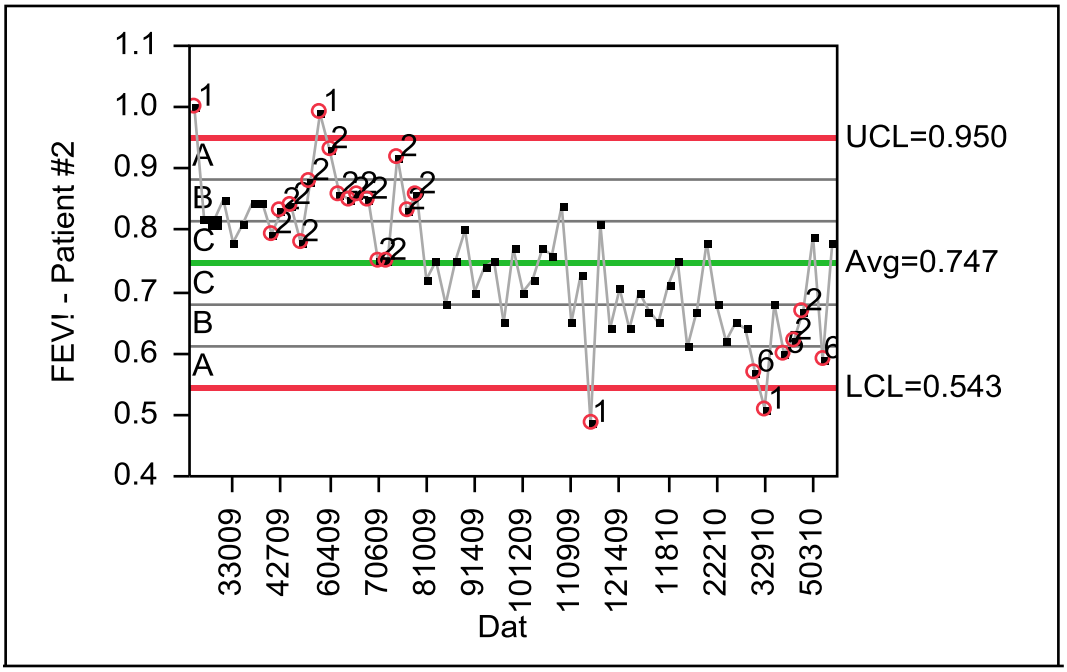


Figure 41. Patient #2 FEV1 Shewart-Western Electric Analysis.

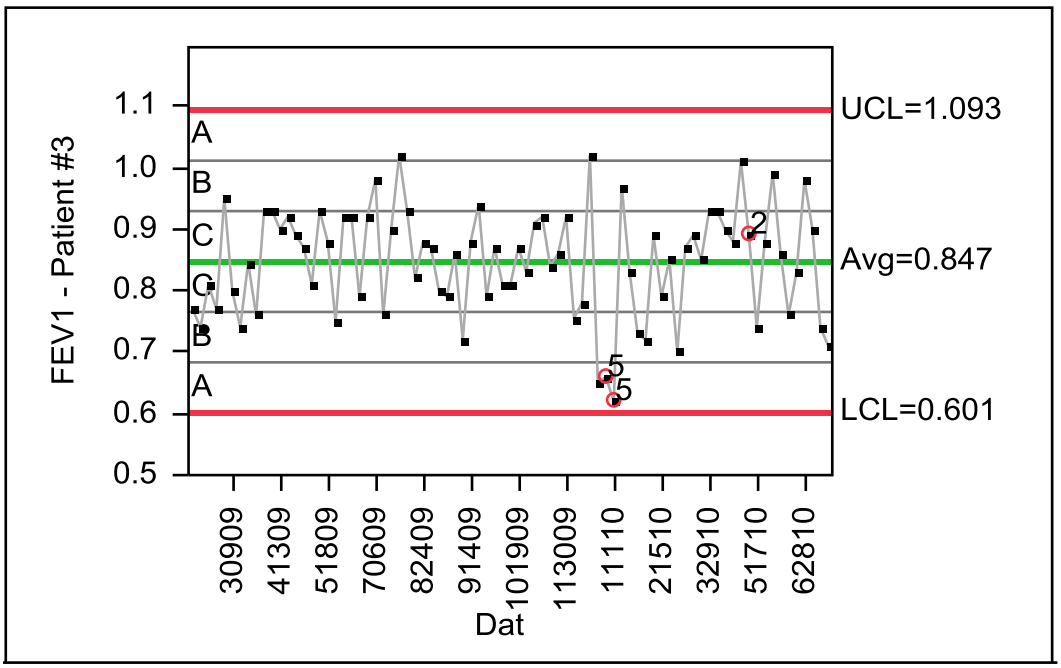


Figure 42. Patient #3 FEV1 Shewart-Western Electric Analysis.

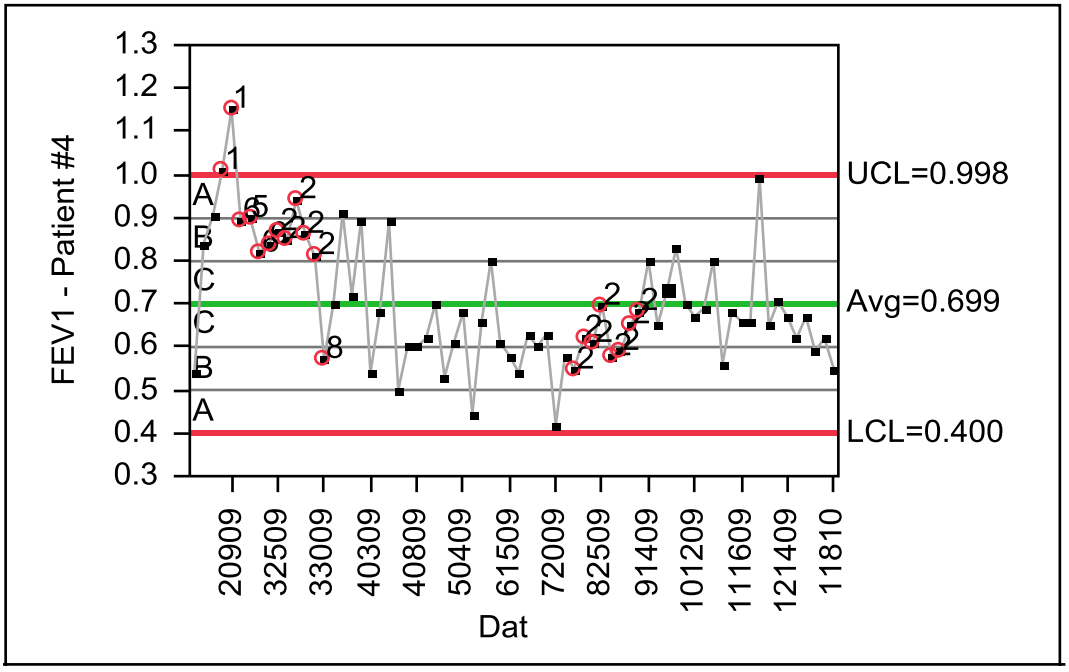


Figure 43. Patient #4 FEV1 Shewart-Western Electric Analysis.

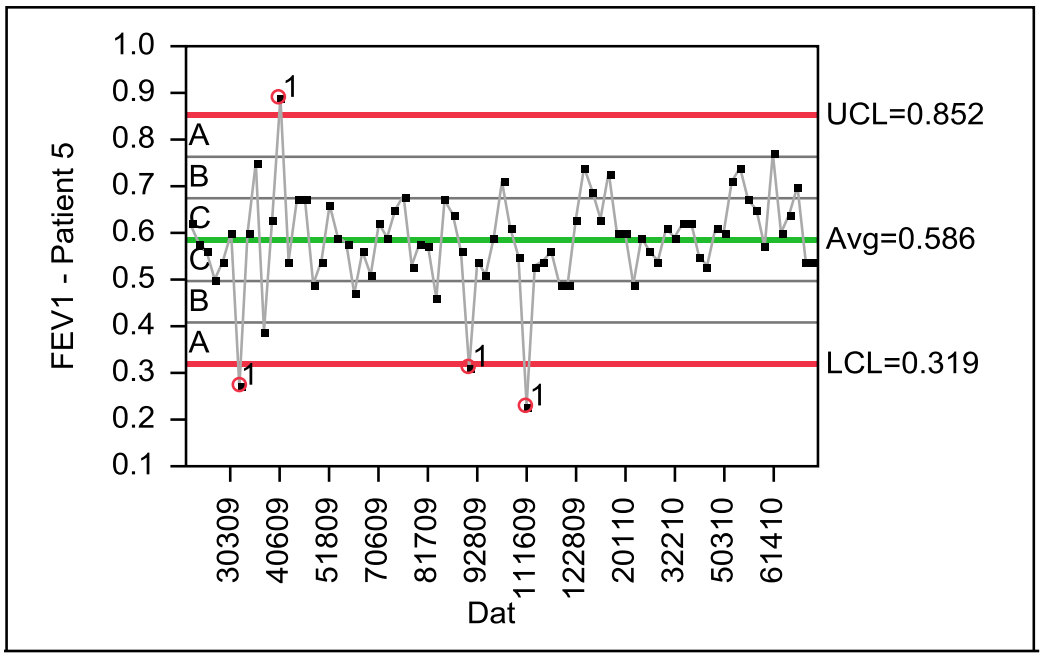


Figure 44. Patient #5 FEV1 Shewart-Western Electric Analysis.

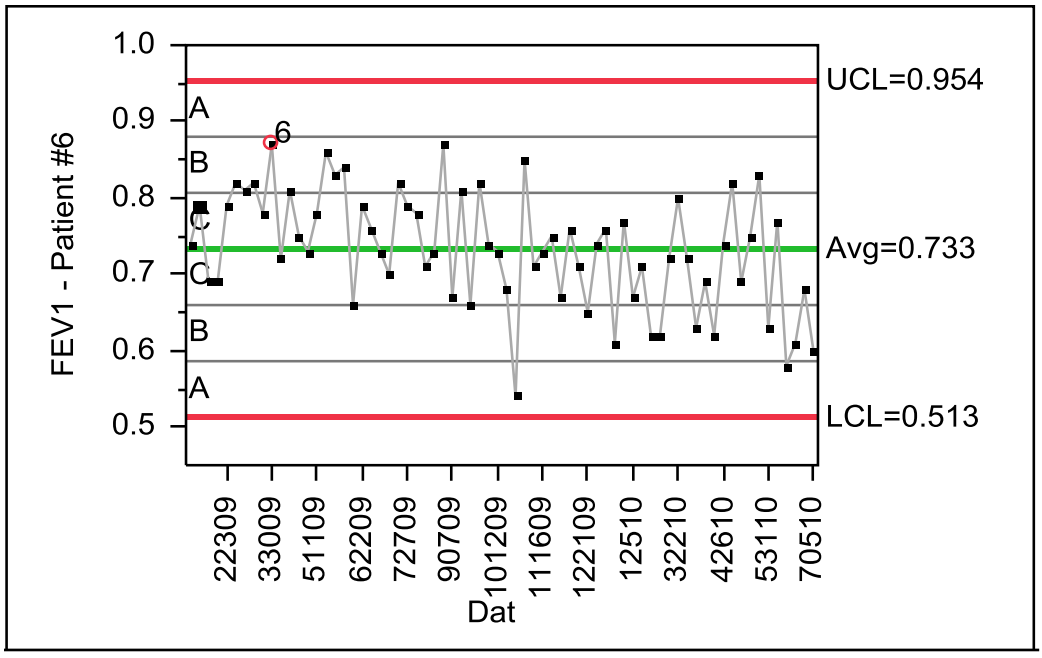


Figure 45. Patient #6 FEV1 Shewart-Western Electric Analysis.

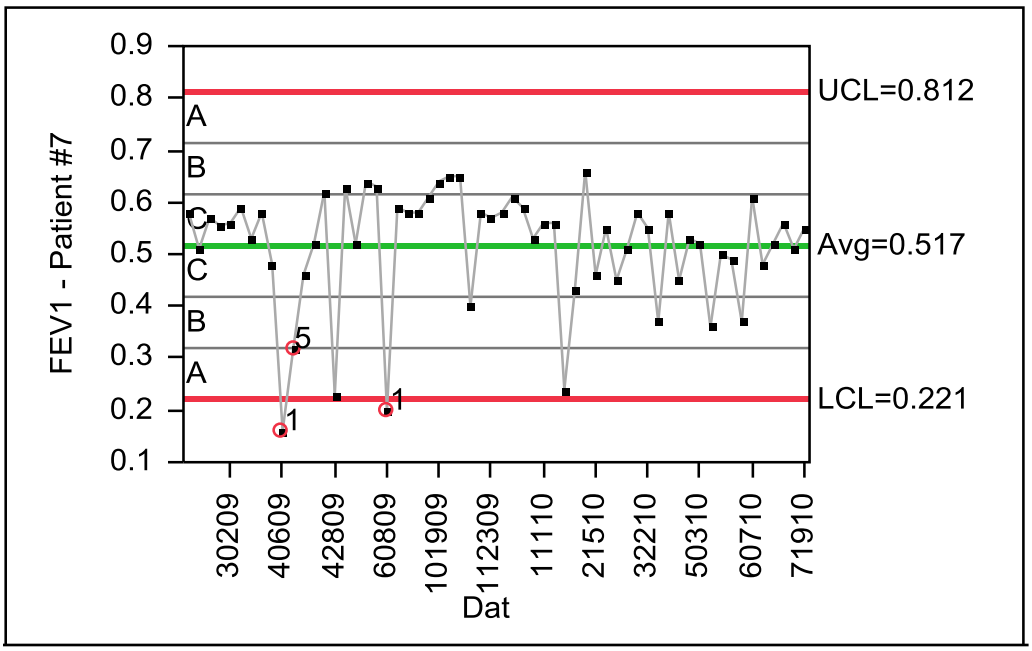


Figure 46. Patient #7 FEV1 Shewart-Western Electric Analysis.

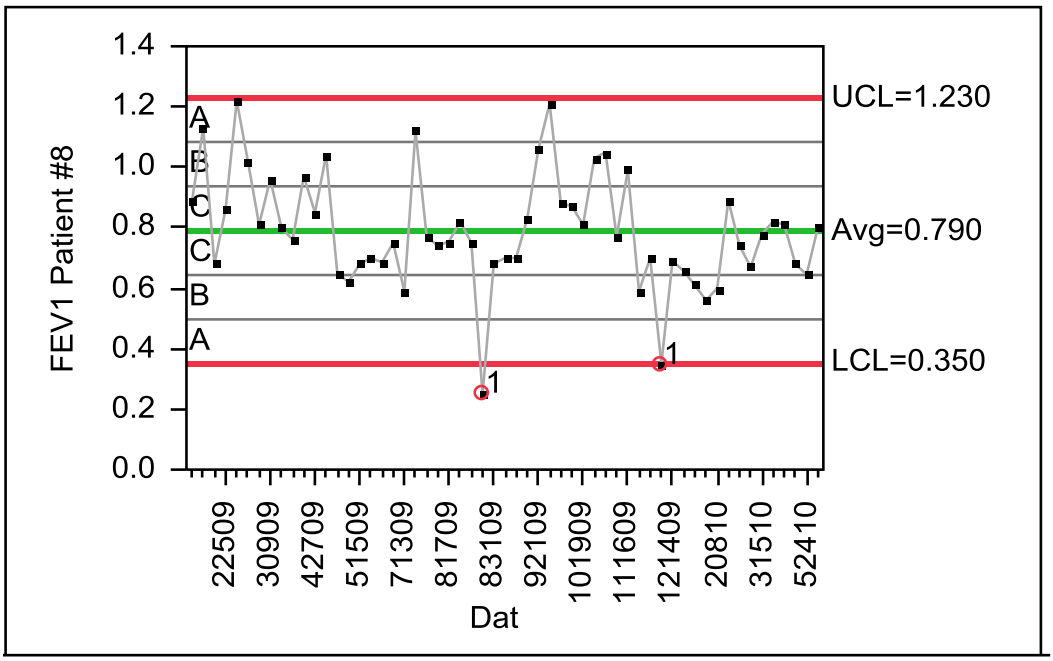


Figure 47. Patient #8 FEV1 Shewart-Western Electric Analysis.

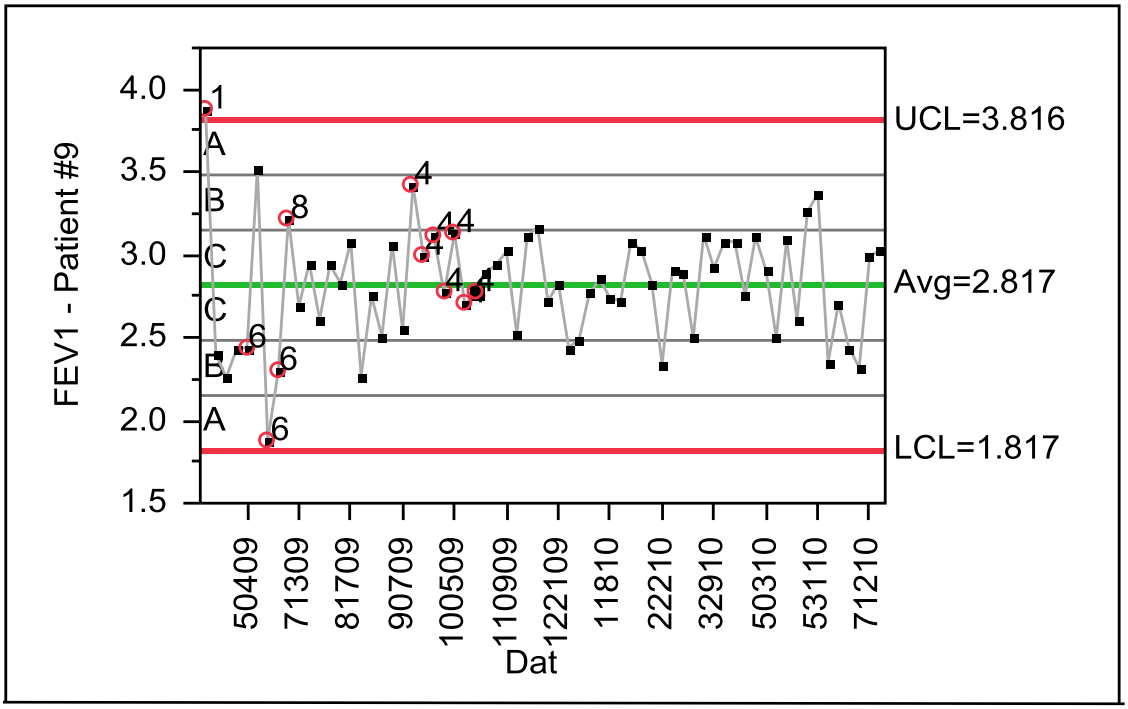


Figure 48. Patient #9 FEV1 Shewart-Western Electric Analysis.

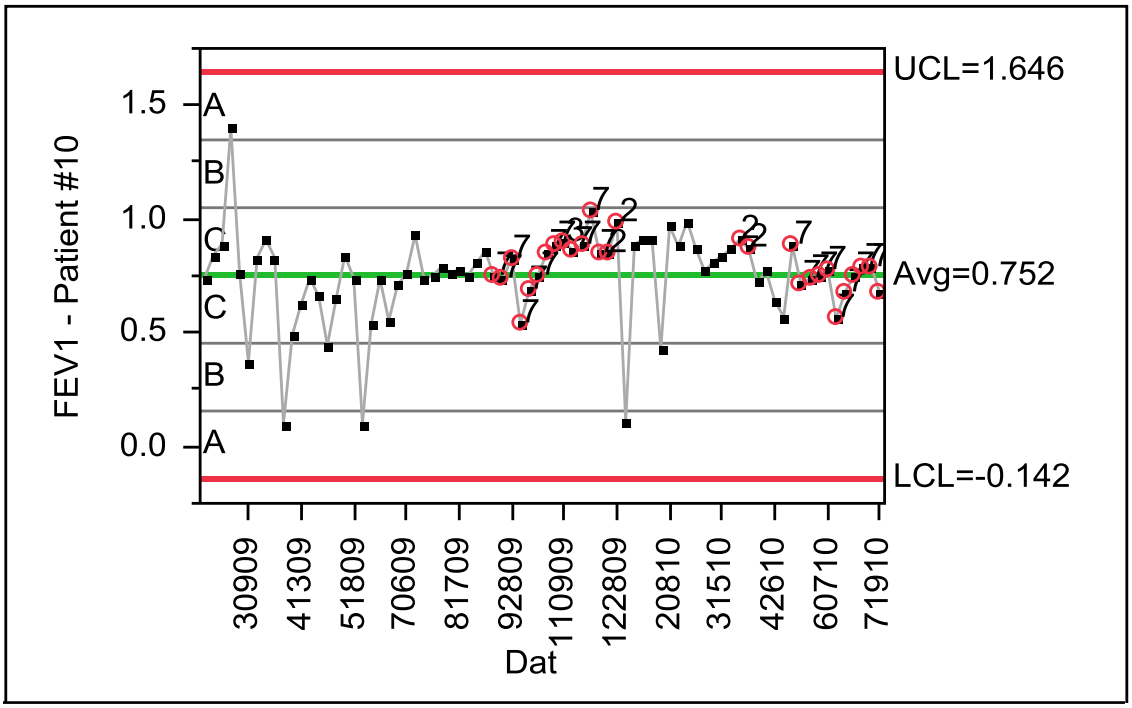


Figure 49. Patient #10 FEV1 Shewart-Western Electric Analysis.

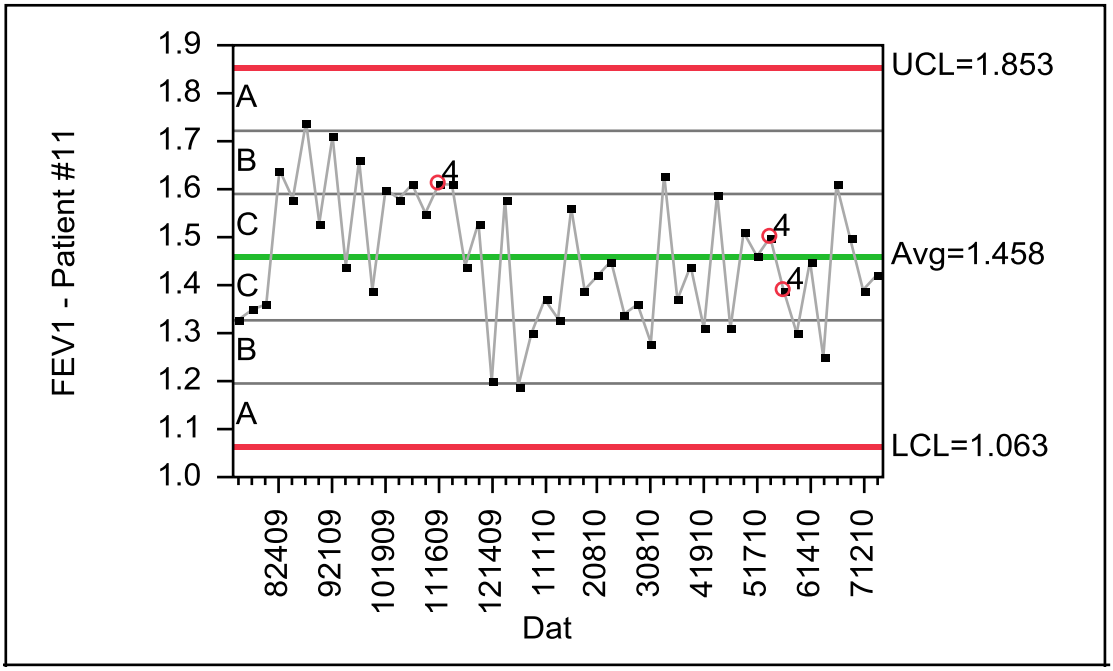


Figure 50. Patient #11 FEV1 Shewart-Western Electric Analysis.

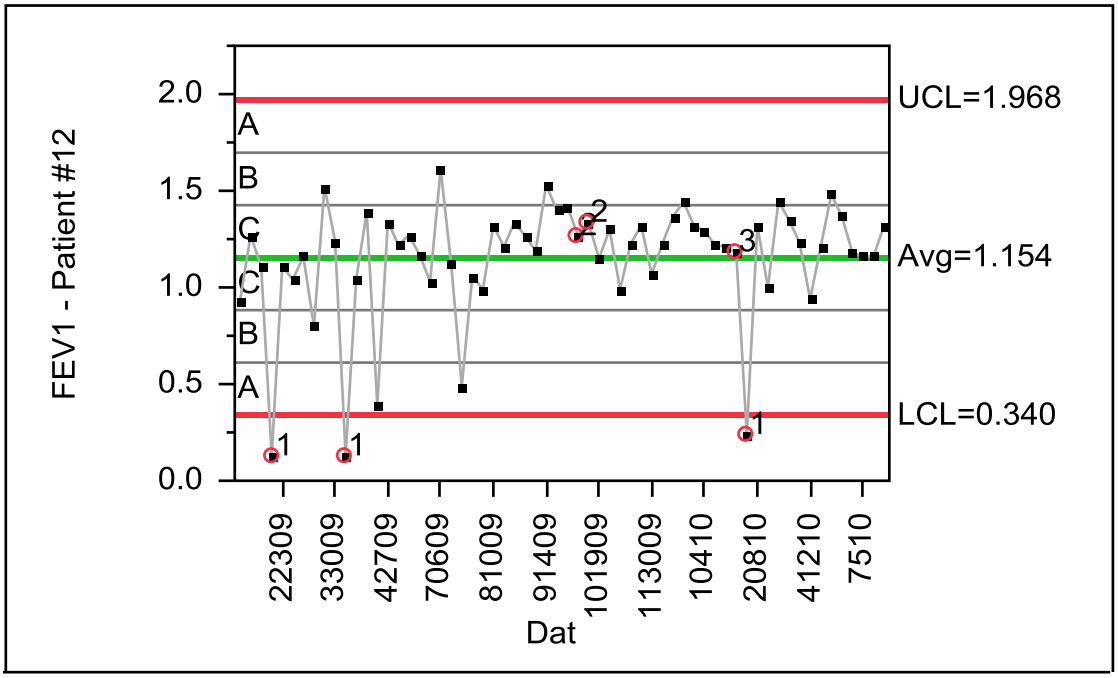


Figure 51. Patient #12 FEV1 Shewart-Western Electric Analysis.

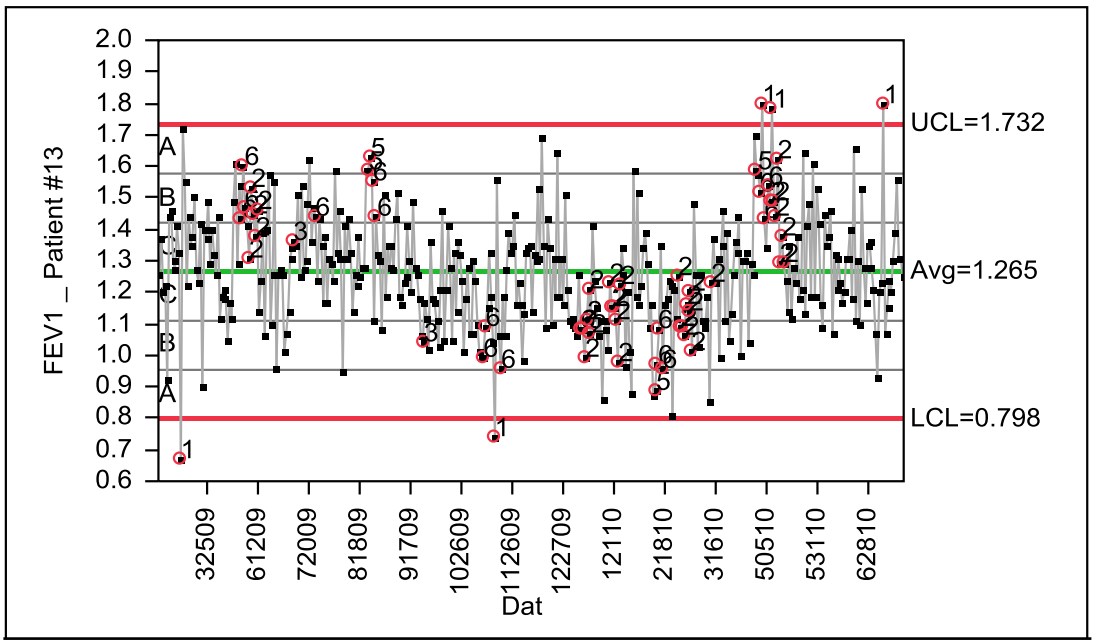


Figure 52. Patient #13 FEV1 Shewart-Western Electric Analysis.

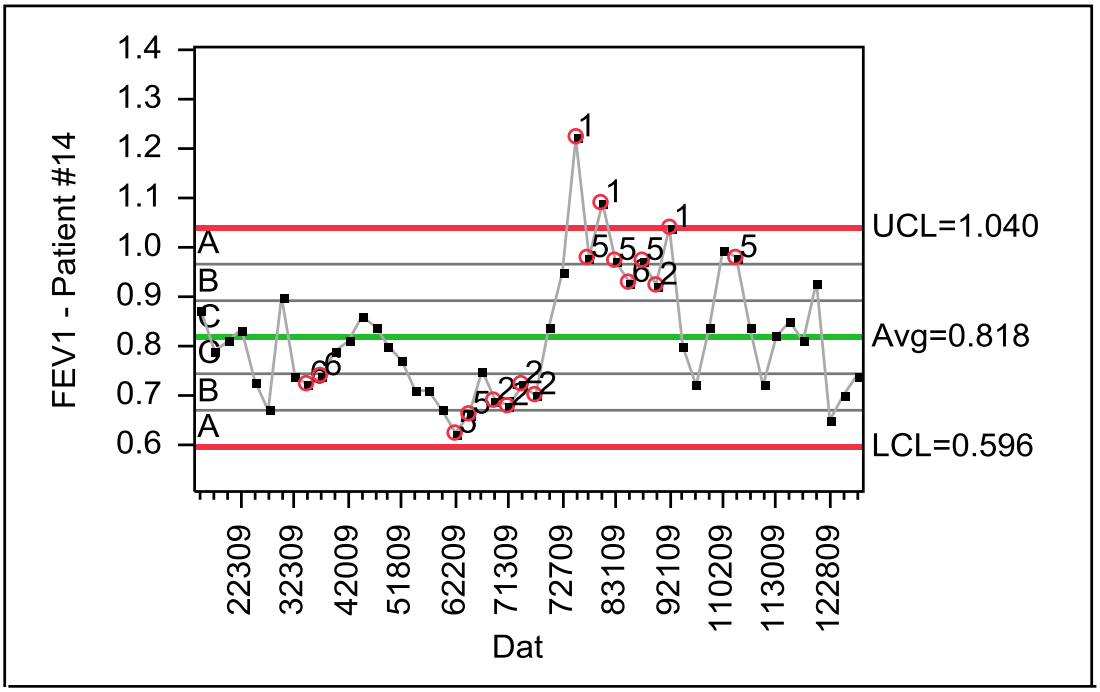


Figure 53. Patient #14 FEV1 Shewart-Western Electric Analysis.

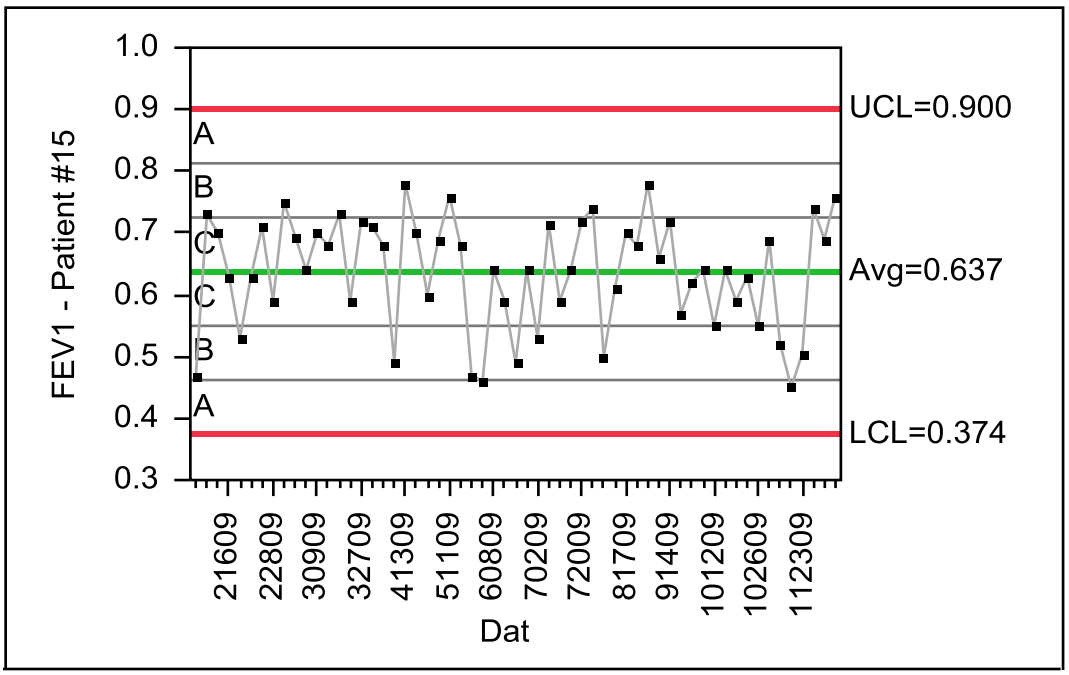


Figure 54. Patient #15 FEV1 Shewart-Western Electric Analysis.

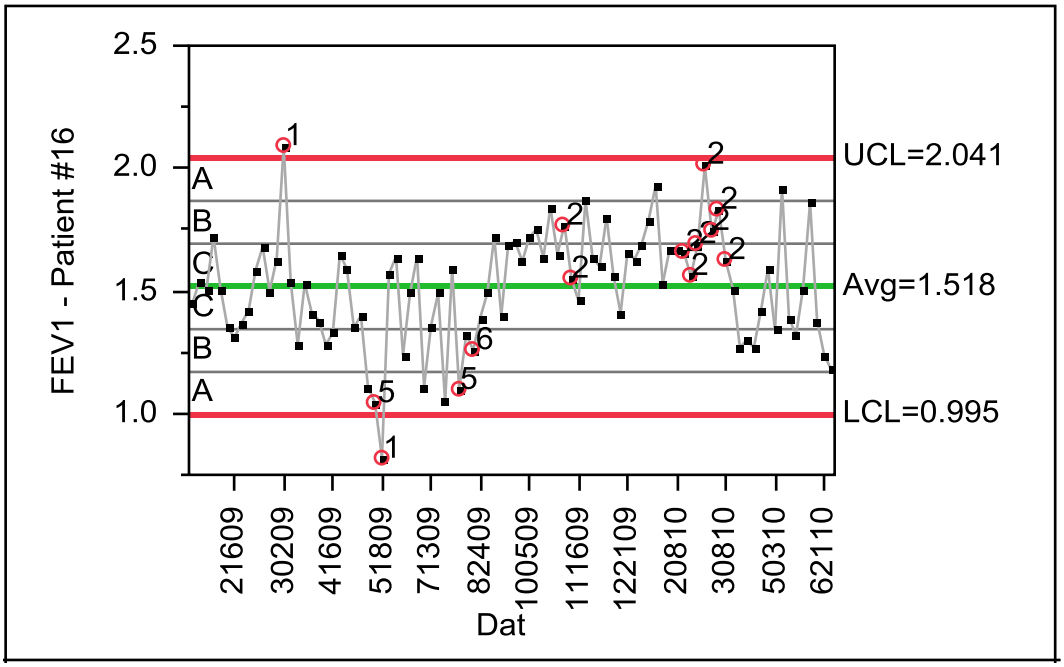


Figure 55. Patient #16 FEV1 Shewart-Western Electric Analysis.

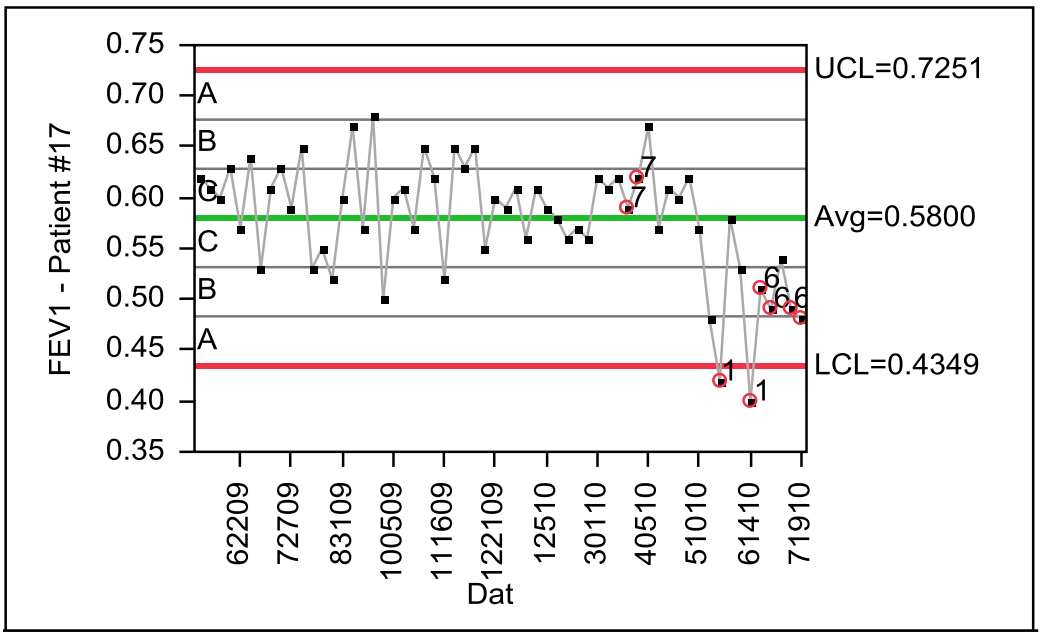


Figure 56. Patient #17 FEV1 Shewart-Western Electric Analysis

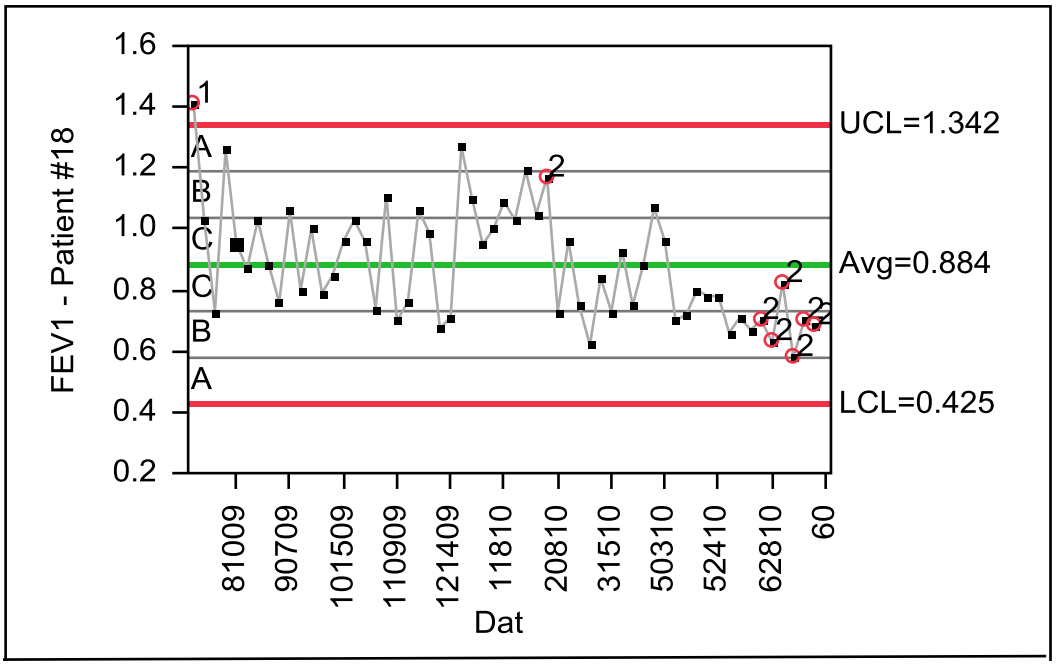


Figure 57. Patient #18 FEV1 Shewart-Western Electric Analysis.

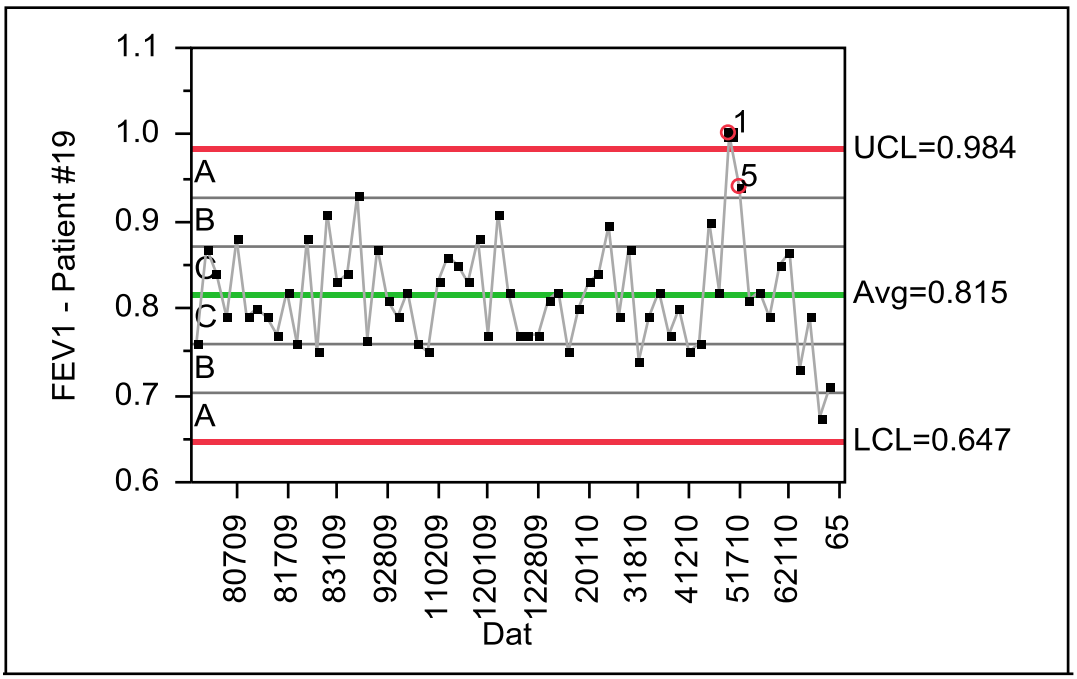


Figure 58. Patient #19 FEV1 Shewart-Western Electric Analysis.

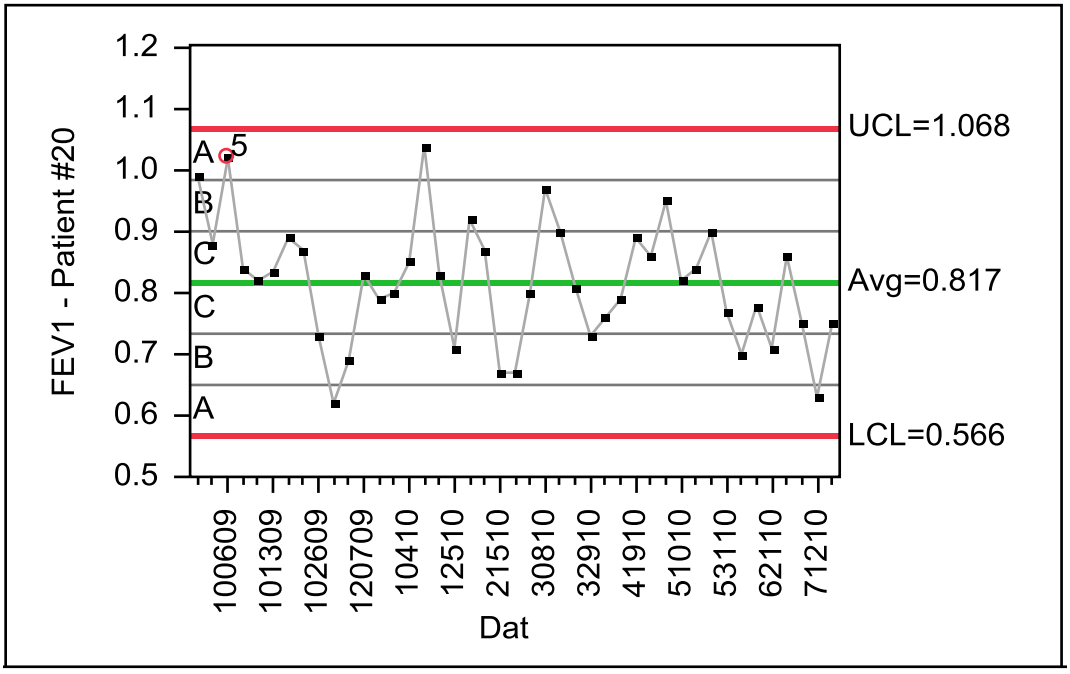


Figure 59. Patient #20 FEV1 Shewart-Western Electric Analysis.

Each patient who had a statistically-derived threshold violation had a corresponding Western Electric Rule violation, per Table 7. Figure 60, Figure 61, and Table 16 indicate that Western Electric Rule 2, followed by Rules 1 and 6, are the most prevalent. Rule 2 detects a shift in the process mean. Rule 1 also detects a shift in the mean, an increase in the standard deviation, or a single aberration in the process (e.g., something unusual happened to the patient, such as mowing grass causing a COPD episode). Rule 6 also detects a shift in the mean; any four out of five points provide a positive test for Rule 6.

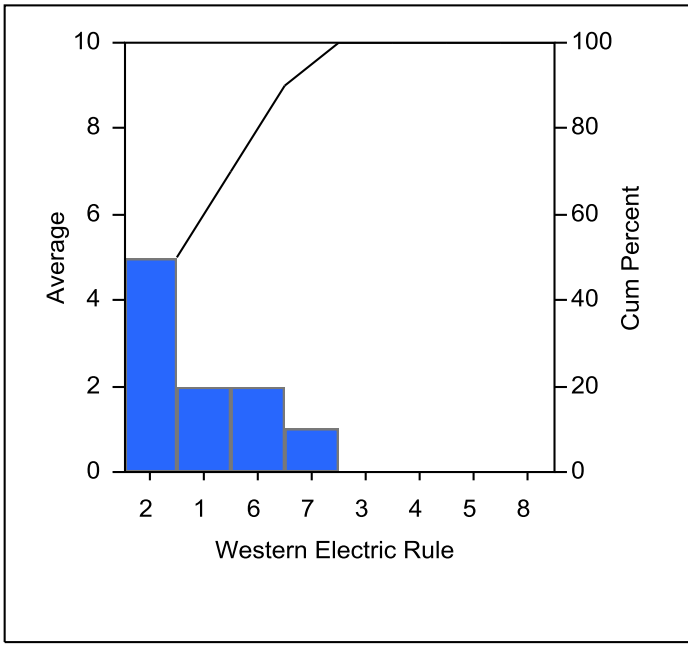


Figure 60. All Patients' FEV1 Pareto of Western Electric Rules. Patients #1-20 FEV1 Pareto of average occurrence for a specific Western Electric Rule (see Table 7).

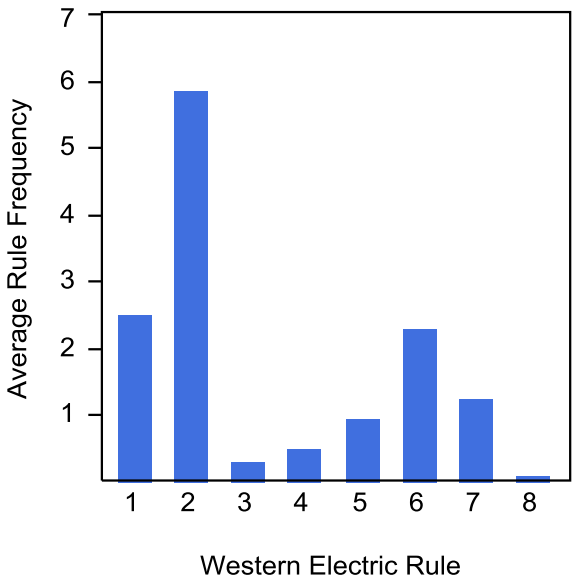


Figure 61. All Patients' FEV1 Average Rule Frequency. Classification of Western Electric Rules versus the average frequency of occurrence in the population of 20 patients.

Table 16. Western Electric Rule Violations. Pre-predictive Algorithm. Note the high degree of process mean shift Rule violations in Rules 2, 1, and 6. Summary statistics are noted below the values (AVG = average).

Patient #	Rule 1	Rule 2	Rule 3	Rule 4	Rule 5	Rule 6	Rule 7	Rule 8
1	8	3	0	0	2	6	0	0
2	3	17	0	0	1	2	0	0
3	0	1	0	0	2	0	0	0
4	2	13	0	0	1	3	0	1
5	4	0	0	0	0	0	0	0
6	0	0	0	0	0	1	0	0
7	2	0	0	0	1	0	0	0
8	2	0	0	0	0	0	0	0
9	1	0	0	7	0	3	0	1
10	0	11	1	0	0	0	23	0
11	0	0	0	3	0	0	0	0
12	4	2	1	0	0	0	0	0
13	5	30	2	0	4	17	0	0
14	3	5	0	0	6	3	0	0
15	0	0	0	0	0	0	0	0
16	13	28	2	0	0	6	0	0
17	1	0	0	0	0	5	2	0
18	1	7	0	0	0	0	0	0
19	1	0	0	0	1	0	0	0
20	0	0	0	0	1	0	0	0
Summary	2.5	5.85	0.3	0.5	0.95	2.3	1.25	0.1
Statistics	AVG	AVG	AVG	AVG	AVG	AVG	AVG	AVG

Section 5. Final Construct: Classifier Design

Patients' classifiers are considered TP, FN, FP, and TN. Each patient's data points have been classified per Table 12.

Section 6. Final Construct: System Evaluation

In order to develop a model for prediction that was general enough for a variety of patients' FEV1 values, a random K-fold was investigated. Random K-fold is a general technique specifically targeted for small samples and serves as a cross-validation technique. In a neural net platform, the data are randomly partitioned into "K" equal subsamples. One subsample is specific for the validation group, and "K-1" subsamples are specific for the training group. The cross-validation is then repeated "K" times and averaged to provide the best estimate. The activation functions are applied (i.e., tanh, linear, and/or Gaussian) at the layers, and a logistic transformation is applied at the response (responses are TP, FN, FP, TN in this study).

A random K-fold with an activation node of 1 Gaussian was applied due to its efficiency with small sample sizes. Unfortunately, as can be seen in Table 17, this model was not adequate (i.e., specific aim 2, hypothesis B, Generalized R-square>.80) in predicting FEV1.

Table 17. Results of Random K-fold with 1 Gaussian Nodes. All patients' FEV1 subjected to random K-fold (5 folds), with 1 Gaussian activation node, 1 activation layer, and no boosting or trans-covariates. Patient #, training Generalized R-square, validation R-square, # of misclassifications and training misclassification rate are presented. Summary statistics are noted below the values (AVG = average).

Patient #	Training Gen R-square	Validation Gen R-square	# of Misclassifications
1	0.786	0.860	11.000
2	0.845	0.864	17.000
3	0.209	-0.006	4.000
4	0.636	0.514	15.000
5	1.000	1.000	0.000
6	0.970	1.000	1.000
7	0.440	0.818	3.000
8	1.000	1.000	0.000
9	0.869	0.700	6.000
10	0.641	0.704	22.000
11	0.803	0.591	2.000
12	0.512	0.630	6.000
13	0.430	0.489	36.000
14	0.682	0.574	9.000
15	0.740	0.598	1.000
16	0.611	0.853	9.000
17	0.700	0.220	5.000
18	0.746	0.766	5.000
19	0.868	0.889	2.000
20	0.871	0.986	1.000
Summary	0.718	0.702	7.750
Statistics	AVG	AVG	AVG

Varying combinations of linear, tanh, and Gaussian activation nodes were considered. All had less than Generalized R-square values desired per hypothesis B. A random K-fold (5-folds) model was initiated with 3 tanh and 3 Gaussian activation nodes. One activation layer (no hidden layers), one tour, and a squared penalty (to mitigate the tendency for the neural network to over-fit the data) were applied. No boosting was

employed. Covariates were transformed (fitting option). The data was randomly divided by the software into training (estimates model parameters) and validation (validates predictive ability of the model) sets, approximately 75% and 25%, respectively. Table 18 summarizes the results.

Table 18. Results of Random K-fold with 3 tanh and 3 Gaussian Nodes. All patients' FEV1 subjected to random K-fold (5 folds), with 3 tanh and 3 Gaussian activation nodes. One layer (no hidden layers), one tour, transformed covariates with a squared penalty were applied. No boosting was employed. Patient #, training Generalized R-square, validation Generalized R-square, # of misclassifications, and training misclassification rate are presented. Summary statistics are noted below the values (AVG = average).

Patient #	Training Gen R-square	Validation Gen R-square	# of Misclassifi-cations	Training Misclass Rate
1	0.791	0.569	7	0.069
2	0.900	0.826	9	0.167
3	0.769	1.000	2	0.031
4	0.693	0.414	12	0.286
5	1.000	1.000	0	0.000
6	0.973	0.995	1	0.018
7	0.482	0.762	3	0.063
8	1.000	1.000	0	0.000
9	0.683	0.661	8	0.151
10	0.359	0.581	20	0.262
11	0.809	0.965	3	0.075
12	0.636	0.868	4	0.080
13	0.367	0.156	44	0.144
14	0.883	0.853	8	0.200
15	0.934	0.936	1	0.026
16	0.799	0.786	7	0.096
17	0.729	0.773	7	0.146
18	0.875	0.938	4	0.083
19	0.995	0.998	0	0.000
20	0.957	1.000	1	0.029
Summary	0.782	0.804	7.05	0.096
Statistics	AVG	AVG	AVG	AVG

Figure 62 diagrams the model. Neural networks randomly use different starting points, and select the iteration with the best validation. The patients were approximately evenly divided between the radial Gaussian and sigmoid tanh for the transformation formula (Table 19) and no significant correlations were found between the transformation formula type and other patient values presented.

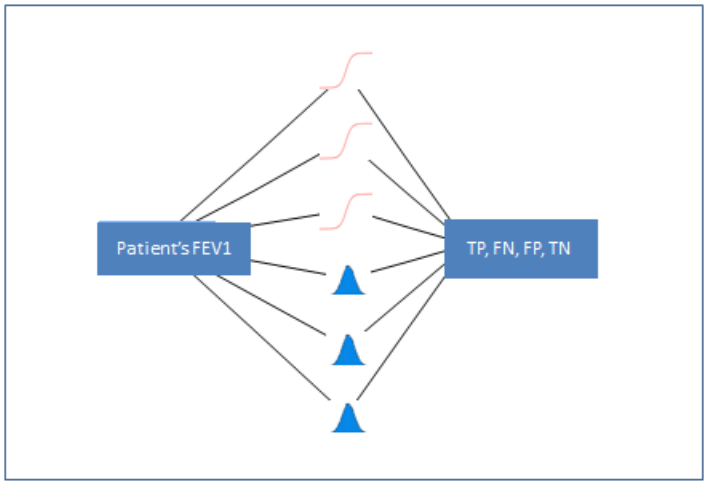


Figure 62. Neural Net Diagram. Diagram of tanh and Gaussian activation nodes in the random K-fold neural net.

Table 19. All Patients' FEV1 Transformation Formulae. Patient #, training Generalized R-square, validation Generalized R-square, # of misclassifications, and training misclassification rate are also presented. Summary statistics are noted below the values (AVG = average).

Patient #	Training Gen R-square	Validation Gen R-square	# of Misclassifi-cations	Training Misclass Rate	Transformation Formula
1	0.791	0.569	7	0.069	ArcSinH
2	0.900	0.826	9	0.167	Log
3	0.769	1.000	2	0.031	Log
4	0.693	0.414	12	0.286	Log
5	1.000	1.000	0	0.000	ArcSinH
6	0.973	0.995	1	0.018	Log
7	0.482	0.762	3	0.063	ArcSinH
8	1.000	1.000	0	0.000	ArcSinH
9	0.683	0.661	8	0.151	Log
10	0.359	0.581	20	0.262	ArcSinH
11	0.809	0.965	3	0.075	Log
12	0.636	0.868	4	0.080	ArcSinH
13	0.367	0.156	44	0.144	ArcSinH
14	0.883	0.853	8	0.200	Log
15	0.934	0.936	1	0.026	Log
16	0.799	0.786	7	0.096	ArcSinH
17	0.729	0.773	7	0.146	ArcSinH
18	0.875	0.938	4	0.083	Log
19	0.995	0.998	0	0.000	ArcSinH
20	0.957	1.000	1	0.029	ArcSinH
Summary	0.782	0.804	7.05	0.096	
Statistics	AVG	AVG	AVG	AVG	

Table 20 details the shift from actual to predicted in this four-schema classification with the random K-fold applied the predicted values, and the confusion matrix of actual and predicted classification of TP, FN, FP, and TN. When the data is reviewed further (Table 21), of the 141 changes (from actual to predicted) of the total 1672 (8.43%[5]),

[5] This total is less than the 9.6% reported in Table 19 due to the Training Misclassification Rate being calculated on a smaller sample size (i.e., Training sample only).

there is no significant difference between the total number of changes to the predicted states of FN and FP. In eleven predicted instances, the clinician intervened but shouldn't have and in eleven additional predicted instances, the clinician did not intervene but should have based on the statistically-derived thresholds. Of note, however, is the high 108 predicted instances that changed from FN (did not intervene but should have) to TN (no intervention necessary). This represents approximately 77% of the total misclassifications. 62% of the FN to TN changes originated from only three of the 20 patients, Patients #4, 10, and 13. In reviewing these patients' analytics, there does not appear to be any cause and effect. Two of the three patients (Patient #10 and 13) had a significant number of outliers removed, but so did an additional two (Patient #1 and 3) that did not exhibit the high predicted change from FN to TN.

In some patients (#1, 2, 4, 9, 13, 14, and 16), it appears that the random k-fold has overfit the data, with the validation Generalized R-square lower than the training Generalized R-square. The squared penalty was intended to mitigate that risk, and in many cases, the data did not have overfitting.

Table 20. All Patients' FEV1 Actual to Predicted Counts.

Patient #	TN to FN	TP to FP	FN to TN	FN to FP	TP to TN	FP to TN	FP to TP	TN to FP	SUM
1	1	1	5	0	0	0	0	0	7
2	3	1	5	0	0	0	0	0	9
3	0	1	1	0	0	0	0	0	2
4	1	0	10	1	0	0	0	0	12
5	0	0	0	0	0	0	0	0	0
6	0	0	1	0	0	0	0	0	1
7	0	0	1	0	1	0	1	0	3
8	0	0	0	0	0	0	0	0	0
9	1	0	7	0	0	0	0	0	8
10	4	0	16	0	0	0	0	0	20
11	0	0	3	0	0	0	0	0	3
12	0	0	3	0	0	0	0	1	4
13	0	0	41	0	0	2	0	1	44
14	1	1	4	0	0	0	2	0	8
15	0	0	0	0	0	1	0	0	1
16	0	1	5	0	0	0	1	0	7
17	0	1	2	0	0	2	0	2	7
18	0	0	3	0	0	0	1	0	4
19	0	0	0	0	0	0	0	0	0
20	0	0	1	0	0	0	0	0	1
	0.55	0.3	5.4	0.05	0.05	0.25	0.25	0.2	
Summary	AVG	AVG	AVG	AVG	AVG	AVG	AVG	AVG	
Statistics	11	6	108	1	1	5	5	4	
	SUM	SUM	SUM	SUM	SUM	SUM	SUM	SUM	

Table 21. All Patients' FEV1 Frequency of Predicted State Changes.

TP		**FN**		**FP**		**TN**	
Act→Pred	Count	Act→Pred	Count	Act→Pred	Count	Act→Pred	Count
FP→TP	5	TN→FN	11	TP→FP	6	FN→TN	108
				FN→FP	1	TP→TN	1
				TN→FP	4	FP→TN	5
Totals:	5		11		11		114

Likely disease state of each patient of Table 13 was compared to their training and validation Generalized R-square in Table 19. As shown in Figures 63 and 64, a change from Mild or Moderate to Severe or Very Severe occurs in model fit. Because these two populations (Mild/Moderate and Severe/Very Severe) are not normal distributions, the non-parametric Wilcoxon/Kruskal-Wallis Tests were performed. In the training set, the Wilcoxon/Kruskal-Wallis Tests resulted in the z statistic (-1.746) and the Chi-Square (3.2043) indicating that, at $p<.05$, there is a significant difference between the two populations. Applying the Wilcoxon/Kruskal-Wallis Tests to the validation set also indicated that there is a significant difference between the two populations (z=1.138, and Chi-Square=1.399).

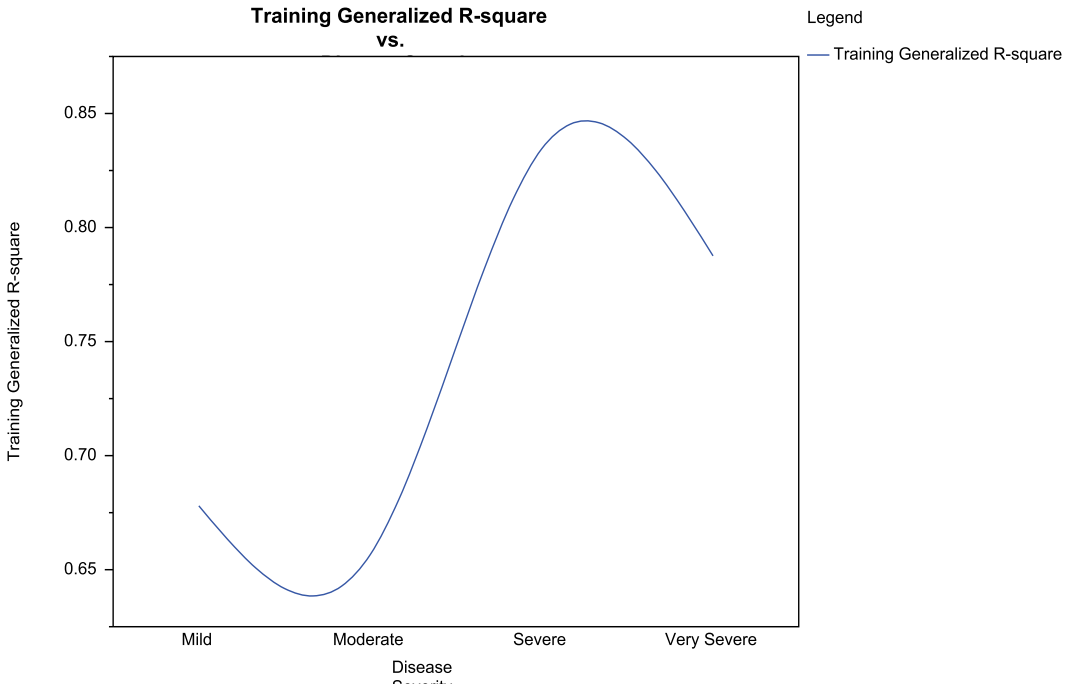


Figure 63. All Patients' Training R-square vs. Disease Severity. Patients #1-20 COPD and/or CHF disease severity compared to the training Generalized R-square resulting from the random K-fold (3tanh and 3 Gaussian).

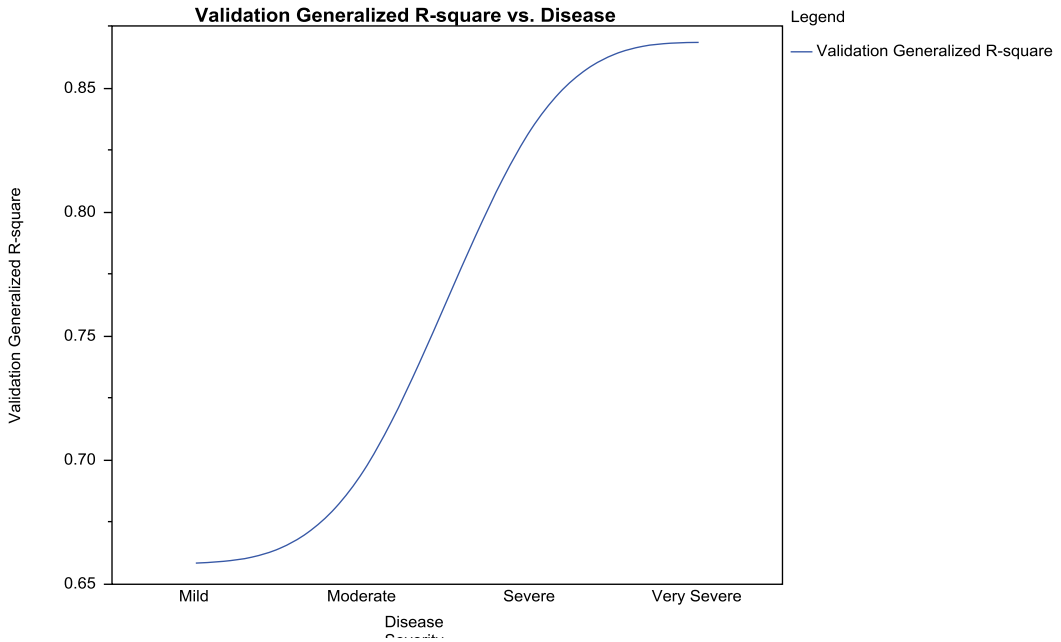


Figure 64. All Patients' Validation R-square vs. Disease Severity. Patients #1-20 COPD and/or CHF disease severity compared to the validation Generalized R-square resulting from the random K-fold (3tanh and 3 Gaussian).

It is now possible to re-test hypothesis B of the second specific aim where H_1 states that FEV1 measurements can be predicted demonstrating ≥ .80 Generalized R-square. JMP software considers Generalized R-square to be equal to R^2 in continuous normal responses, which is how FEV1's data property is identified. R^2 is equal to 1 - (*Sum of Squares* of the error divided by the *Sum of Squares* of the total). A very weak model is reflected in a Generalized R-square of zero, whereas a perfect fit of the model is a Generalized R-square of one. It describes the proportion of variability in a data set that is provided by the random K-fold model. In the training set seen in Table 19, the ranges of Generalized R-square are from .359 to 1, with an average of .782. This does not meet H_1, and, therefore, H_1 is rejected for the training set and H_0 is accepted. In the validation

set seen in Table 19, the ranges of Generalized R-square are from .156 to 1, with an average of .804. This does meet H_1, and, therefore, H_1 is accepted for the validation set and H_0 is rejected.

A summary of each patient's actual and predicted FEV1 values using the random K-fold model as described above are presented in Figures 65 to 84, with selected statistics included. Note the value shift from actual to predicted on the *y* axes, as well as each data point, resulting from the transformation (3tanh and 3Gaussian). A smoother is included, which is a cubic spline (λ = 0.05) and standardized x values. The cubic spline is a third degree polynomial spliced together so the resulting curve is smooth.

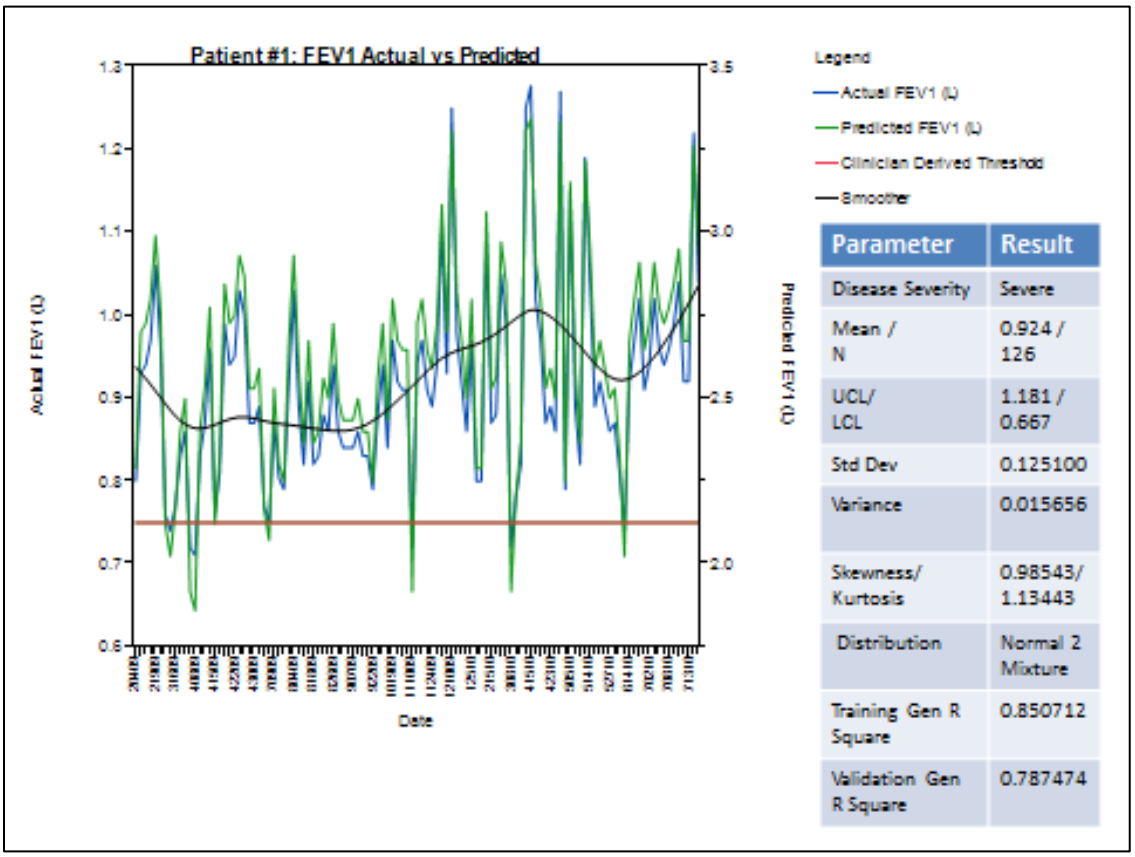


Figure 65. Patient #1 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

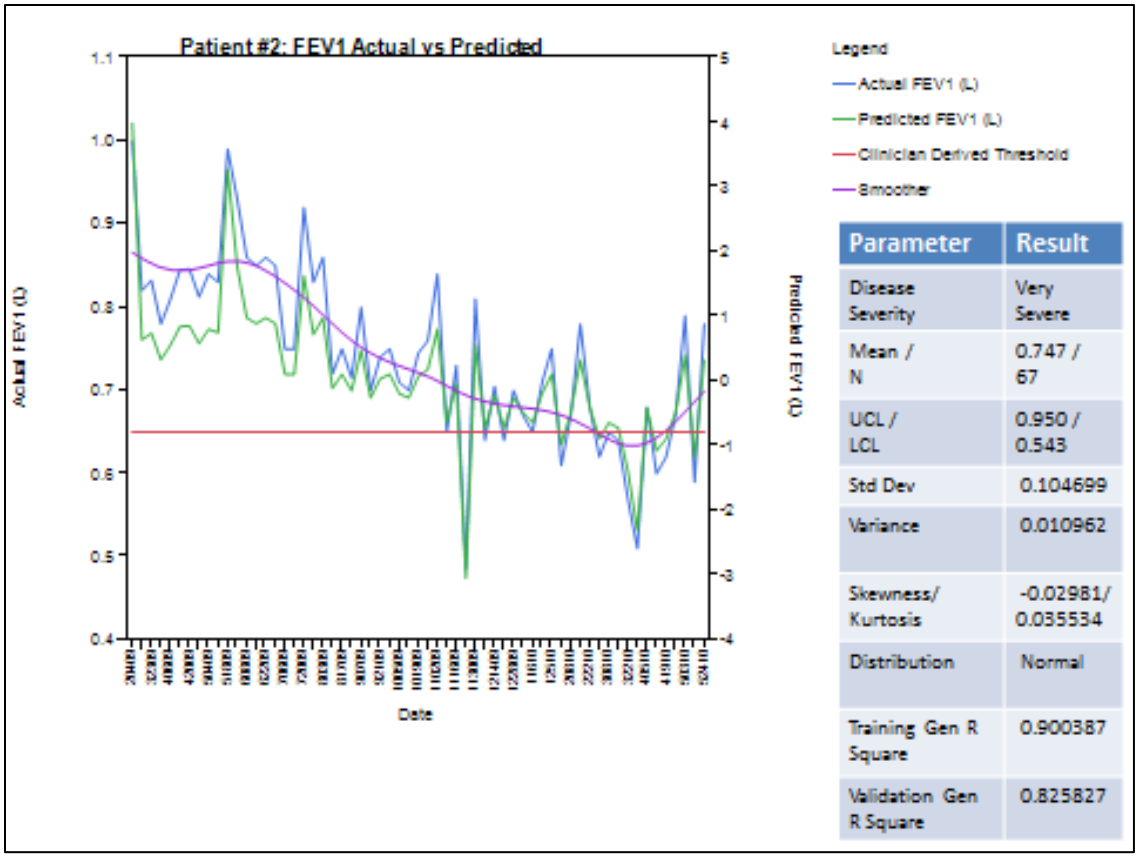


Figure 66. Patient #2 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

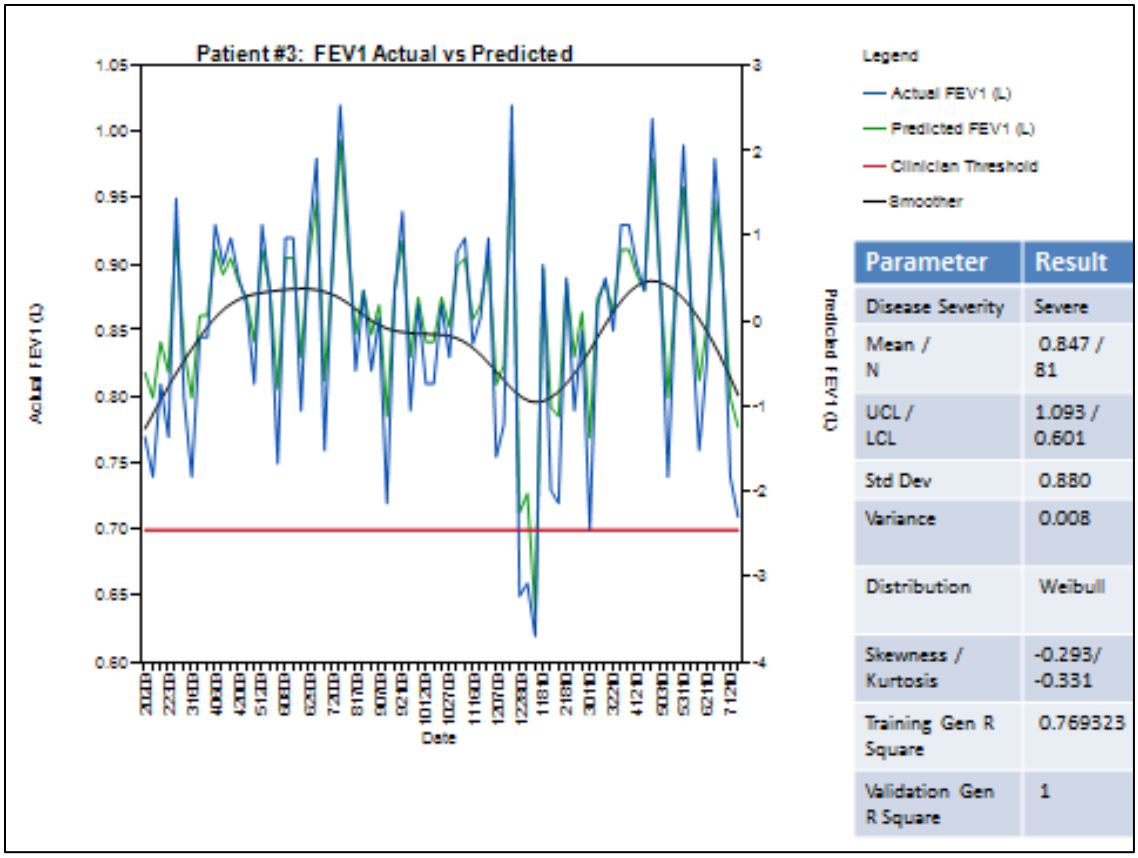


Figure 67. Patient #3 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

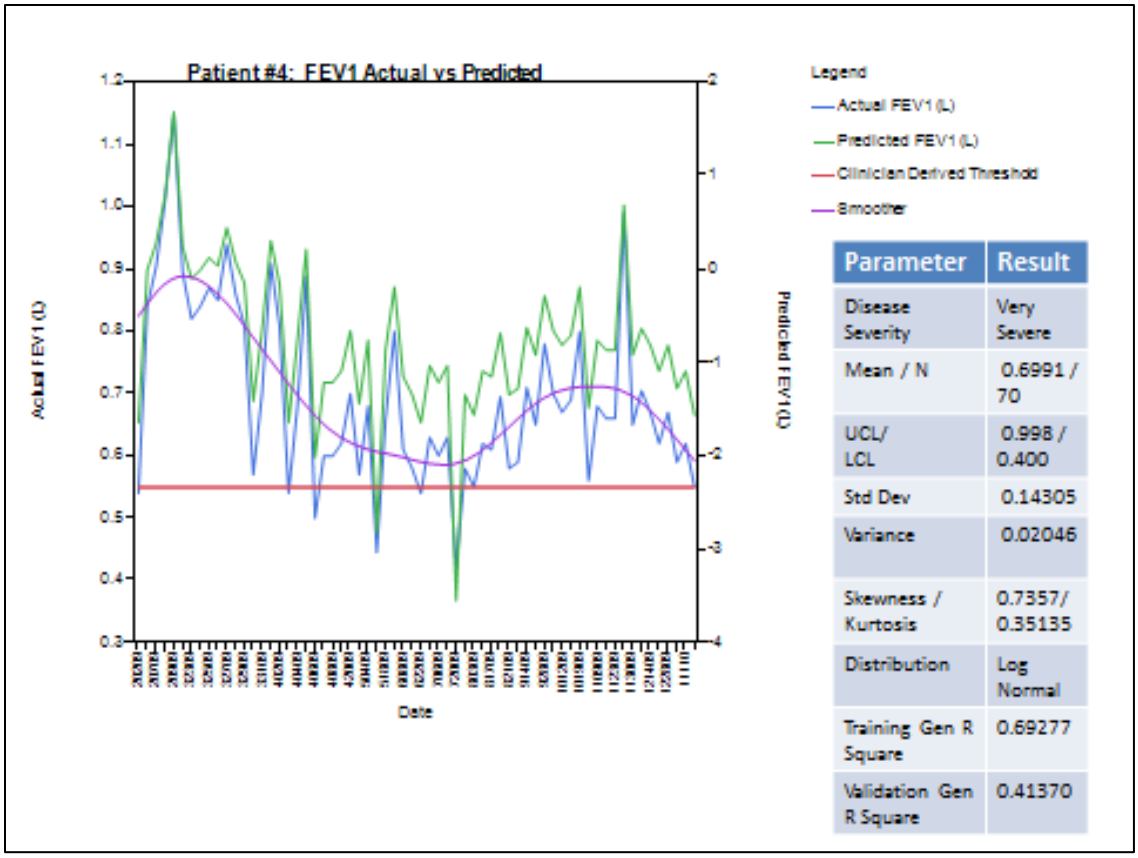


Figure 68. Patient #4 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

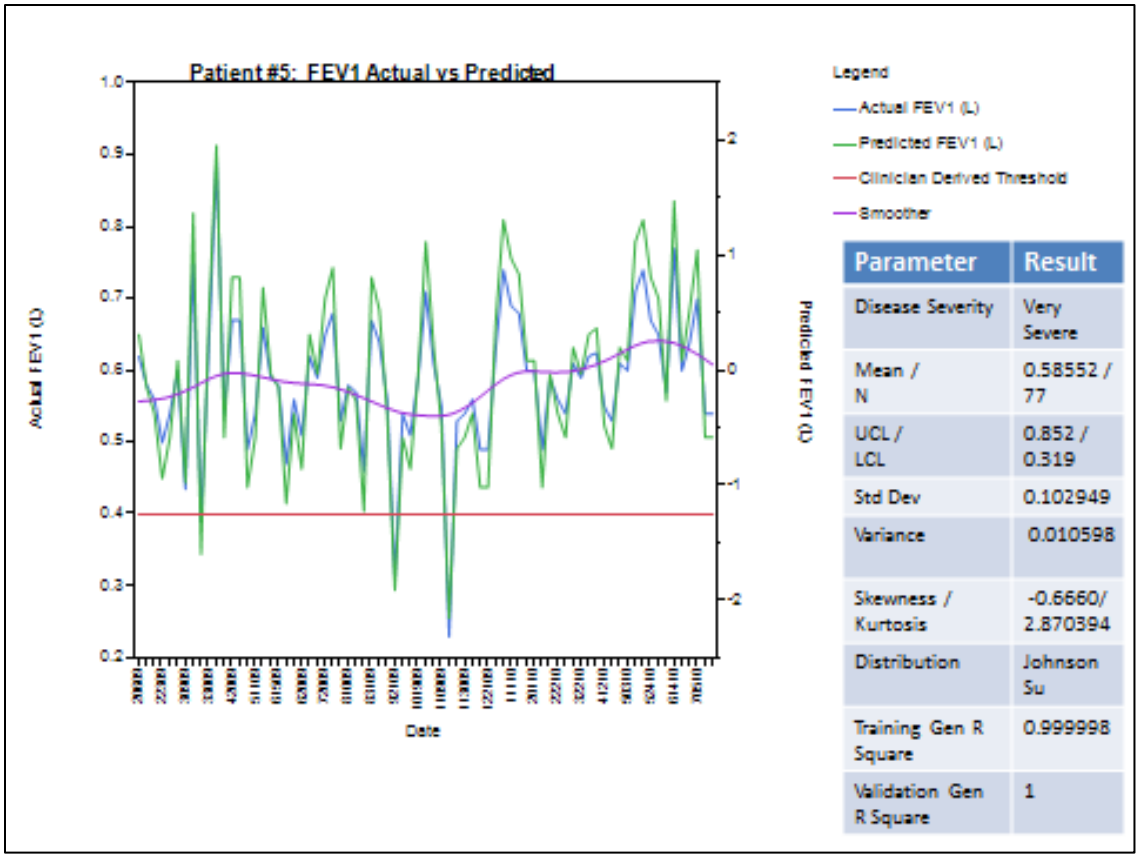


Figure 69. Patient #5 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

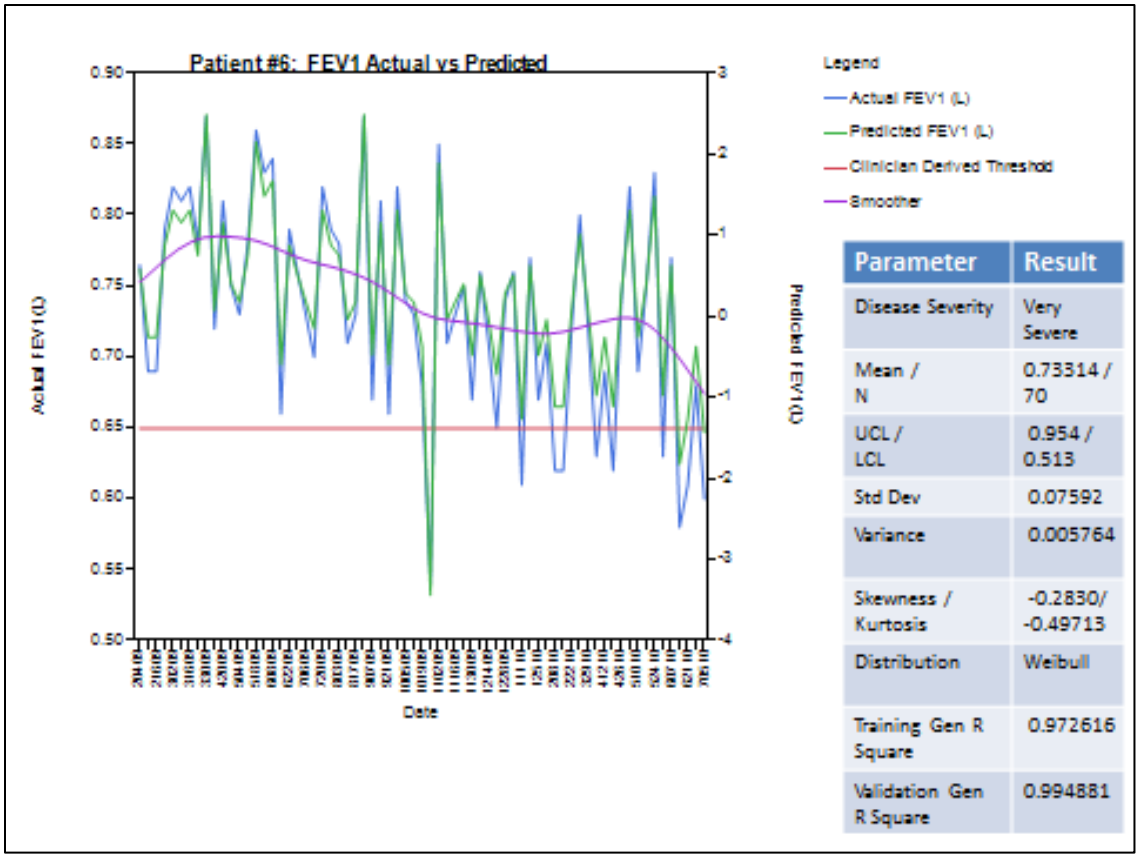


Figure 70. Patient #6 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

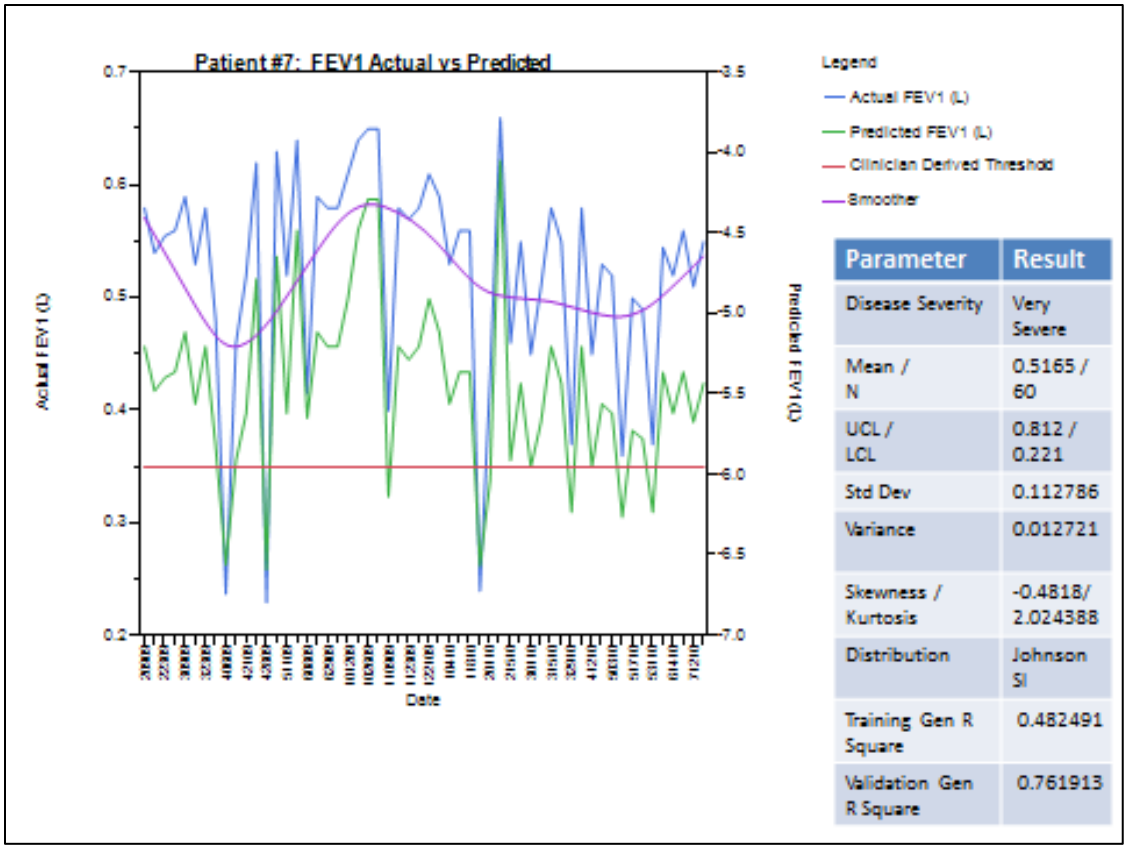


Figure 71. Patient #7 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

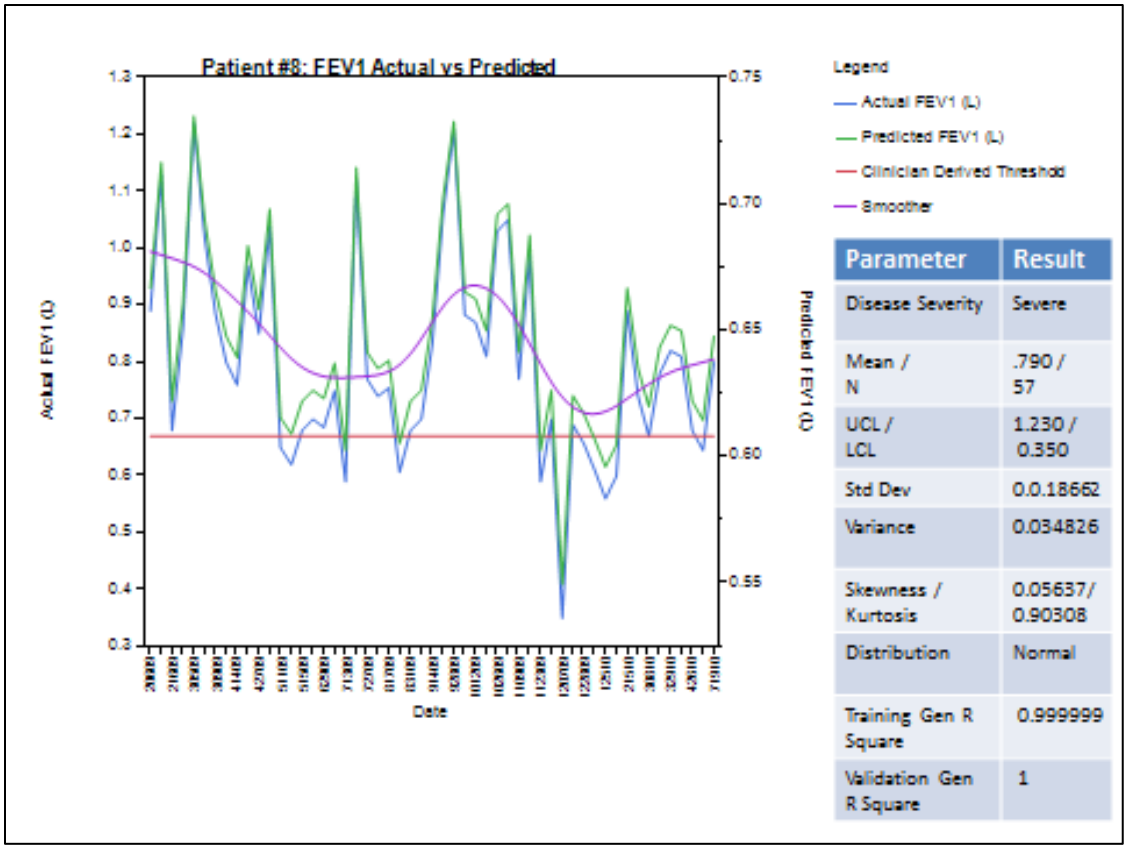


Figure 72. Patient #8 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

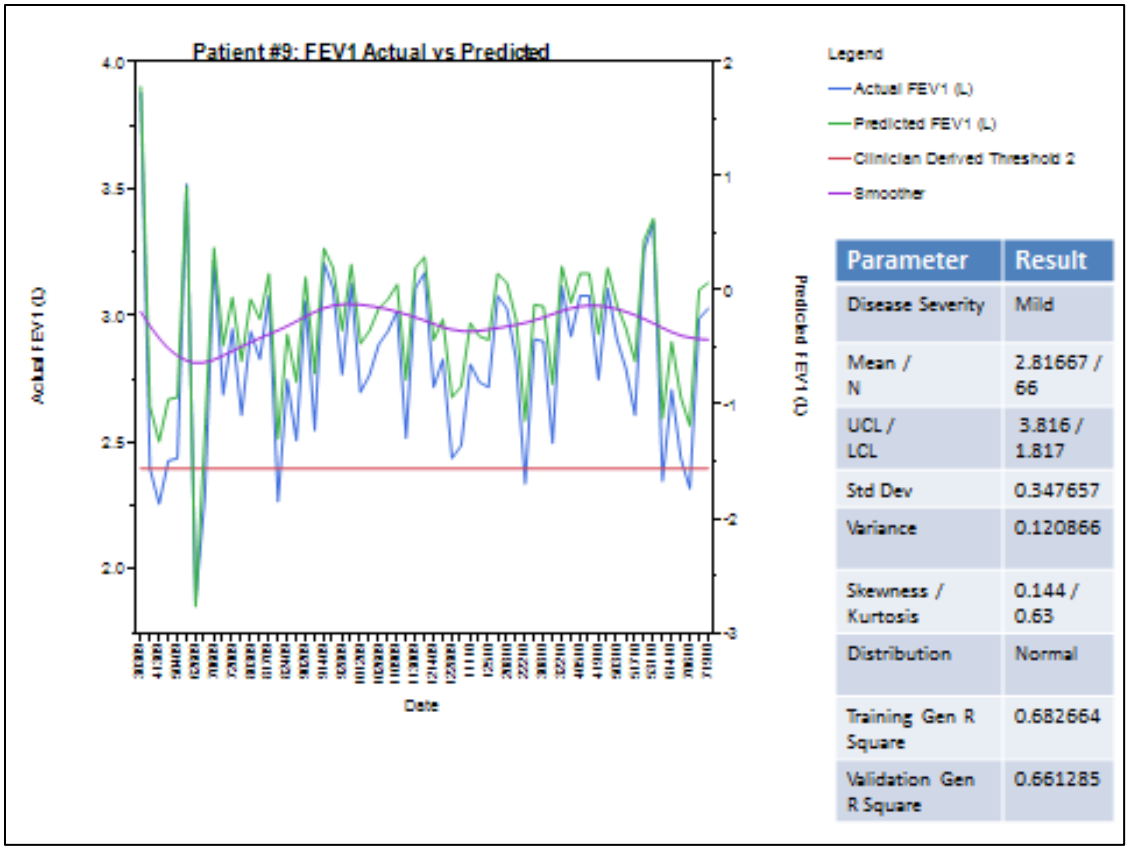


Figure 73. Patient #9 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

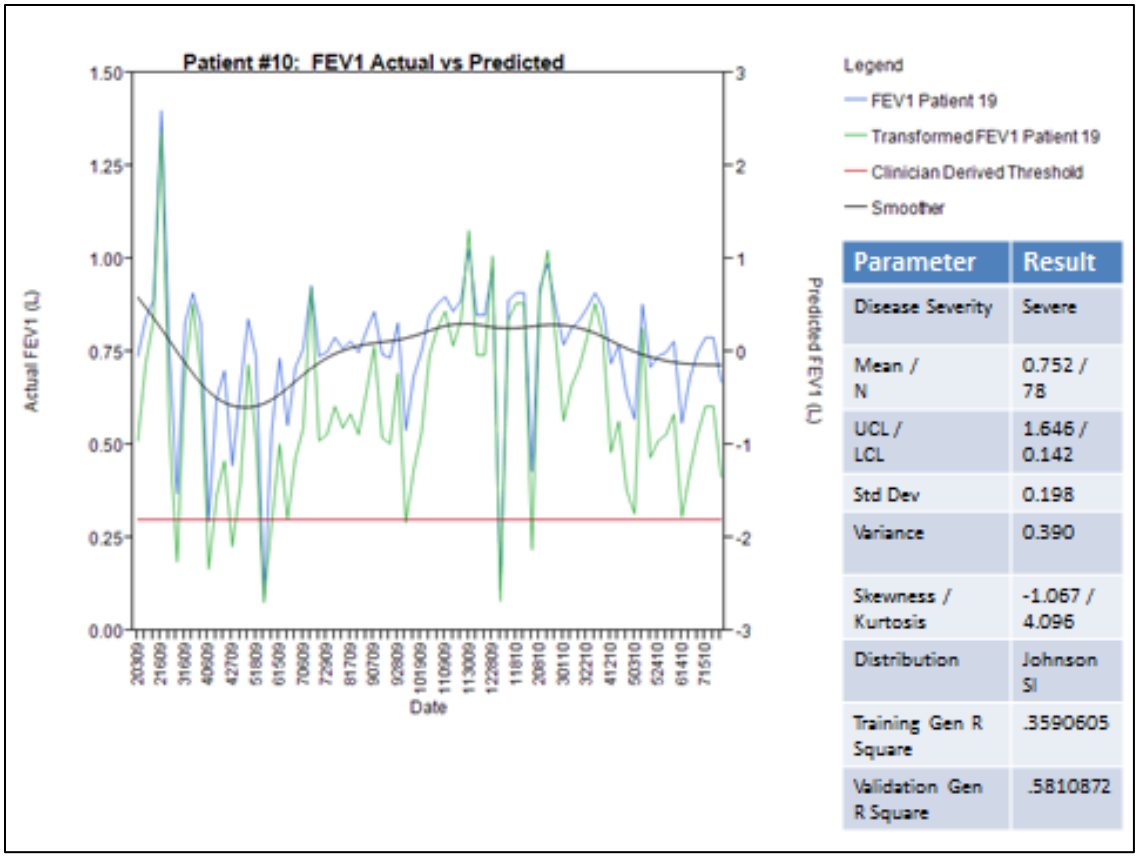


Figure 74. Patient #10 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

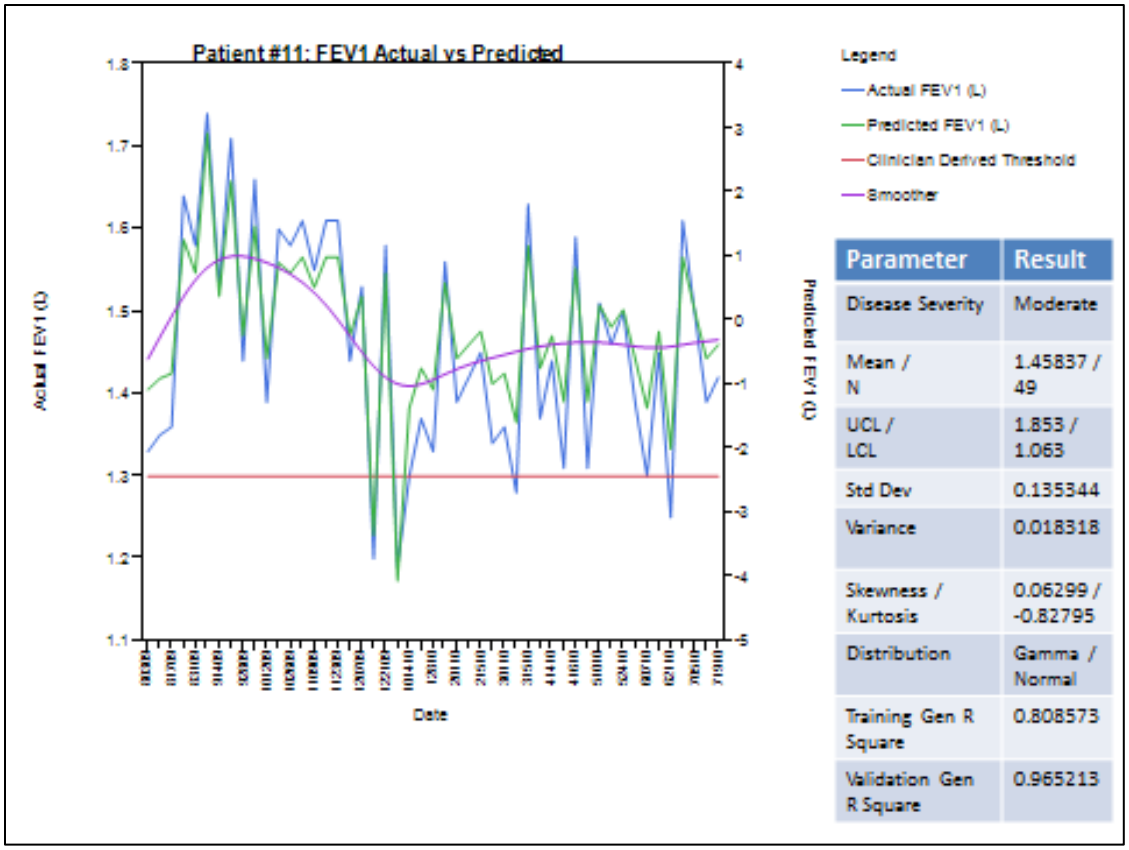


Figure 75. Patient #11 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

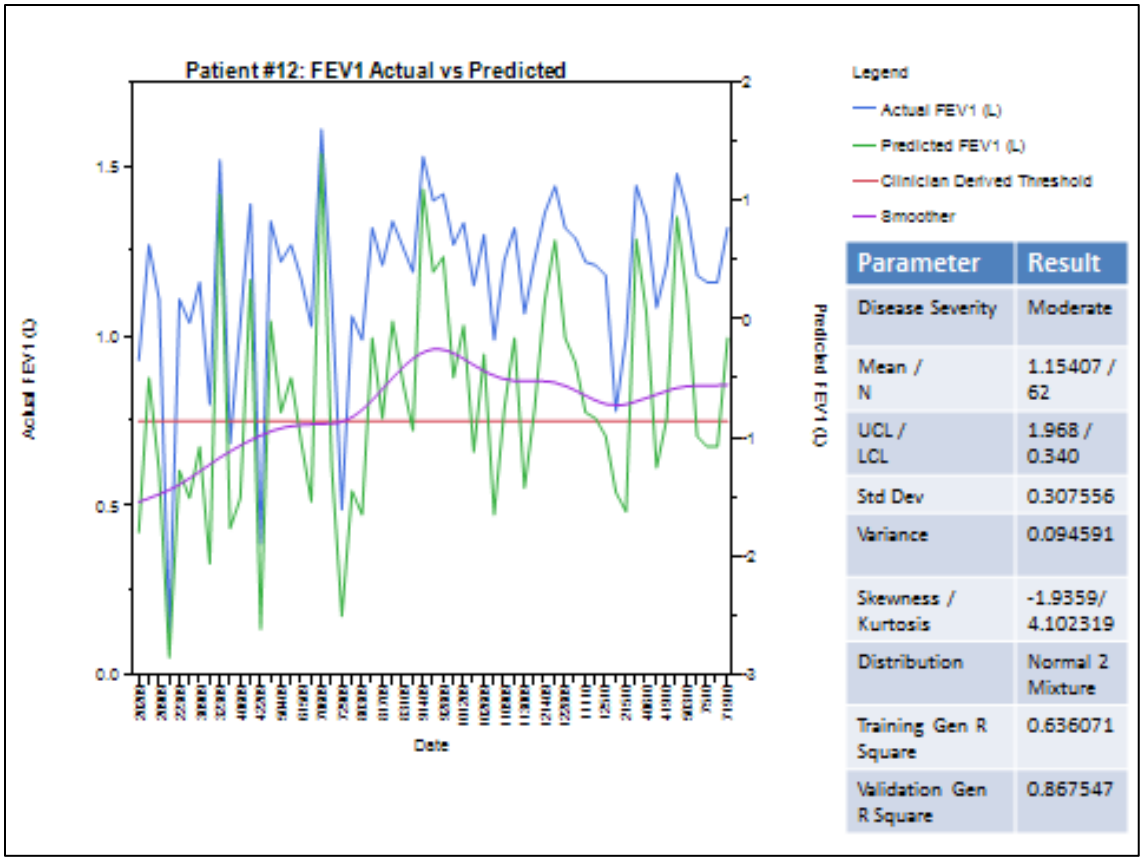
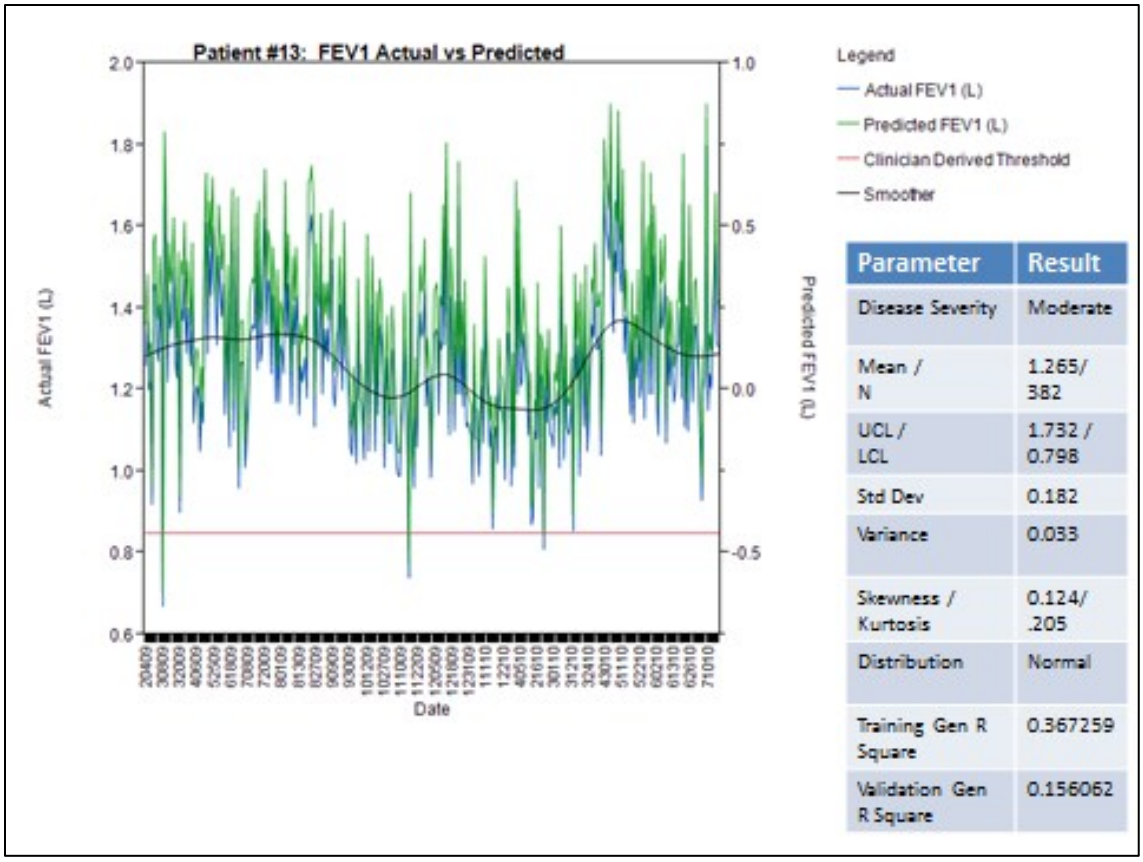


Figure 76. Patient #12 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.



Parameter	Result
Disease Severity	Moderate
Mean / N	1.265/ 382
UCL / LCL	1.732 / 0.798
Std Dev	0.182
Variance	0.033
Skewness / Kurtosis	0.124/ .205
Distribution	Normal
Training Gen R Square	0.367259
Validation Gen R Square	0.156062

Figure 77. Patient #13 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

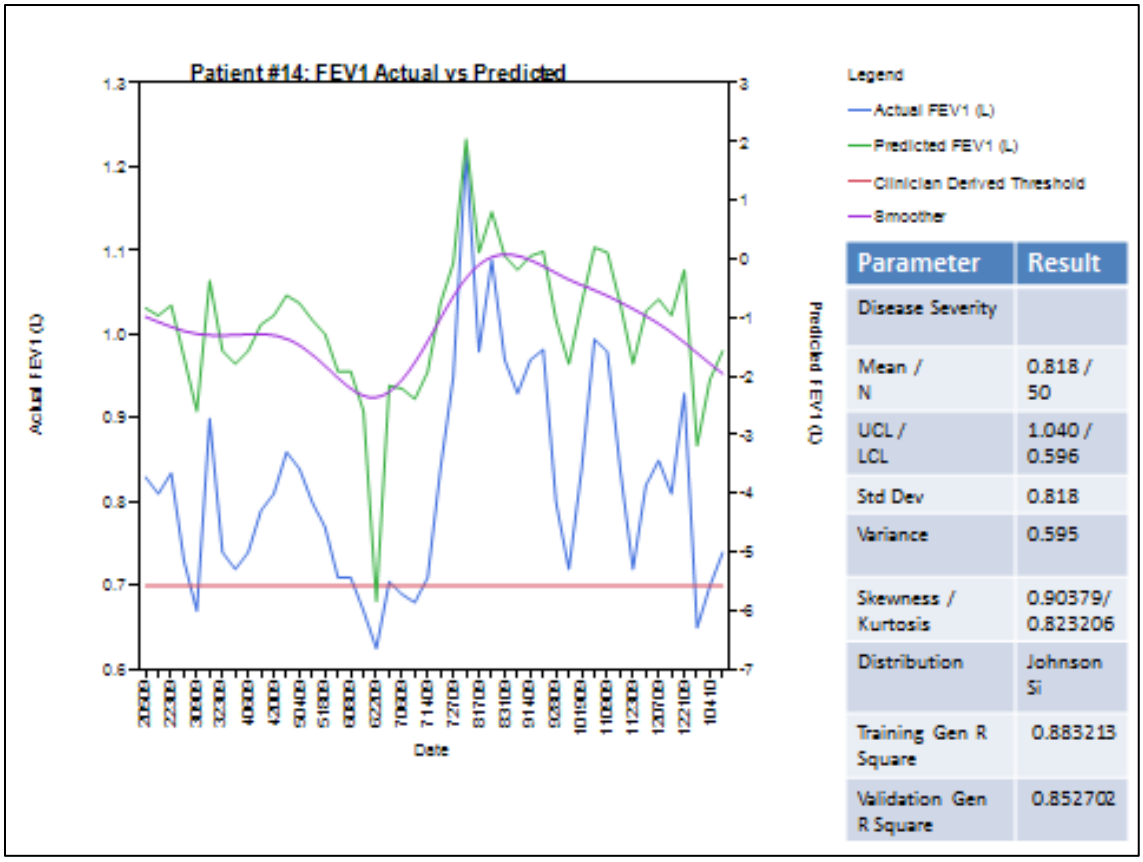


Figure 78. Patient #14 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

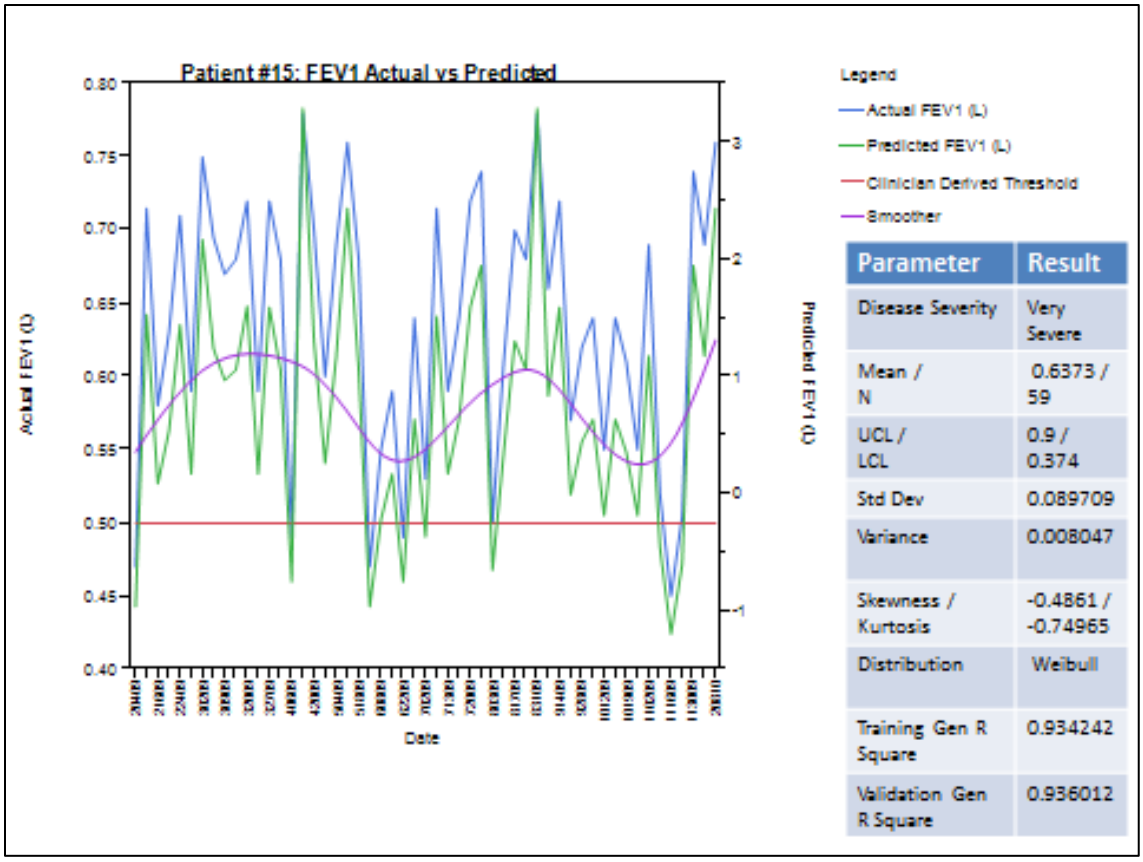


Figure 79. Patient #15 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

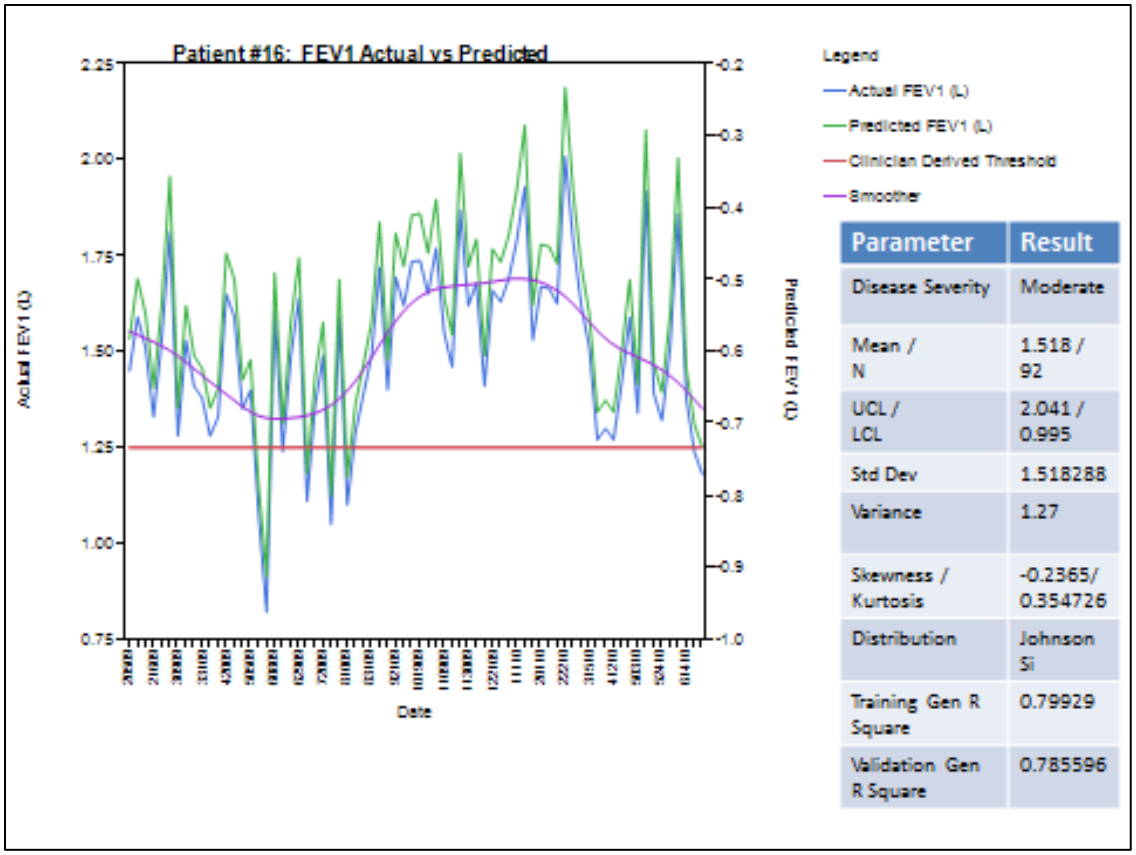


Figure 80. Patient #16 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

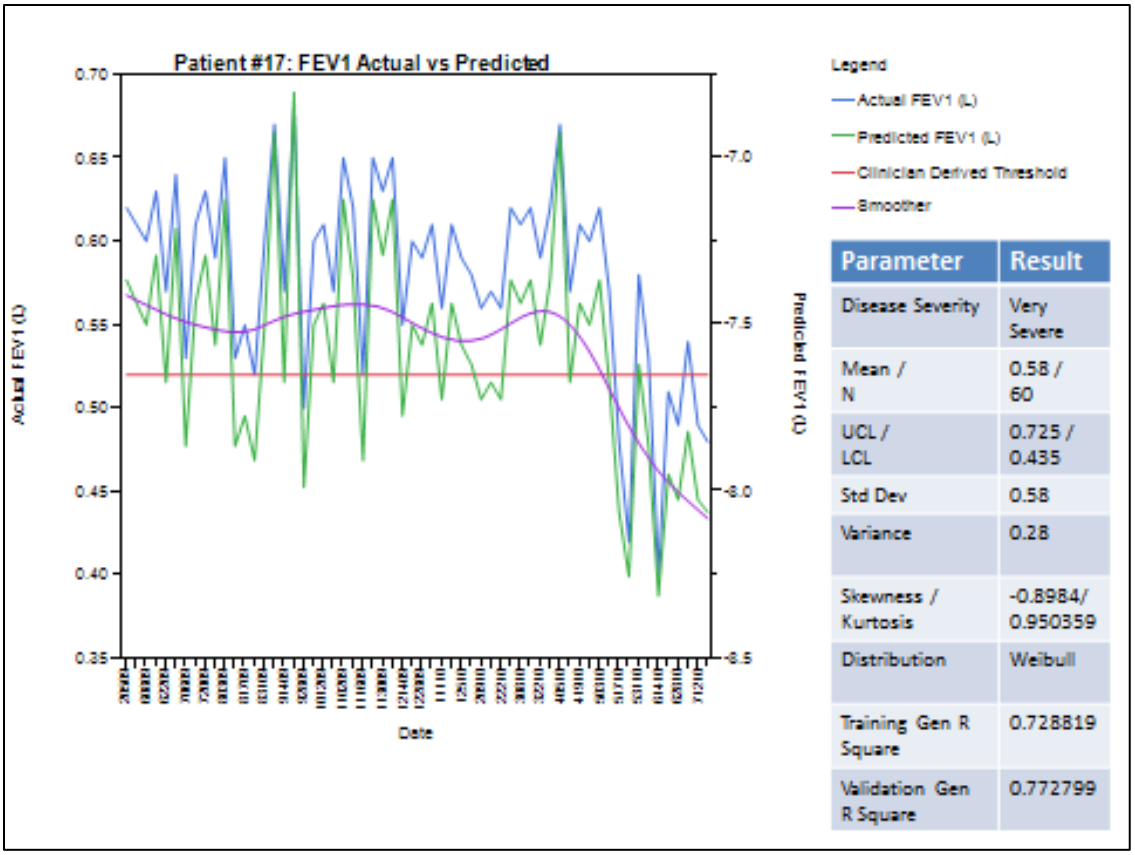


Figure 81. Patient #17 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

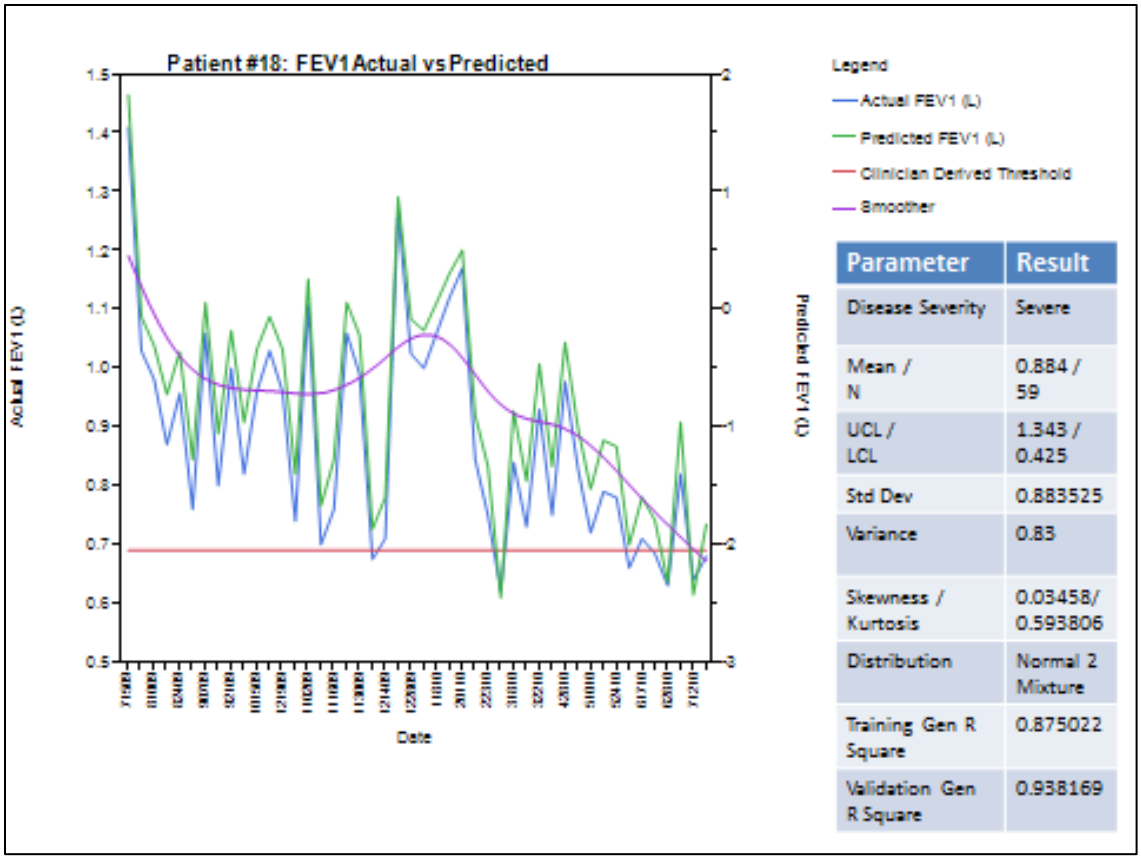


Figure 82. Patient #18 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

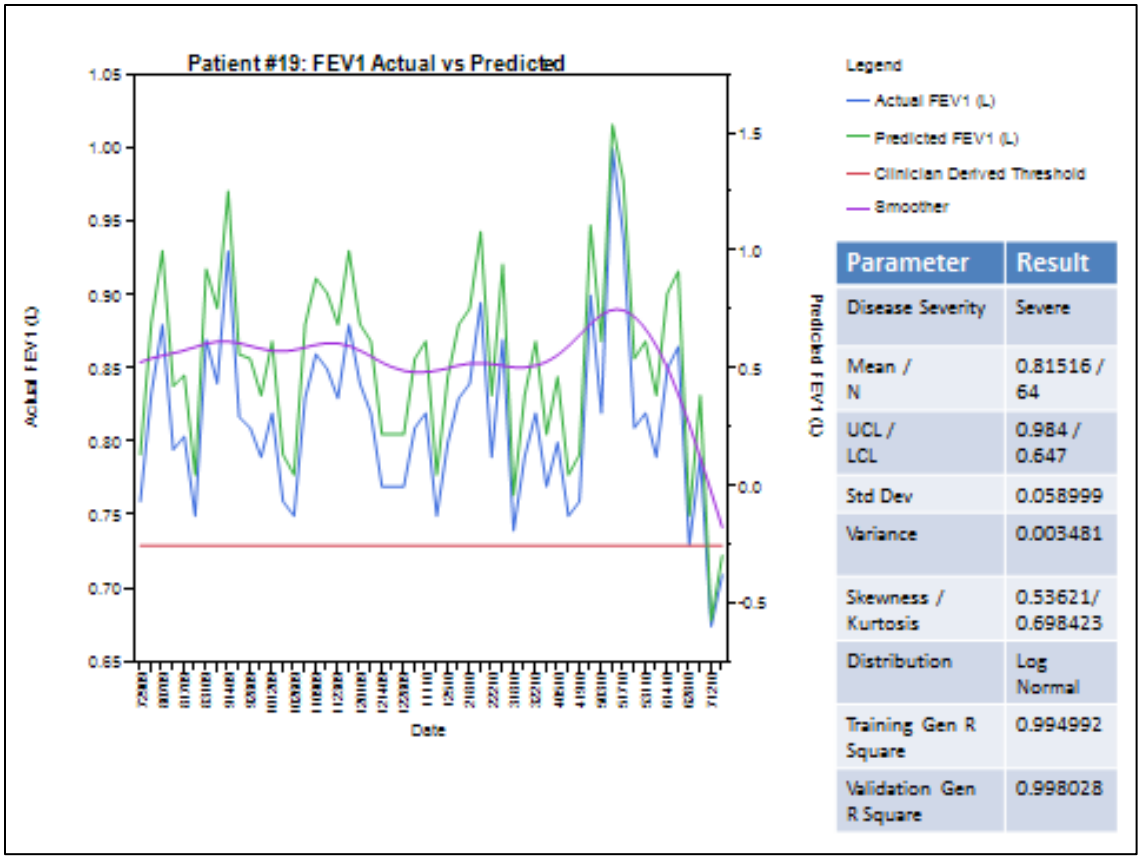


Figure 83. Patient #19 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

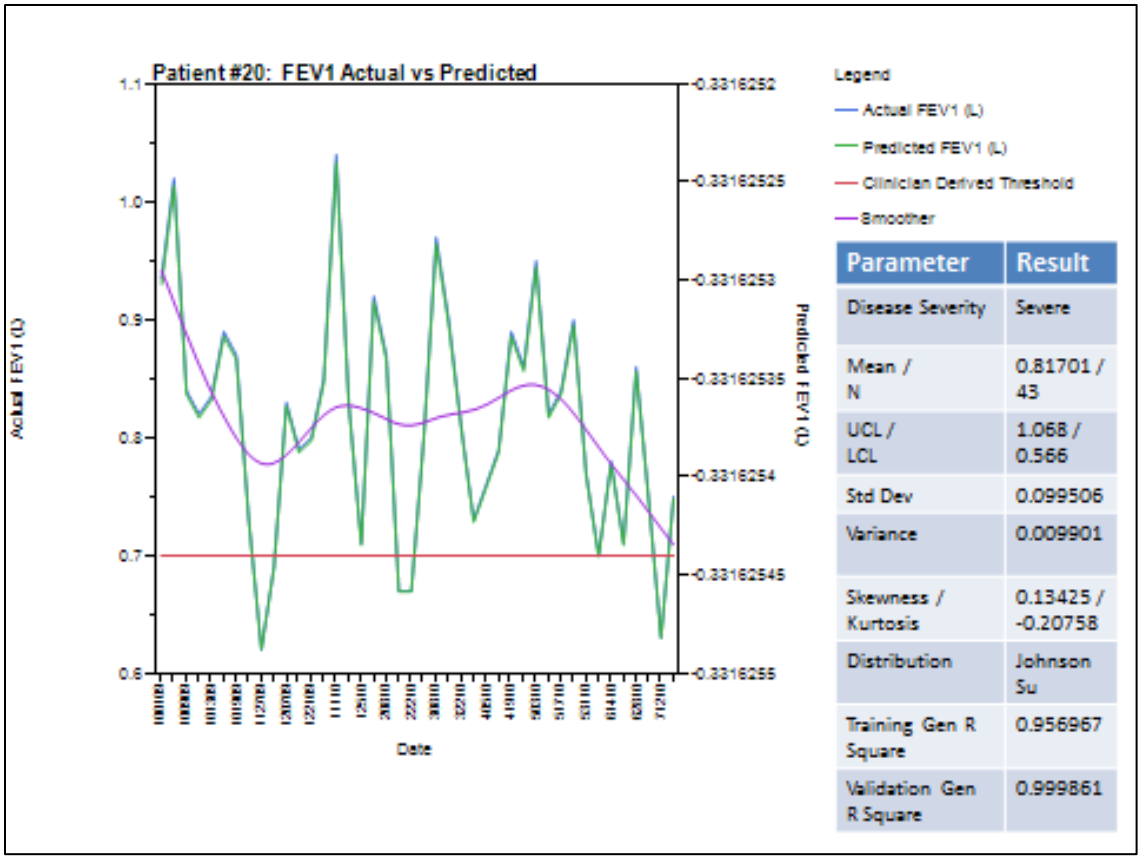


Figure 84. Patient #20 Actual vs. Predicted FEV1. Smoother and the clinician derived threshold noted.

In order to test hypothesis C of the second specific aim which states that the pre- and post-predicted data point will be statistically different (p<.005) in SPC zone designation, the corresponding +/- 3σ zone point locality was compared before and after the random K-fold was applied. The value of the zones was determined for each population (pre- and post-prediction), and then each data point was compared between the two. Zones are designated per SPC convention: A (+/- 3σ), B (+/- 2σ), C (+/- 1σ), and this study added O (Out-of-Control per Western Electric Rules).

Patient #1's FEV1 SPC zone data were analyzed (Figure 85) for pre- and post-prediction agreement. While a significant amount of data points were the same (as seen in the linear axis), there were a large number of data points that were pre-prediction zone A, B, C, and O that were not the same as their corresponding post-prediction zone. Of Patient #1's 126 total FEV1 data points, 25 (19.8%) were incorrectly predicted for the zone, resulting in an R-square of 0.573 and a Kappa Coefficient (degree of agreement) of 0.703 (p<0.005). To understand why, the individual pre- and post-prediction points were designated 0 (points agree pre- and post-prediction) or 1 (points did not agree pre- and post-prediction). The 0/1 designations were then plotted against the pre-prediction FEV1 values (Figure 86) and post-prediction FEV1 values (Figure 87), and contingency analysis performed (Table 22) for Patient #1. Points designated 1 in Table 22 that fell on or near the sigma values ("edge sigma") were unremarkable. The FEV1 zone data from Patient #1 coupled with low R-square and Kappa Coefficient do not indicate that pre- and post-prediction data are the same; therefore, for hypothesis C of the second specific aim H_0 is rejected and H_1 is accepted (p<0.005). The pre- and post-predicted data points are different.

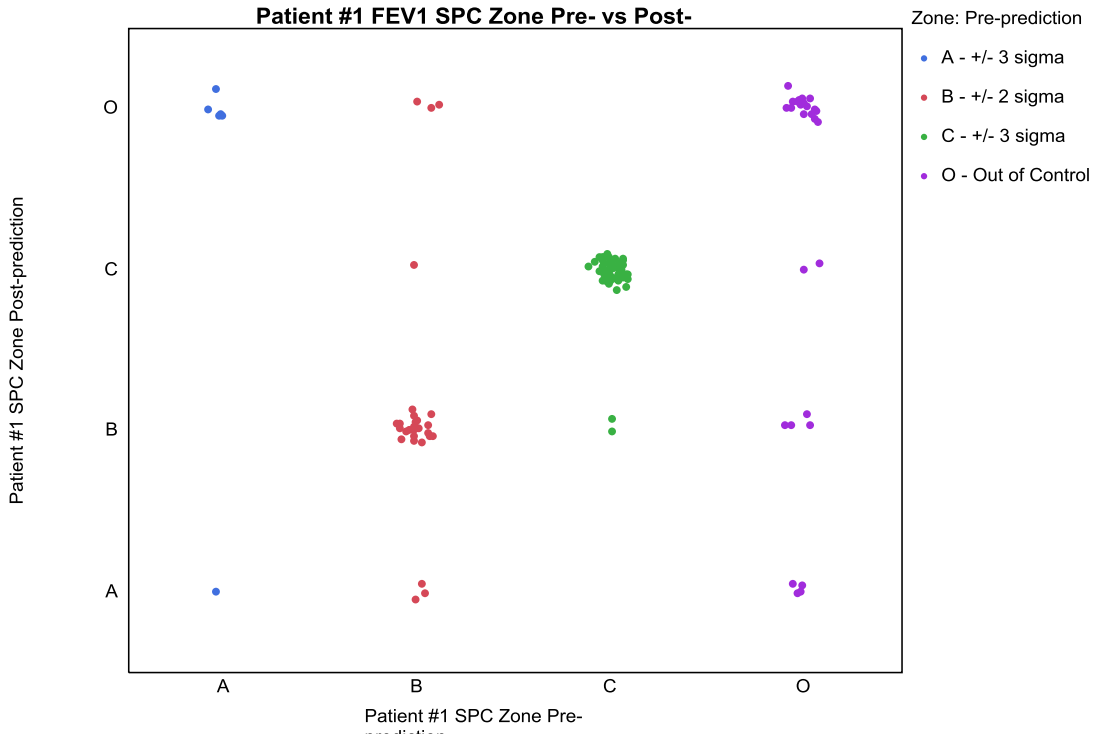


Figure 85. Patient #1 FEV1 SPC Zone Pre- vs. Post-prediction. Patient #1 FEV1 SPC zone, pre-predicted vs. post-predicted following administration of a random K-fold as described earlier. 19.8% are mis-matched, resulting in a R-square of 0.573 and a Kappa Coefficient of 0.703 ($p<0.005$). Chi-square is suspect due to small counts.

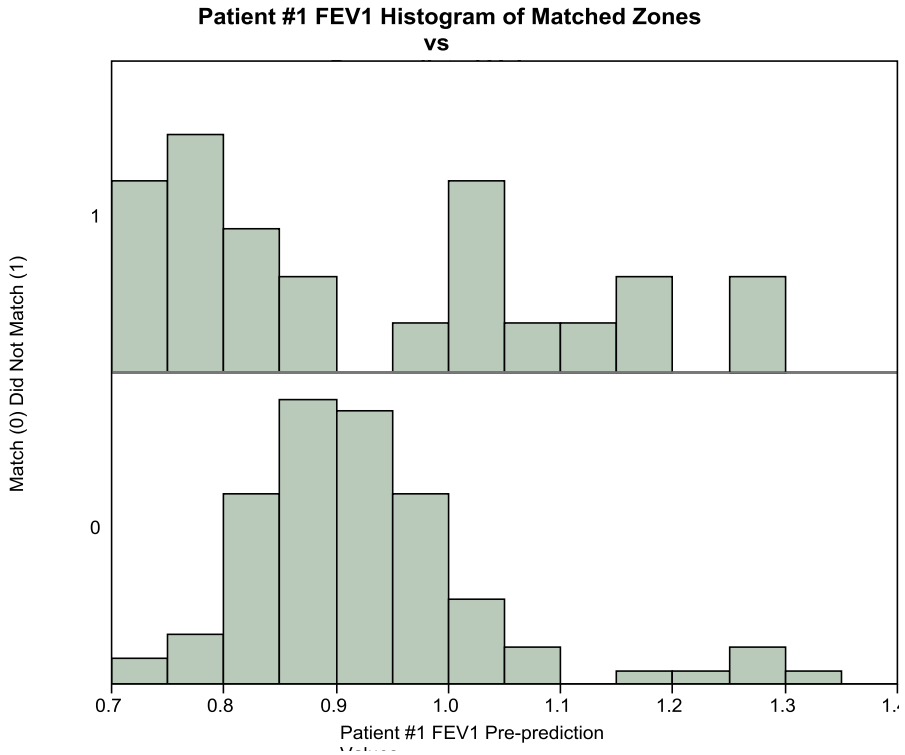


Figure 86. Patient #1 FEV1 Matched Zones vs. Pre-prediction. Patient #1 histogram of zone match (0 = match; 1 = did not match) vs. FEV1 pre-prediction values.

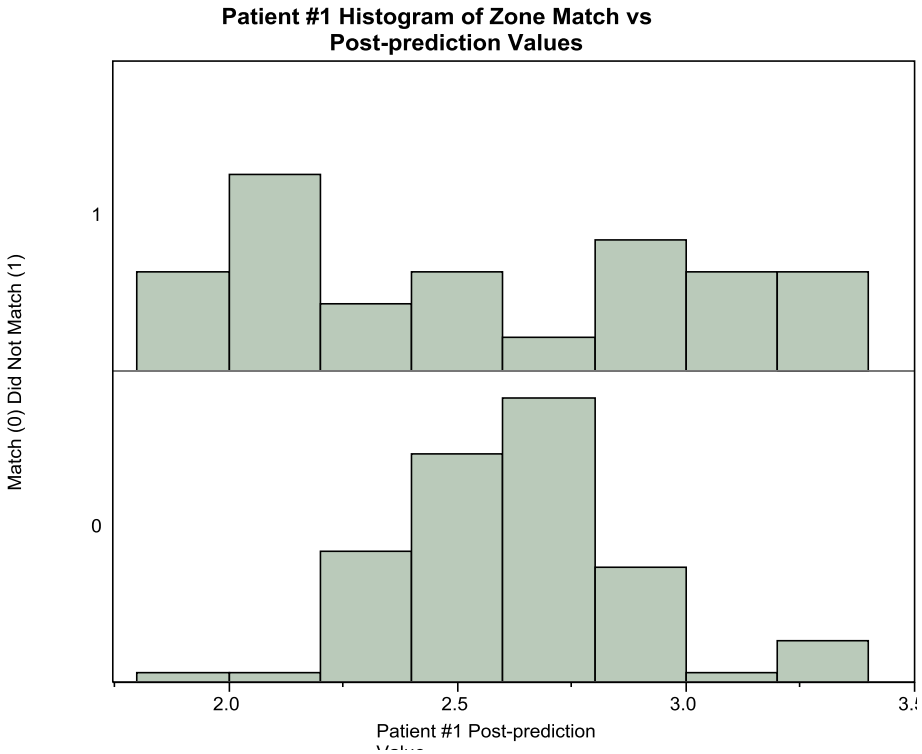


Figure 87. Patient #1 FEV1 Matched Zones vs. Post-prediction. Patient #1 FEV1 histogram of zone match (0 = match; 1 = did not match) vs. FEV1 post-prediction values.

Table 22. Patient #1 FEV1 Contingency Analysis. Pre- and post-prediction zones of A, B, C, and O.

Zone-Pre-prediction	Cell	A	B	C	O
A	Count	1	0	0	5
A	Total %	0.793651	0	0	3.968254
A	Col %	12.5	0	0	20
A	Row %	16.66667	0	0	83.33333
A	Expected	0.380952	1.333333	3.095238	1.190476
A	Deviation	0.619048	-1.33333	-3.09524	3.809524
A	Cell Chi^2	1.005952	1.333333	3.095238	12.19048
B	Count	3	22	1	3
B	Total %	2.380952	17.46032	0.793651	2.380952
B	Col %	37.5	78.57143	1.538462	12
B	Row %	10.34483	75.86207	3.448276	10.34483
B	Expected	1.84127	6.444444	14.96032	5.753968
B	Deviation	1.15873	15.55556	-13.9603	-2.75397
B	Cell Chi^2	0.729201	37.54789	13.02716	1.318106
C	Count	0	2	62	0
C	Total %	0	1.587302	49.20635	0
C	Col %	0	7.142857	95.38462	0
C	Row %	0	3.125	96.875	0
C	Expected	4.063492	14.22222	33.01587	12.69841
C	Deviation	-4.06349	-12.2222	28.98413	-12.6984
C	Cell Chi^2	4.063492	10.50347	25.44472	12.69841
O	Count	4	4	2	17
O	Total %	3.174603	3.174603	1.587302	13.49206
O	Col %	50	14.28571	3.076923	68
O	Row %	14.81481	14.81481	7.407407	62.96296
O	Expected	1.714286	6	13.92857	5.357143
O	Deviation	2.285714	-2	-11.9286	11.64286
O	Cell Chi^2	3.047619	0.666667	10.21575	25.30381

Validating the conclusion from Patient #1 that H_1 of hypothesis C is accepted, another patient's data were examined. Patient #3 FEV1 SPC zone data were also analyzed

(Figure 88) for pre- and post-prediction. Similar to Patient #1, a significant amount of data points were the same (as seen on the linear axis), but a large number of data points exist that were pre-prediction zone A, B, C, and O that were not the same as their corresponding post-prediction zone. Of Patient #3's 81 total FEV1 data points, 36 (44.4%) were incorrectly predicted for the zone, resulting in an R-square of 0.249 and a Kappa Coefficient (degree of agreement) of 0.147 (p<0.005). To understand why, the individual pre- and post-prediction points were designated 0 (points agree pre- and post-prediction) or 1 (points did not agree pre- and post-prediction). The 0/1 designations were then plotted against the pre-prediction FEV1 values (Figure 89) and post-prediction FEV1 values (Figure 90) for Patient #3. Points designated 1 falling on or near the sigma values ("edge sigma") were, again, unremarkable. Patient #3's FEV1 zone data coupled with low R-square and Kappa Coefficient do not significantly (p<0.005) indicate that pre- and post-prediction data are the same; therefore, H_0 of hypothesis 3 of the second specific aim continues to be rejected and H_1 continues to be accepted.

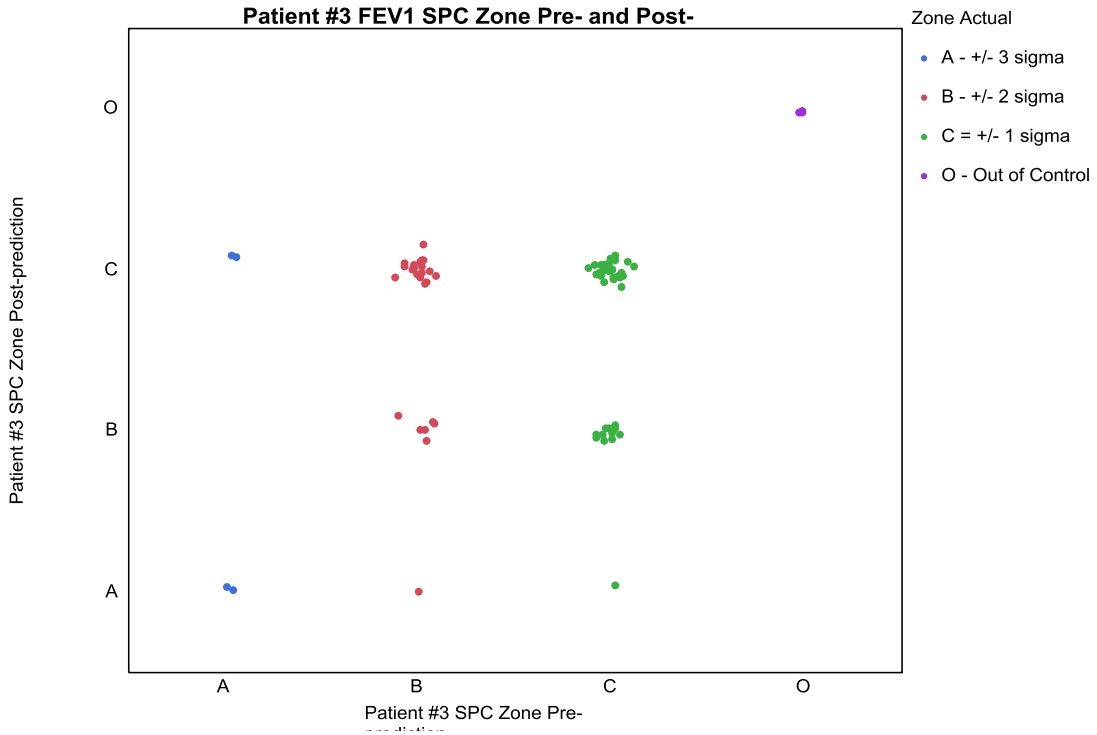


Figure 88. Patient #3 FEV1 SPC Zone Pre- vs. Post-prediction. Patient #3 FEV1 SPC zone, pre-prediction vs. post-prediction following administration of a random K-fold as described earlier. 44.4% are mis-matched, resulting in an R-square of 0.249 and a Kappa Coefficient of 0.147 ($p<0.005$). Chi-square is suspect due to small counts.

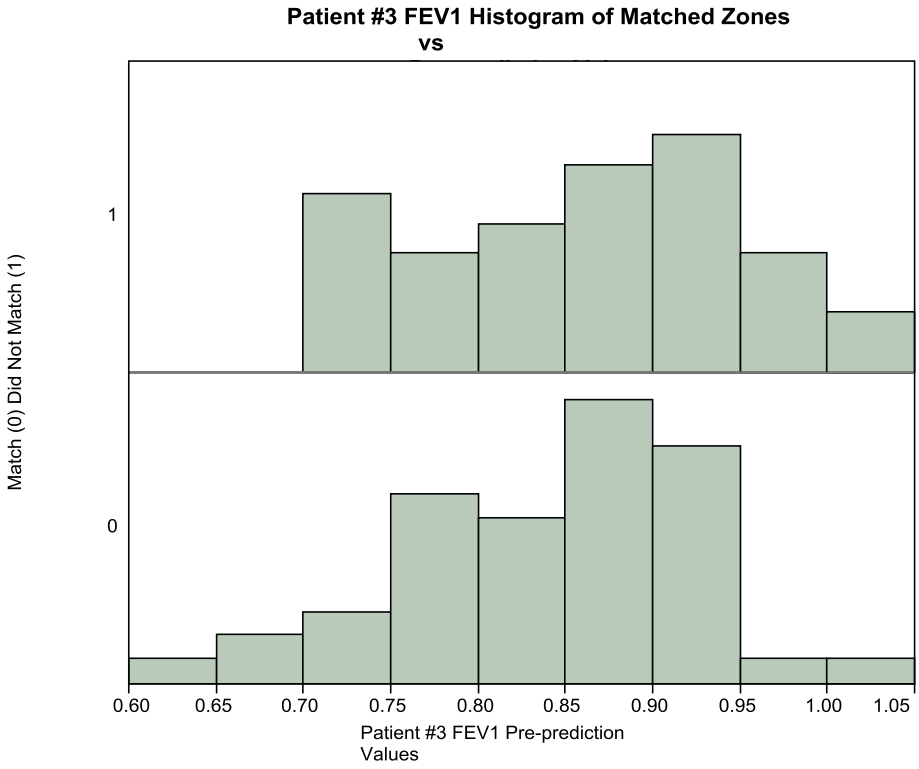


Figure 89. Patient #3 FEV1 Matched Zones vs. Pre-prediction. Patient #3 histogram of zone match (0 = match; 1 = did not match) vs. FEV1 pre-prediction values.

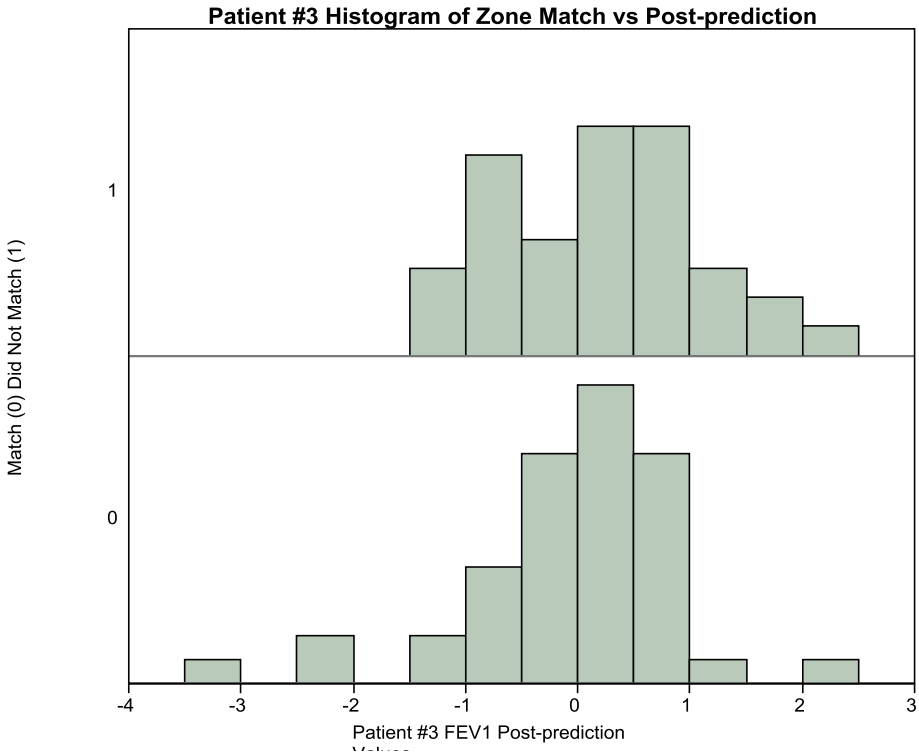


Figure 90. Patient #3 FEV1 Matched Zones vs. Post-prediction. Patient #3 FEV1 histogram of zone match (0 = match; 1 = did not match) vs. FEV1 post-prediction values.

To address the third specific aim and its hypothesis D, forecasting a future value in time in an FEV1 series would provide the clinician and the patient insight into COPD and/or CHF, providing an opportunity to manage disease outcome prior to its actual occurrence. JMP 9.0.3 software was limited in its modeling platform for neural nets (which provided the prior random K-fold prediction for classification), and, therefore, the modeling platform was changed to autocorrelation. Dates in this platform are required to have equidistance between them, and, therefore, each FEV1 data was assigned 1-n and labeled “equivalent date”. Utilizing 25 autocorrelation lags, 10 forecasts, a confidence interval of 0.95, constraints 0-1, and (where applicable) periods/season of 12, various smoothing methods were attempted for forecasting the next value in a series for Patient

#3. Table 23 summarizes the smoothing methods and resulting negative R-squares which indicate a poor fit for the models.

Table 23. Patient #3 FEV1 Smoothing Methods. Smoothing methods used to forecast a future value in the time series. The use of seasonal models is not applicable for the FEV1 data but is included for completeness.

Smoothing Method	**R-square**
Winter's Method	-0.498
Seasonal Exponential Smoothing	-0.498
Damped Trend Linear Exponential Smoothing	-0.052
Linear (Holt) Exponential Smoothing	-0.198
Double Exponential Smoothing	-0.207
Simple Exponential Smoothing	-0.050

Following the unsuccessful smoothing methods for forecasting a future value in the time series of Patient #3's FEV1, an autoregressive integrated moving average (ARIMA) was investigated. The autoregressive order (p), non-seasonal differences/difference order (d), and moving average order (q) were varied using an intercept and constrained fit. 24 combinations with 245 iterations each were completed, and resulted in a mixed ARIMA (p, d, q) of (45, 0, 49) with a .595 R-square. Figure 91 is the resulting graph comparing the original n=81 pre-ARIMA FEV1 data points to the post-ARIMA (45, 0, 49) n=91 data values.

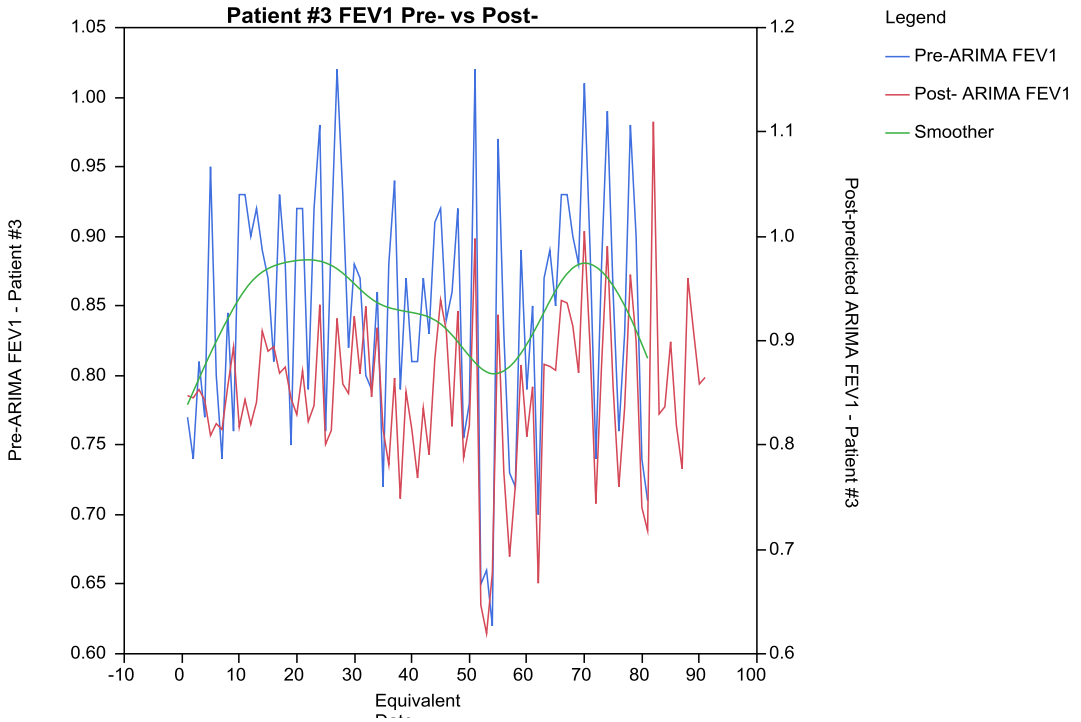


Figure 91. Patient #3 FEV1 ARIMA. Pre- vs. post-ARIMA (45, 0, 49) with smoother.

Examining the SPC charts for pre- and post-ARIMA (45, 0, 49) results in Figures 92 and 93. The three original (pre-ARIMA) points in Figure 92 continue to be rejected in the post-ARIMA in Figure 93. However, additional points are also rejected previous to data point 81; mean, UCL, and LCL have tightened in the post-ARIMA state. This may provide clinical confidence that future points beyond data point 81 are realistic and should be addressed with the patient – before the patient has actually taken their FEV1 value from their peak flow device, the next day or up to ten days in advance (Figure 91).

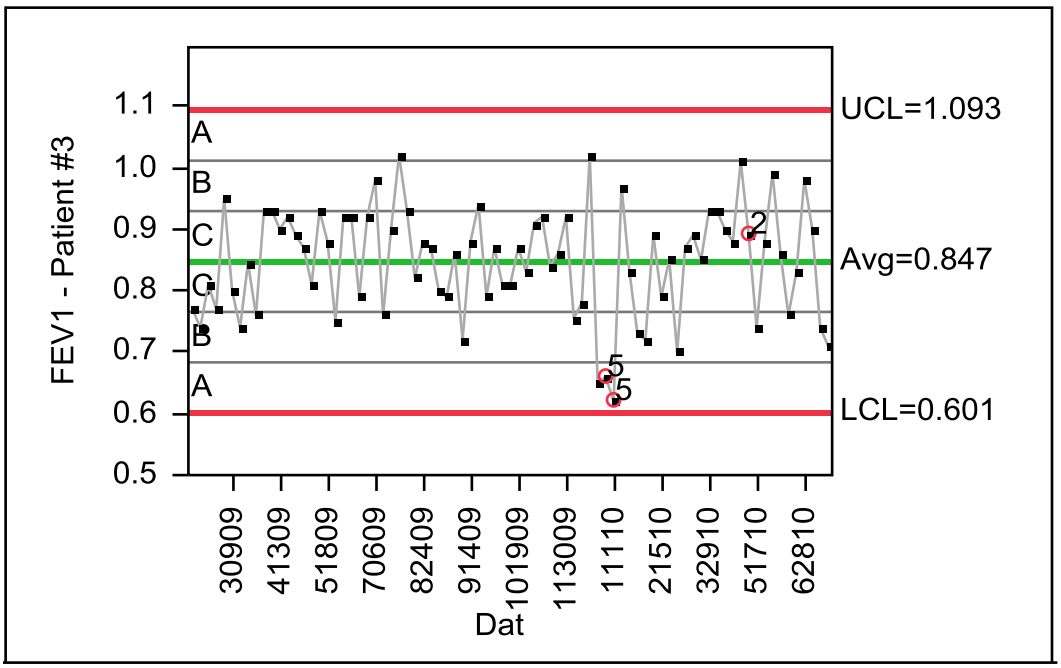


Figure 92. Patient #3 FEV1 Shewart-Western Electric Analysis Pre-ARIMA. Patient #3 FEV1 Shewart-Western Electric analysis, using the original, pre-ARIMA FEV1 values. Note the three out-of-control points marked 5, 5, and 2 corresponding to the Western Electric Rules in Table 7.

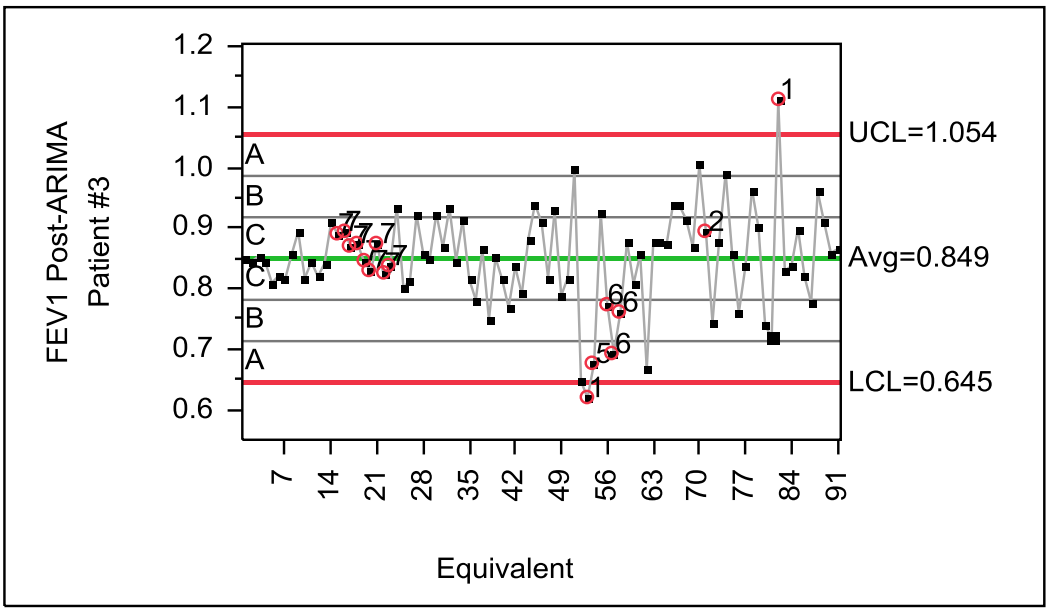


Figure 93. Patient #3 FEV1 Shewart-Western Electric Analysis Post-ARIMA. This analysis is the post-ARIMA (45, 0, 49) for Patient #3. Note the out-of-control points before Equivalent Date data point 81 compared to Figure 92. From data point 81 forward, only one point was out-of-control, data point 82, failing Western Electric Rule 1, per Table 7.

The model summary for ARIMA (45, 0, 49) for forecasting the future point is shown in Table 24. Model is considered stable and invertible. Maximum iterations exceeded and Hessian is not positive. A comparison of all models is listed in Table 25. ARIMA (45, 0, 49) Akaike's Information Criteria (AIC) is better than many of the models, as is Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE). This satisfies the third specific aim and, with the ARIMA (45, 0, 49) R-square = .596, H_1 is accepted and H_0 is rejected.

Table 24. Patient #3 Model Summary for ARIMA (45, 0, 49).

Model Summary Parameter	Value
DF	-14
Sum of Squared Errors	0.007
Akaike's A Information Criterion	-101.222
Schwarz's Bayesian Criterion	126.252
R-square	0.595
R-square Adjusted	3.309
MAPE	5.064
MAE	0.042
-2LogLikelihood	-291.221

Table 25. Comparison of Models. Results of attempted time series autocorrelation models. ARIMA (45, 0, 49) is dark bordered.

Model	DF	Variance	AIC	SBC	RSquare	-2LogLH	Weights	MAPE	MAE
ARMA(60, 60)	40	-2.076e-7	-180.7284	108.99999	0.479	-422.7284	0.999650	5.141015	0.042
ARIMA(0, 0, 0)	80	0.0077507	-162.7964	-160.4019	0.000	-164.7964	0.000128	8.830931	0.072
AR(1)	79	0.0077672	-161.6315	-156.8426	0.010	-165.6315	0.000071	8.819699	0.072
ARMA(1, 1)	78	0.0076881	-161.4564	-154.2730	0.032	-167.4564	0.000065	8.712183	0.071
ARMA(2, 2)	76	0.0075686	-160.7106	-148.7383	0.070	-170.7106	0.000045	8.520167	0.069
ARMA(1, 2)	77	0.0077118	-160.2364	-150.6586	0.042	-168.2364	0.000035	8.668685	0.071
Simple Exponential Smoothing(Zero to One)	79	0.0078488	-155.3736	-152.9916	-0.05	-157.3736	0.000003	8.871462	0.073
ARMA(5, 5)	70	0.0072135	-152.9262	-126.5873	0.122	-174.9262	0.000001	8.255124	0.067
ARIMA(1, 1, 1)	77	0.007968	-152.4501	-145.3040	-0.04	-158.4501	0.000001	8.903106	0.073
Damped-Trend Linear Exponential Smoothing	77	0.0080527	-151.3736	-144.2275	-0.05	-157.3736	0.000000	8.871462	0.073
ARMA(10, 10)	60	0.0061602	-150.2985	-100.0151	0.285	-192.2985	0.000000	6.973500	0.057
ARMA(53, 53)	26	-8.305e-6	-149.2597	106.94633	0.549	-363.2597	0.000000	4.781040	0.040
Double (Brown) Exponential Smoothing	78	0.0086378	-142.6072	-140.2377	-0.21	-144.6072	0.000000	9.630992	0.078
Linear (Holt) Exponential Smoothing	77	0.0084945	-141.2297	-136.4908	-0.20	-145.2297	0.000000	9.588847	0.078
ARMA(51, 51)	22	-2.076e-5	-134.7255	111.90271	0.474	-340.7255	0.000000	5.173569	0.042
ARMA(15, 15)	50	0.00651	-133.9344	-59.70644	0.313	-195.9344	0.000000	6.962235	0.057
ARMA(50, 50)	20	-4.684e-5	-130.5128	111.32656	0.592	-332.5128	0.000000	4.811297	0.040
ARMA(49, 49)	18	-6.587e-5	-129.1002	107.95030	0.569	-327.1002	0.000000	4.879244	0.040
ARMA(20, 20)	40	0.0058566	-127.2516	-29.07920	0.403	-209.2516	0.000000	6.361471	0.052
ARI(1, 1)	78	0.0121836	-123.4736	-118.7096	-0.55	-127.4736	0.000000	10.422970	0.087
ARMA(55, 55)	30	-1.045e-5	-123.2838	142.50008	0.463	-345.2838	0.000000	5.192873	0.043
ARMA(51, 53)	24	-0.000027	-111.3781	140.03904	0.463	-321.3781	0.000000	5.313901	0.044
ARMA(45, 45)	10	-0.001103	-102.1348	115.76011	0.574	-284.1348	0.000000	5.084743	0.042
Seasonal Exponential Smoothing(12, Zero to One)	66	0.0101807	-101.3475	-96.90853	-0.50	-105.3475	0.000000	10.843574	0.088
ARMA(45, 49)	14	-0.000539	-101.2215	126.25117	0.596	-291.2215	0.000000	5.063732	0.042344
ARMA(40, 40)	0	.	-100.1649	93.785460	0.566	-262.1649	0.000000	5.248708	0.043
Winters Method (Additive)	65	0.0103616	-99.34915	-92.69063	-0.50	-105.3492	0.000000	10.843449	0.088
ARMA(47, 49)	16	-0.000435	-97.46418	134.79739	0.594	-291.4642	0.000000	4.952270	0.041
ARMA(46, 49)	15	-0.000613	-96.19599	133.67113	0.586	-288.196	0.000000	5.106025	0.042
ARMA(40, 49)	9	-0.003286	-84.01115	131.48928	0.580	-264.0111	0.000000	5.240258	0.043
ARMA(1, 49)	30	306.96439	866.42773	988.54464	-3e+4	764.42773	0.000000	1251.0717	10.47

Section 7. Validation with Clinician: Establishing the Ground Truth

Six clinicians were contacted regarding this study. The first, a primary care physician in Phoenix, AZ, consulted with the author at length regarding how to approach the clinicians, what type of words to use (e.g., minimize the “tech talk”), and general considerations, such as limiting the face-to-face interviews to only 15 minutes. A copy of the scripted interview and associated charts used during the interview are included in Appendix B. The student ensured clinician anonymity to lend candidness in their answers. Selections a, b, and c were transformed to 1, 2, and 3 respectively to provide quantitative analysis.

The remaining five clinicians were interviewed (four face to face, and one by phone). All five are practicing physicians from Texas Tech Health Sciences Center, Lubbock, TX. Four are pulmonologists and one (Clinician #1) is an anesthesiologist. Tables 26-29 summarize the quantitative responses to Questions 1-6 of the scripted interview. It was encouraging to have results equal or better than 2 on a scale of 1-3, especially Question 3. While increased sensitivity might have been helpful (i.e., a scale of 1-5), the author found that the scale of 1-3 to be distinctive and minimalist enough to provide time for each clinician to ask questions and provide comments within the time allotment. While 15 minutes was the minimal interview time, the average was about 30 minutes. The clinicians were engaged, curious, and anxious to help. An overall average of 2.23 points resulted from the interviews.

Table 26. Interview Questions and Clinician Responses.

#	Question	Clinician					Total	Avg
		1	2	3	4	5		
1	If statistical process control were provided to you in addition to a clinical threshold you provide, would you	3	2	2	2	2	11	2.2
	1) Definitely not use it for interpreting a patient's status							
	2) Might use it for interpreting a patient's status							
	3) Definitely would use if for interpreting a patient's status							
2	If statistical process control alone were provided to you without your clinical threshold, would you.	3	2	3	1	3	12	2.4
	1) Definitely not use it for interpreting a patient's status							
	2) Might use it for interpreting a patient's status							
	3) Definitely will use if for interpreting a patient's status							
3	What are your initial thoughts about using statistical process control with vital signs?	3	3	2	3	2	13	2.6
	1) This type of approach is not appropriate for vital signs							
	2) It's an interesting idea, but I will stay with current practice							
	3) I would like to see more of this type of approach in vital signs							
4	If you used statistical process control, would your clinical interpretation be	3	2	2	1	2	10	2.0
	1) No different than it is now							
	2) May be different than it is now							
	3) Very different than it is now							
5	If predicted points for vital signs were available to you, would you	3	2	1	2	2	10	2.0
	1) Definitely not use them							
	2) Might use them							
	3) Definitely would use them							
6	If predicted points for vital signs were available with statistical process control so that you could tell whether the patient was within his/her normal, would you	3	3	2	1	2	11	2.2
	1) Definitely not use them							
	2) Might use them							
	3) Definitely would use them							
	SUMMARY STATISTICS	18	14	12	10	13	67	2.23 TOTAL AVG
		SUM	SUM	SUM	SUM	SUM	SUM	
		3.0	2.3	2.0	1.7	2.2		
		AVG	AVG	AVG	AVG	AVG		

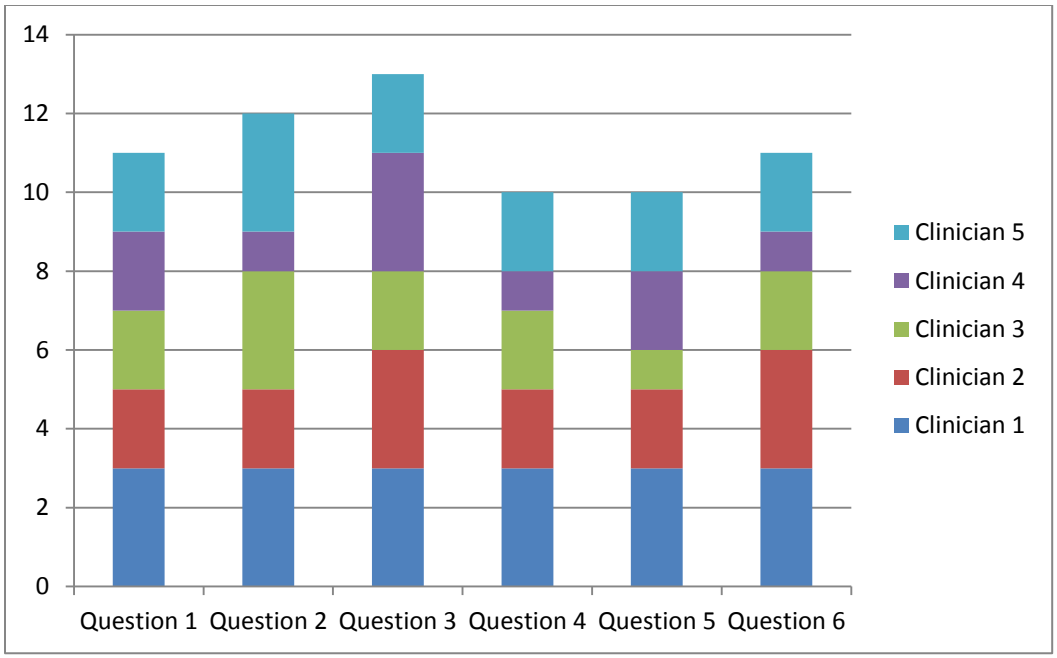


Figure 94. Stacked Chart Analysis of Clinician Responses to Questions 1-6. Minimum total response is 5, mean is 10, and maximum is 15 points for all clinicians for each question.

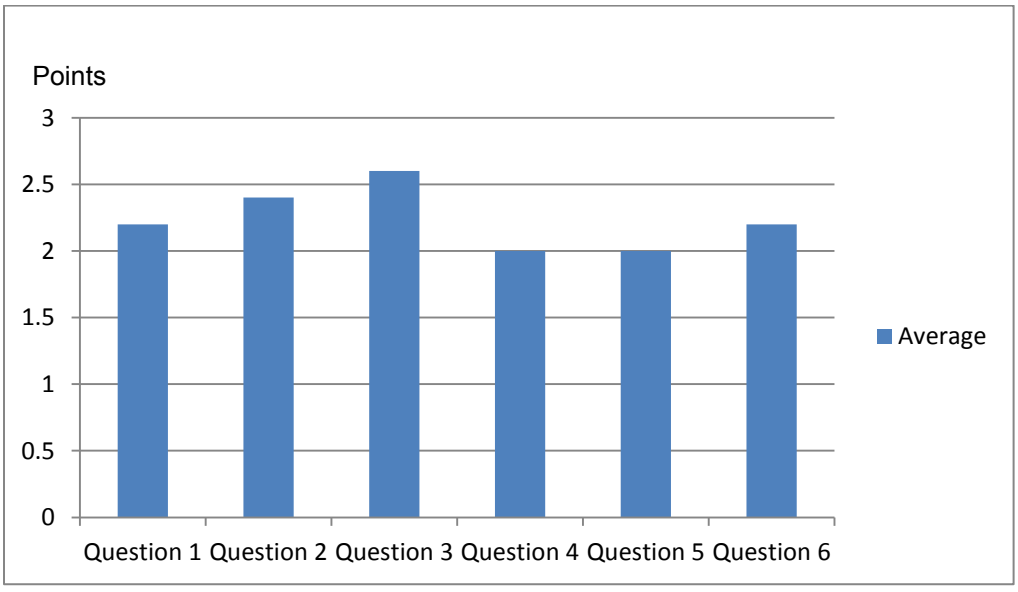


Figure 95. Average Point Response to Questions 1-6 by All Clinicians. Average is 2 for each question. Overall average was 2.23 points.

Most revealing were the qualitative comments (see Table 27) during the interviews, including that of Question 7: "What are your general thoughts about the use of this type of tool with vital signs to manage your patient's clinical outcomes?" General feeling about the topics of statistical process control and predictive analytics were "interesting" and ranged from decidedly excited about the possibilities (Clinician #1) to noticeably reserve or cautious (Clinician #4). Each clinician was definitely unique in understanding and embracing the entirety of the concept. Training on these type of tool(s) appeared to be a common theme, and reiterates what the original consulting physician observed.

Table 27. Qualitative Responses by Clinicians.

Clinician #	Question #	Comments/Quotes	Clinician Ranked the Question (Scale 1-3)
1	7	I'm very positive. This has been needed for a long time. Confidence using these type of tools will increase with time, especially with behavioral prompts for the patients. This would be extremely beneficial if I could have this and the other vital signs combined in the operating room.	n/a
2	3	We need experience to know if it's useful.	3
2	4	We need to look at the increases carefully.	2
2	6	I definitely would use them if we had a good track record.	3
2	7	I think it could help variability in the patient, especially when they have flare-ups but don't report them. If you had this type of information, you could intervene. Some patients are just indifferent.	n/a
3	1	We have to take into account reimbursement. It has the potential to be very useful. When a peak flow device company wants to give me free devices to give to all my patients, I don't take them as it is taking away my bread and butter of earnings.	2
3	2	It's a learning opportunity	3
3	3	Largely for financial reasons. Potentially very valuable but doesn't mesh well with current reimbursement model	2
3	4	Definitely expect to learn something. If fact, even my clinical ideas might help the statistics learn. We have a predicted value that's based on age, sex, and height and it doesn't change over time.	2
3	5	This isn't terribly helpful. I am more interested in a general trend, especially over shift changes, rather than a point here or there in time.	1
3	6	I am more interested in present and past then the future	2
3	7	There is a big future for this, but not today. It can become very valuable as medicine is changing. One on one patient care is going to have to go. This tool could decrease the number of patients and still generate revenue. It can be used as a physician extender; for example, in nursing homes there is an increased concern about outliers. Outliers can be potentially very valuable but current reimbursement models don't support. It is potentially viable under managed care scenarios, yes, now it will work. It is not good for solo practitioners, much better for group practice.	n/a
4	7	If we could account for variables, it would be great but even us physicians don't know. But not sure how I would use it in a clinical practice. Might be good for devices but not on a patient. We may or may not know the variables. It's interesting. ICU is a plethora of values, and it would be a useful tool and helpful in a controlled environment as I would know how my intervention impacts the numbers. Doing this for septic shock would be best, along with electronic ICU (remote), as we need to have many points with long term trend changes, not just a flag for a point as is done now.	n/a
5	1	I would also consider where they live, like out in rural New Mexico, this would be helpful.	2
5	4	I don't want to ignore the data but it is not good	2
5	7	It's complicated and hard to understand, versus something I can glance at. I would probably be retrained in order to use it. I'd like to learn more about it and learn about the clinical outcome.	n/a

Chapter 5. Conclusions and Recommendations

This study provided an opportunity to examine twenty COPD and/or CHF elders' peak flow FEV1 measurement in a telemedicine environment using an Intel Health Guide System® and its associated software. It attempted to meet three specific aims and prove four hypotheses, summarized in Table 28.

The first aim was to perform a Shewart-Western Electric analyses to examine statistically-derived threshold violations on FEV1 and compare to clinician-derived threshold violations for individual patients. This was accomplished by analyzing each FEV1 data point captured remotely from an Intel Health Guide System® following pre-processing (i.e., removal of outliers, double readings, etc.). Each data point was classified into one of four schemas: TP, FN, FP, and TN. The computed *z* value (1.1062) indicated a statistically valid difference ($p<.005$), rejecting H_0. Average increase from the clinician intervening when he shouldn't have statistically to the clinician not intervening when he should have statistically was 1.914% (% Increase = (FN-FP/total data points)*100). This results in less than an hour for a 40hr work week.

However, the seemingly low average increase is deceiving, particularly when the range is examined (-17.544 to 32.051% increase). This speaks to the wide variation seen in other statistics as well (i.e., skewness, mean, distribution, etc.) and may be minimized in the future by controlling sources of error. Sources of error may include age, gender, prescription drugs, clinician intervention, compliance, co-morbidity, sample size, and other COPD and/or CHF vital sign measurements cited in Table 4 and 8. It has been demonstrated that applying statistical process control to individual patient FEV1 measurements may provide an option to the clinician, whether replacing or adding to the

clinician-derived threshold, but additional studies to validate improved patient outcome using SPC is advised.

This study's second aim was to apply predictive algorithms to individual FEV1 measurements and compare the statistically- and clinician-derived threshold violations for the same population as described above. The algorithms and models show promise for training and validating the predictive model for future FEV1 vital signs. The training set resulted in an average Generalized R-square of 0.782, less than the targeted .80 for accepting H_1, and, therefore, H_0 is accepted for the training set. The validation Generalized R-square at .804, on the other hand, did meet the target of .80, and, therefore, H_1 is accepted for the validation set. It can be said, therefore, that a statistically significant learning classifier in the form of a random K-fold was achieved for the thresholds.

Once again, the variation observed in the Generalized R-square is noticeable: a range from .359 (extremely poor fit) to 1 (perfect fit) for the training set, and .156 to 1 for the validation set. Determining the root cause of this significant variation in developing the system model was elusive and one may conclude that sources of error played a key role in the wide range. Average number of misclassifications (7.05) and corresponding average training misclassification rate (.096, or 9.6%) were better than informally expected, and provides some confidence that predictive models for illnesses are possible. Further studies to experimentally block sources of variation and error are recommended.

Hypothesis C of Table 28, also supported the second aim of this study. This hypothesis compared pre- and post-predicted statistical zones (A (+/- 3σ), B (+/- 2σ), C (+/- 1σ), and this study added O (Out-of-Control per Western Electric Rules)). Two patients, #1 and #3, were analyzed and both R-square (.573 and .249, respectively) and Kappa (.703 and .147, respectively) for agreement between the pre- and post-predicted zone failed to meet H_0 hypothesis that the zones were the same; H_1 is accepted in both patients, indicating that there is a difference between pre- and post-predicted zones.

Additional analyses of the patient population were enlightening. Western Electric Rules 1, 2, and 6 accounted for 77.4% of the total Rule violations, indicating shifts in process means. Rule 2, nine points in a row in a single (upper or lower) side of Zone C or beyond, accounted for 42.5% of the total Rule violations alone. This information may be helpful to clinicians if they were cognizant that the probability of nine points in a row occurring in a single side of the mean occurring was less than .0039.

The relationship between disease severity and model fit was an interesting discovery. Patients segregated by disease state (mild, moderate, severe, and very severe) and their FEV1 data subjected to random K-fold indicated that severe and very severe were statistically a better fit to the model. Mild and moderate Generalized R-squares were compared to severe and very severe Generalized R-squares, and found to be statistically different, with Wilcoxon/Kruskal-Wallis *z* values and Chi-Squares at -1.746 and 3.204 for the training set, respectively, and at 1.138 and 1.399 for the validation set, respectively. The difference between the two groups (mild/moderate and severe/very severe) may be explained by the significant physiological consumption of COPD and/or CHF at later disease stages. A larger sample size with varying stages of disease and

sources of error controlled is recommended to validate these results found on twenty patients.

Wilcoxson/Kruskal-Wallis Tests were also calculated for variance vs. disease severity (Figure 29), standard error vs. disease severity (Figure 30), and percent increase from FP to FN vs. disease severity (Figure 39). All showed a significant difference between the mild/moderate and severe/very severe groups. The significance seen in the percent increase from FP to FN demonstrated that the severe/very severe population was actually less (mean=1.13) than the mild/moderate group (mean=4.28). However, with the $p>|z| = -0.4314$ and $p>$Chi-Square = 0.4063, the difference is not strongly significant and further tests with more patients are recommended. Other Wilcoxson/Kruskal-Wallis Tests were calculated for disease severity and coefficient of variance (Figure 31), skewness (Figure 32) , Kurtosis (Figure 33), precision (Figure 34), and recall (Figure 35) did not show a significant difference between mild/moderate and severe/very severe groups.

In answering hypothesis D of the third specific aim as to whether an individual's FEV1 measurement can be forecasted prior to its occurrence, many models were explored (Table 25). The mixed ARIMA (45, 0, 49) model with an optimized R-square fit of 0.596 showed encouraging results in forecasting up to 10 points in advance. Fewer points to forecast were examined, but no improvement in the model occurred. It is recommended that, as larger sample sizes of patients are secured with sources of error identified and blocked for analyses, additive models be explored.

Table 28. Summary of Hypotheses, Specific Aims, and Results.

Hypo-thesis	Specific Aim	H_0	H_1	Source	Result
A	1	There will not be a statistically significant increase of clinical interventions when statistical thresholds are applied (False Negatives) prior to application of predictive algorithms. FN=FP	There will be a statistically significant (p<.005) increase of clinical interventions when statistical thresholds are applied (False Negatives) prior to application of predictive algorithms. FN≠FP	Final Construct, Feature Extraction and Selection using statistical process control	H_0 is rejected, based on z = 1.106. H_1 is statistically valid and true. There is a statistically significant increase of clinican interventions when statistical thresholds are applied. All patients' data examined.
B	2	An FEV1 classification cannot be predicted, demonstrating less than an .80 Generalized R-square.	An FEV1 classification can be predicted demonstrating ≥ .80 Generalized R-square.	Final Construct, System Evaluation using random k-fold	H_0 is accepted for the training set (Generalized R-square = .782), but H_0 is rejected for the validation set (Generalized R-square = .804). All patients' data examined.
C	2	The pre- and post-predicted data point will not demonstrate a statistical difference in zone designation.	The pre- and post-predicted data point will be statistically different (p<.005) in zone designation.	Final Construct, System Evaluation	Patient #1: H_1 is accepted based on R-square of .573 and Kappa of .703 (p<.005) for agreement. Therefore, H_0 is rejected. Patient #3: H_1 is accepted based on R-square of .249 and Kappa of .147 (p<.005) for agreement. Therefore, H_0 is rejected.
D	3	FEV1 cannot be forecasted, demonstrating less than .60 R-square	FEV1 can be forecasted, demonstrating ≥.60 R-square	Final Construct, System Evaluation, using ARIMA (45,0,49)	Patient #3: H_1 is accepted based on R-square = .596. Therefore, H_0 is rejected.

Clinician interviews were positive and generally above average for the tools' potential capabilities; therefore, ground truth is validated based on the sample size (five clinicians). Increased clinician input would be critical for development of this tool(s). It would be imperative for developing such software to insure ease of use for physicians to enhance user experience as well as integration into the physicians' course of patient analysis and outcome; it was clear during the interviews that statistics and/or analytics were areas that clinicians were not entirely convinced nor comfortable, but definitely willing to consider new approaches with further data and training. The clinician accepted statistical process control (Questions 1-4) more than predictive analytics (Questions 5-6), but further validation is needed given the small sample size of clinicians. A prototype software package may be a solution to increased acceptance of predictive FEV1 or other vital signs.

Based on this study, therefore, it can be concluded that it is possible and significant to compare statistically-derived thresholds to clinician-derived thresholds. Although the increase in work effort by the clinician is less than 2%, the clinical implications to patient outcome do need to be explored further. It can also be concluded that, by using a random K-fold neural net, validation of the predicted classification can be a relatively strong learning model. While this study was not able to demonstrate agreement in zones before and after the predicted model, the ARIMA model was able to forecast future points with some confidence. Lastly, it can also be concluded that disease states do have a statistical impact on model fit.

Other analytic approaches may be considered for increase in model fit. Applying support vector machines (SVMs) may be an important option to consider for

classification. SVMs provide an optimal hyperplane for linearly separable patterns, but can be extended to patterns that are not linearly separable by transformation of original data to map into a new space (e.g., Kernel Function). It is a pattern recognition and classification tool that may be a more elegant representation than random K-fold or ARIMA models. Bundling different models, particularly with the grouping of many vital signs to ascertain patient status, may be an area for further study.

By controlling sources of error and exploring other algorithms, the role of statistics and advanced analytics can and should be explored further in better managing chronic diseases such as COPD and CHF. With the increase of captured patient data from new sources such as mobile and telemonitoring form factors, it may be one of the more plausible options to deal with the growing number of elders in the world.

REFERENCES

Addicam, Sanjay. *Building Intelligent Systems.* China: Intel Press, 2012.

Akselrod, S. "Power Spectrum Analysis of Heart Rate Fluctuations: A Quantitative Prove of Beat-to-Beat Cardiovascular Control." *Science*, 1981: 220-222.

AMA. *Find-a-Code: HCPCS Procedure and Supply Codes.* 2012. http://www.findacode.com/code.php?set=HCPCS&c=S9110 (accessed February 6, 2013).

American Heart Association. *About Heart Failure.* August 20, 2012. http://www.heart.org/HEARTORG/Conditions/HeartFailure/AboutHeartFailure/About-Heart-Failure_UCM_002044_Article.jsp (accessed January 31, 2013).

American Telemedicine Association. *What is Telemedicine.* 2012. http://www.americantelemed.org/learn/what-is-telemedicine (accessed January 26, 2013)
.

Anderson, G. "The Growing Burden of Chronic Disease in America." *Public Health Reports*, 2004: 263-270.

Bansal, Dipali. "Statistical Analysis if Human Physiological Paramter Usign Six Sigma Techniques." *Proceedings of the World Congress on Engineering.* London, United Kingdom: World Congress of Engineering, 2008.

Barbieri, Riccardo. "A Point-process Model of Human Heartbeat Intervals: New Definitions of Heart Rate and Heart Rate Variability." *American Journal of Physiology Heart: Circulation Physiology*, 2005: 209-212.

Benatar, Daniel. "Outcomes of Chronic Heart Failure." *Archives of Internal Medicine*, 2003: 347-352.

Clark, David E. "Monitoring Hospital Trauma Mortality Using Statistical Process Control Methods." *Journal of the American College of Surgeons*, June 1998: 630-635.

Clinical Cases and Images. January 24, 2012. http://clinicalcases.org/2005/07/pulmonology-and-critical-care-cases.html (accessed January 30, 2013).

Dale, J. "An Evaluation of the West Surrey Telemedicine Monitoring Project." *Journal of Telemedicine and Telecare*, 2003: 39-41.

Dishman, Eric. "Written Testimony to Health I.T. Policy Committee, Meaningful Use Wokrgroup." *Panel on Meaningful Use of HIT in Real Lives of Patients and Familieis.* 2010.

Dishongh, Terrance J. *Wireless Sensors for Healthcare Applications.* Norwood, MA: Artech House, 2010.

Dolan, Brian. "Health Apps, Devices Generated $718M in 2011." *Mobihealthnews.com.* January 12, 2012. http://mobihealthnews.com/15803/health-apps-devices-generated-718m-in-2011/ (accessed February 6, 2013).

Dolan, Brian. "Report: 13K iPhone Consumer Health Apps in 2012." *mobihealthnews.com.* September 22, 2011. http://mobihealthnews.com/13368/report-13k-iphone-consumer-health-apps-in-2012/ (accessed February 6, 2013).

Duda, Richard O. *Pattern Classification, Second Edition.* New York City, New York: Wiley Interscience, 2001.

Ekeland, Anne G. "Effectiveness of Telemedicine: A Systematic Review of Reviews." *International Journal of Medical Informatics* 79, no. 11 (2010): 746-771.

Eren, Ali. "A Decision Support System for Telemedicine THrough the Mobile Telecommunication Platform." *Journal of Medical Systems*, 2008: 31-35.

Evans, Melanie. "Coming Up Short." *Mondern Healthcare*, 2004: 14.

Falas, T. "A Review of Decision Support Systems in Telecare." *Journal of Medical Systems*, 2008: 347-356.

Fodor, Adriana. "Telemedicine for the Diabetic Foot: A Model for Improving Medical Care, Developing Decision Support Systems, and Reducing Medical Cost." *Web Based Applications in Healthcare and Biomedicine, Annals of Information Systems, Spring Science and Business Media*, 2010: 243-258.

Fralick, Celeste R. "Phoenix Chapter of American Society of Quality." *Scottsdale Mayo Clinic Report on Lean Implementation in Laboratory.* 2006.

Garg, Manoj. "Multimodal Analysis of Body Sensor Network Data Streams for Real-time Healthcare." *MIR '10 Proceedings of the International Conference on Multimedia Information Retrieval.* Philadelphia, Pennsylvania: MIR, 2010. 469-477.

Global Initiative for Chronic Obstructive Lung Disease. "Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease." *GOLD - The Global Initiative for Chronic Obstructive Lung Disease.* 2013. http://www.goldcopd.org/uploads/users/files/GOLD_Report_2013.pdf (accessed February 12, 2013).

Growth House, Inc. *Caregiver Guides.* 2012. http://www.growthhouse.org/chfcopd.html (accessed January 31, 2013).

Haesum, Lisa K.E. "Cost-Utility Analysis of a Telerehabilitation Program: A Case Study of COPD Patients." *Telemedicine and e-Health*, 2012: 688-692.

Health Data Management. *Report: Telemedicine Market to Hit $3.6B.* October 8, 2009. http://www.healthdatamanagement.com/news/telemedicine-39184-1.html (accessed January 26, 2013).

Healthcare IT News. *Global Telemedicine on Growth Track.* August 30, 2011. http://www.healthcareitnews.com/news/global-telemedicine-market-growth-track (accessed January 26, 2013).

Helal, Abdelsalam. "Smart Home-Based Health Platform for Behavioral Monitoring and Alteration of Diabetes Patients." *Journal of Diabetes Science and Technology*, 2009: 141-148.

Hokan, Ali. "A Low-cost Mobile Adaptive Tracking System for Chronic Pulmonary Patients in Home Enviornement." *Telemedicine and e-Health*, 2013: 24-30.

Intel Corporation. *Intel Chips Timeline.* 2012. http://www.intel.com/content/www/us/en/history/history-intel-chips-timeline-poster.html (accessed March 30, 2013).

Intel Corporation. *Intel Healthcare.* 2010. www.intel.com/healthcare/telehealth/index.htm (accessed 2010).

Jensen, Morton Hasselstrom. "Clinical Impact of Home Telemonitoring on Patients with Chronig Obstructive Pulmonary Disease." *Telemedicine and e-Health*, 2012: 674-776.

Khoor, S. "Internet-based, GPRS, Long-term ECG Monitoring and Non-linear Heart-rate Analysis for Cardiovascular Telemedicine Management." *Computers in Cardiology*, 2003: 424-435.

Kleiger, Robert L. "Decreased Heart Rate Variability and Its Association With Increased Mortality After Acute Myocardial Infarction." *The American Journal of Cariology*, 1987: 256-262.

Kotagal, Meera. "Improving Quality in Resource Poor Settings: Observational Study From Rural Rwanda." *British Medical Journal*, 2009: 3488.

Kyriacou, E. "Multi-purpose Healthcare Telemedicine Systems with Mobile Communication Link Support." *Biomedical Engineering OnLine*, 2003: 2003.

Latifi, R. *Current Principles and Practices of Telemedicine and e-Health.* IOS Press, 2000.

Lee, Sang-goo. "Telemedicine: Challenges and Opportunities." *Journal of High Speed Networks*, 2000: 15-30.

Li, Huaming. "Body Sensor Networks-based Sensor Fusion forCardiovascular Biosignal Prediction." *HealthNet*, 2008.

Lisetti, Christine Laetitia. "Using Noninvasive Wearable Computers to Recognize Human EMotions from Physiological Signals." *EURASIP Jounral on Applied Signal Processing*, 2004: 1672-1687.

Lucas, James M. "Combined Shewart-CUSUM Quality Control Schemes." *Journal of Quality Technology*, 1982: 51-59.

Lutz, Wolfgang. "The Coming Acceleration of Global Population Ageing." *Nature*, 2008: 716-719.

Lymberis, A. "Smart Wearables for Remote Health Monitoring, From Prevention to Rehabilitation: Current R&D, Future Challenges." *Proceedings of the 4th Annual IEEE Conference on Information Technology Applications in Biomedicine.* IEEE, 2003. 272-275.

Malic, M. "Heart Rate Variability." *Annals of Noninvasive Electrocardiology*, 2006: 151-181.

McNames, James. "Cardiovascular Signal Decomposition and Estimation with the Extended Kalman Filter." *Proceedings of the 28th IEEE EMBS Annual International Conference.* New York City, New York: IEEE, 2006. 3708-3711.

McNames, James. "Statistical Modeling of Cardiovascular Signals and Parameter Estimation Based on the Extended Kalman Filter." *IEEE Transactions on Biomedical Engineering*, 2008: 119-129.

Mechanic, David. "The Changing Elderly Population and Future Health Care Needs." *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, 1999: 24-38.

Merritt, David. *Paper Kills.* Washington, D. C. : CHT Press, 2007.

Montgomery, Douglas. *Applied Statistics and Probabilities for Engineers.* New York City, NY: John Wiley & Sons, 1994.

Montgomery, Douglas. *Applied Statistics and Probability for Engineers.* New York City: John Wiley and Sons, 1994.

Montgomery, Douglas. *Introduction to Statistical Quality Control, Fifth Edition.* Hoboken, New Jersey: John Wiley & Sons, 2004.

Mower, Emily. "Investigating Inplicit Cues for User State Estimation in Human-robot Interaction Using Physiological Monitors." *16th IEEE International Conference on Robot & Human Interactive Communication.* Jeju, Korea: IEEE, 2007. 1125-1130.

Myers, Sue. "Impact of Home-Based Monitoring on the Care of Patients with Congestive Heart Failure." *Home Health Care Management & Practice*, 2006: 444-451.

Net Wellness. *Consumer Health Information for COPD.* May 20, 2008. http://www.netwellness.org/healthtopics/copd/overview.cfm (accessed January 31, 2013).

Oeff, M. "Monitoring Multiple Cardiovascular Parameters Using Telemedicine in Patients with Chronic Heard Failure." *Herzschrittmacherther Elektrophysiol*, 2005: 150-158.

Ollapp, SwEng LLC. *HeartWise Blood Pressure Tracker.* 2012. http://www.ollapp.com/app/heartwise-blood-pressure-tracker/android (accessed January 29, 2013).

Pantelopoulos, Alexandros. "Prognosis - A Wearable Health-Monitoring System for People at Risk: Methodology and Modeling." *IEEE Transactions on Information Technology in Biomedicine*, 2010: 610-623.

Pare, G. "Cost-minimization Analysis of a Telehomecare Program for Patients with Chronic Obstructive Pulmonary Disease." *Telemedicine Journal and e-Health*, 2006: 114-121.

PricewaterhouseCoopers Health Research Institute. *What Works: Healthing the Healthcare Staffing Shortage.* 2007. www.pwchealth.com/hrj (accessed 2012).

Pyzdek, Thomas. *The Six Sigma Handbook.* New York: McGraw-Hill, 2003.

Quinn-Curtis, Inc. *Western Electric (WECO) and other Named Rule Sets.* 2012. http://www/quinn-curtis.con/SPCNamedRuleSets.htm (accessed January 29, 2013).

Rodriguez, J. "MOLEC: Anywhere and At Any Time Arrhythmia Classification." *Computers in Cardiology*, 2005: 607-610.

Rodriguez, Jimena. "Real-Time Classification of ECGs on a PDA." *IEEE Transactions on Information Technology in Biomedicine*, 2005: 25-34.

Society of Technical Communication. "Usability Toolkit." *Usability and User Experience, An STC Community.* 2013. http://www.stcsig.org/usability/resources/toolkit/toolkit.html (accessed February 15, 2013).

Stachura, Max. *Telehomecare and Remote Monitoring: An Outcomes Overview.* Industry, Washington, D.C.: Advamed, 2007.

Stetson, P. F. "Independent Component Analysis of Pulse Oximetry Signals." *Proceedings of the 26th Annial International Conference of the IEEE EMBS.* San Francisco, California: IEEE, 2004. 231-234.

The Merck Manual Home Health Handbook. *Chronic Obstructive Pulmonary Disease.* December 2008. http://www.merckmanuals.com/home/lung_and_airway_disorders/chronic_obstructive_pulmonary_disease_copd/chronic_obstructive_pulmonary_disease.html?qt=copd&sc=&alt=sh (accessed January 31, 2013).

Theodoridis, Sergios. *Pattern Recognition, Fourth Edition.* Burlington, Massachusettes: Elsevier, Inc., 2009.

United States Government. *Office of the National Coordinator in Health Care.* 2012. www.hhc/healthit/http://healthit.hhs.gov.portal (accessed 2012).

Western Electric Company. *Statistical Quality Control Handbook.* Charlotte, North Caroline: Delmar Printing Company, 1956.

WinterGreen Research. *Telemedicine Monitoring: Market Shares, Strategies, and Forecasts Worldwide, 2012-2018.* Market Research Report, Lexington, Massachusettes: WinterGreen Research, Inc, 2012.

Yang, Ping. "An Adaptive Cusum Test Based on a Hidden Semi-Markov Model Change Detection in Non-invasive Mean Blood Pressure Trend." *Proceedings of the 28th IEEE EMBS Annual International Conference.* New York City, New York: IEEE, 2006. 3395-3398.

Yang, Ping. "The Use of Mathematical Models and Clinical Knowledge in Interoperative Vital Sign Trend Monitoring." *30th Annual International IEEE EMBS Conference.* IEEE, 2008. 1028.

Young, Terry. "Using Industrial Processes to Improve Patient Care." *British Journal of Medicine*, 2004: 162-164.

Zarocostas, John. "Use of Telemedicine Will Double In The Next Decade, Report Predicts." *British Medical Journal*, 2010: 1634.

Zimmerman, Sheryl. *Assisted Living: Needs, Practices, and Policies in Residential Care for the Elderly.* Baltimore: The Johns Hopkins University Press, 2001.

APPENDIX A

INTERVIEW SCRIPT FOR CLINICIAN VALIDATION

Hello, my name is Celeste Fralick. I am a doctoral student in Biomedical Engineering at Arizona State University. I am studying the impact of statistics and predictive analytics in remote telemonitoring of elders with COPD and/or CHF. My data is limited to only peak flow output, FEV1 (Forced Expiratory Volume in 1 second), and I have no other information about the patients. The FEV1 was captured automatically by their telemonitoring unit at home, and the time span is about 1 – 1.5 years.

I would appreciate your candid answers to 7 questions. This should take about 15 minutes to complete. Your answers will be used, but you will remain anonymous. If you have any questions during the exercise, please don't hesitate to ask. I will record your answers, and may take notes during our conversation.

The first three charts (show Charts 1, 2, and 3) show three different patient's FEV1 average (mean) in green and 3 standard deviations away from the average (mean) in red. Upper Control Limit or UCL is +3 standard deviations. Lower Control Limit or LCL is -3 standard deviations. Each chart is personalized, as each person has their own average and standard deviation.

A medical threshold was placed on each patient, so the clinician could be notified if the patient went below this number and he/she could intervene. The next three charts (of the same patients) include that threshold (show Charts 4, 5, 6).

in the 1950's a method known as statistical process control (aka SPC) was developed that identified the probability of a point occurring in a chart. Eight Rules were created to

help the reader of a chart know when to intervene, as some probabilities of a point occurring are less than .4%.

In the same three patients, these Rules were applied and are indicated by a Rule number that failed (show Charts 7, 8, 9). Most of the Rules that failed in these charts are Rules that indicate a shift in the mean or a shift in the standard deviation. (If asked to see the Rules, provide them.) The clinician is expected to intervene when a Rule fails, as the probability of that point occurring is so small. Note that some clinician threshold points do not fail if statistical process control and the Rules are applied. This infers that the patient's data is within their normal variation according to statistical process control and the Rules about probability.

The following questions assume that statistical process control and its Rules are provided in an automated format for ease of use.

Questions:

1. If statistical process control were provided to you *in addition* to a clinical threshold you provide, would you
    a) Definitely not use it for interpreting a patient's status
    b) Might use it for interpreting a patient's status
    c) Definitely would use if for interpreting a patient's status
2. If statistical process control *alone* were provided to you *without* your clinical threshold, would you.
    a) Definitely not use it for interpreting a patient's status
    b) Might use it for interpreting a patient's status
    c) Definitely will use if for interpreting a patient's status

3. What are your initial thoughts about using statistical process control with vital signs?
    a) This type of approach is not appropriate for vital signs
    b) It's an interesting idea, but I will stay with current practice
    c) I would like to see more of this type of approach in vital signs
4. If you used statistical process control, would your clinical interpretation be
    a) No different than it is now
    b) May be different than it is now
    c) Very different than it is now

One more chart, and we'll be done. Chart 10 shows a patient with their actual FEV1 in blue provided for points 1 – 81, and their predicted FEV1 in red for points 1 – 81, and new predictions 82 - 92 (show Chart #10). So the blue line is before prediction and the red line is after prediction.

The following questions assume that predicted points and statistical process control would be automated for ease of use

5. If predicted points for vital signs were available to you, would you
    a) Definitely not use them
    b) Might use them
    c) Definitely would use them
6. If predicted points for vital signs were available with statistical process control so that you could tell whether the patient was within his/her normal, would you
    a) Definitely not use them
    b) Might use them
    c) Definitely would use them

7. What are your general thoughts about the use of this type of tool with vital signs to manage your patient's clinical outcomes?

Do you have any final comments or questions? Thank you for your time and consideration. This concludes the interview.

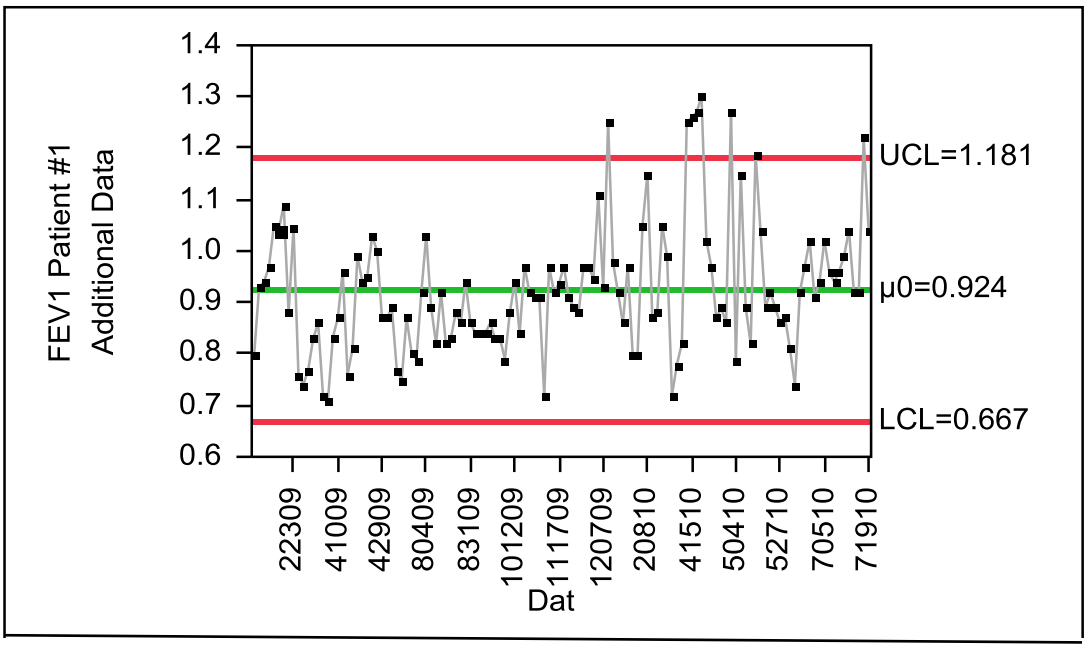


Chart 1. Patient #1

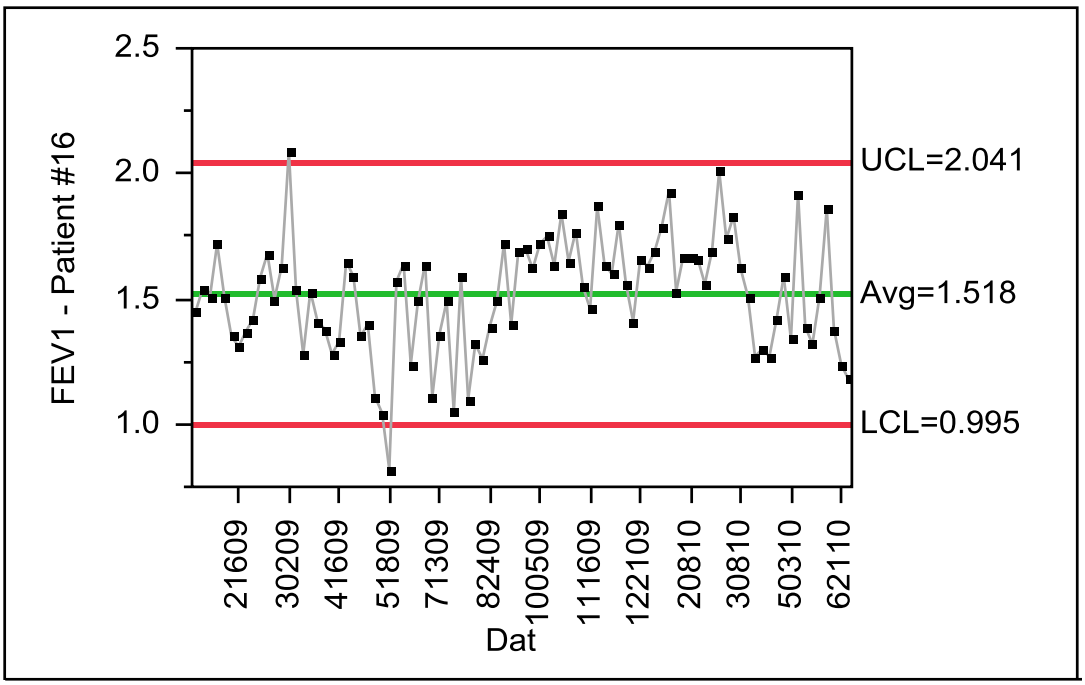


Chart 2. Patient #16

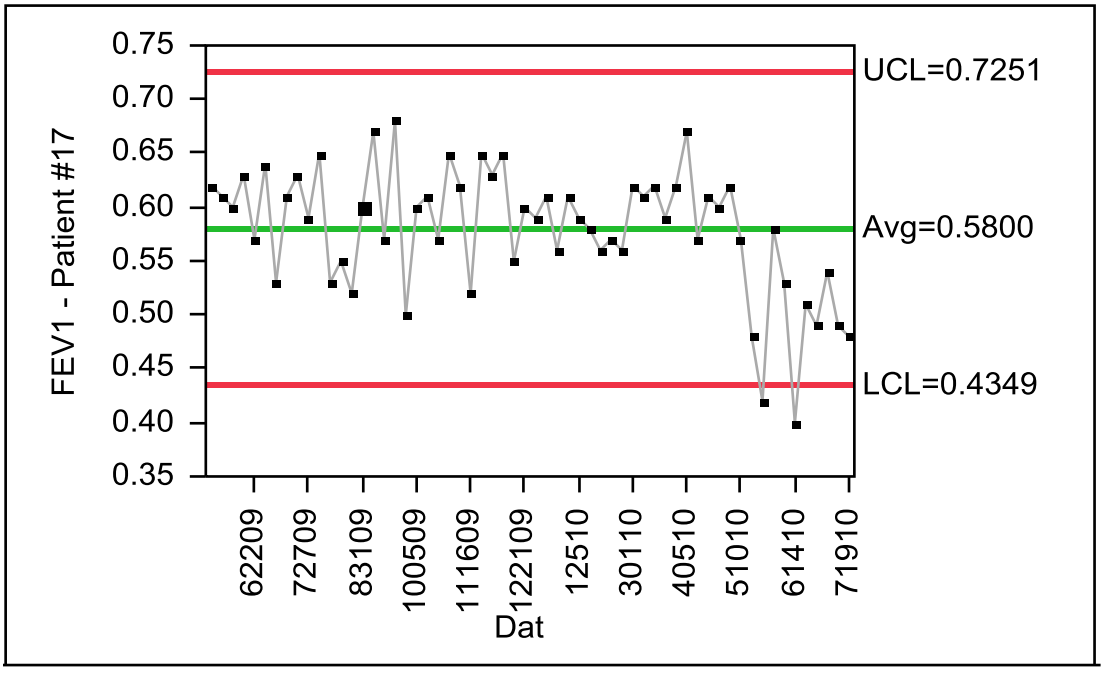


Chart 3. Patient #17

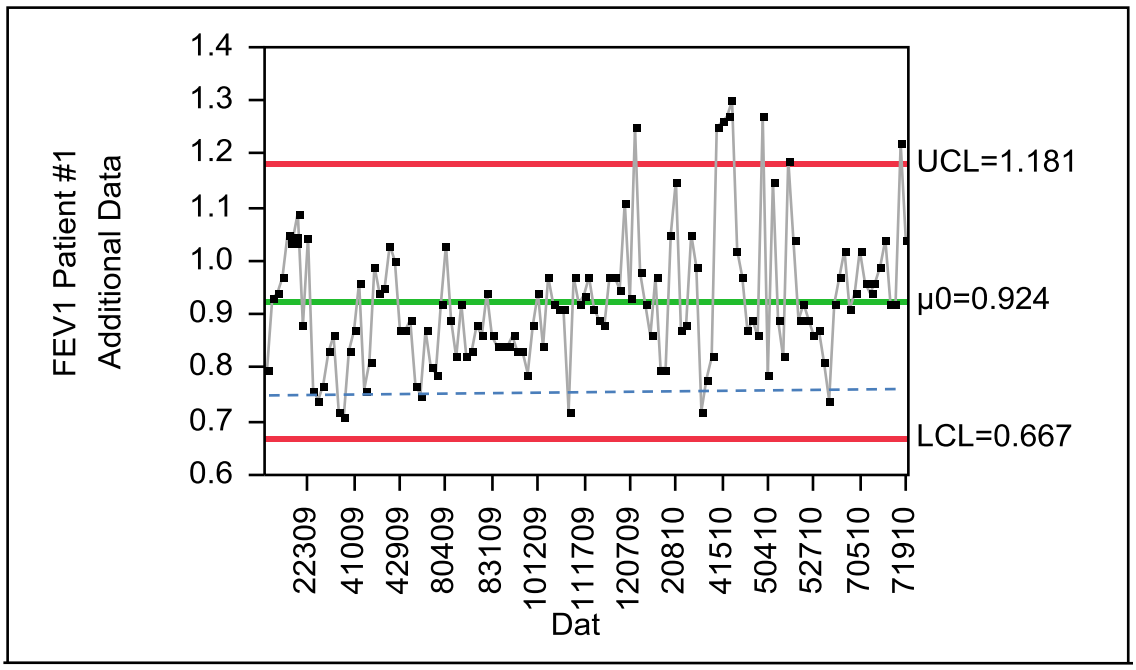


Chart 4. Patient #1 (Clinician Threshold = .75)

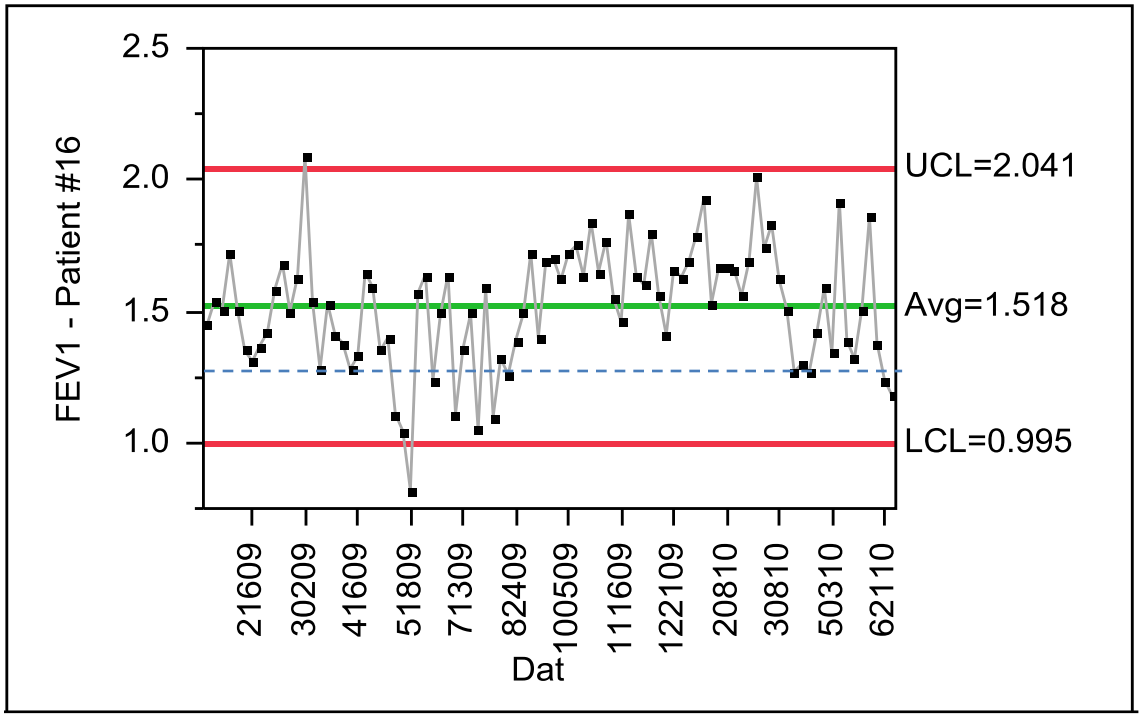


Chart 5. Patient #16 (Clinician Threshold = 1.25)

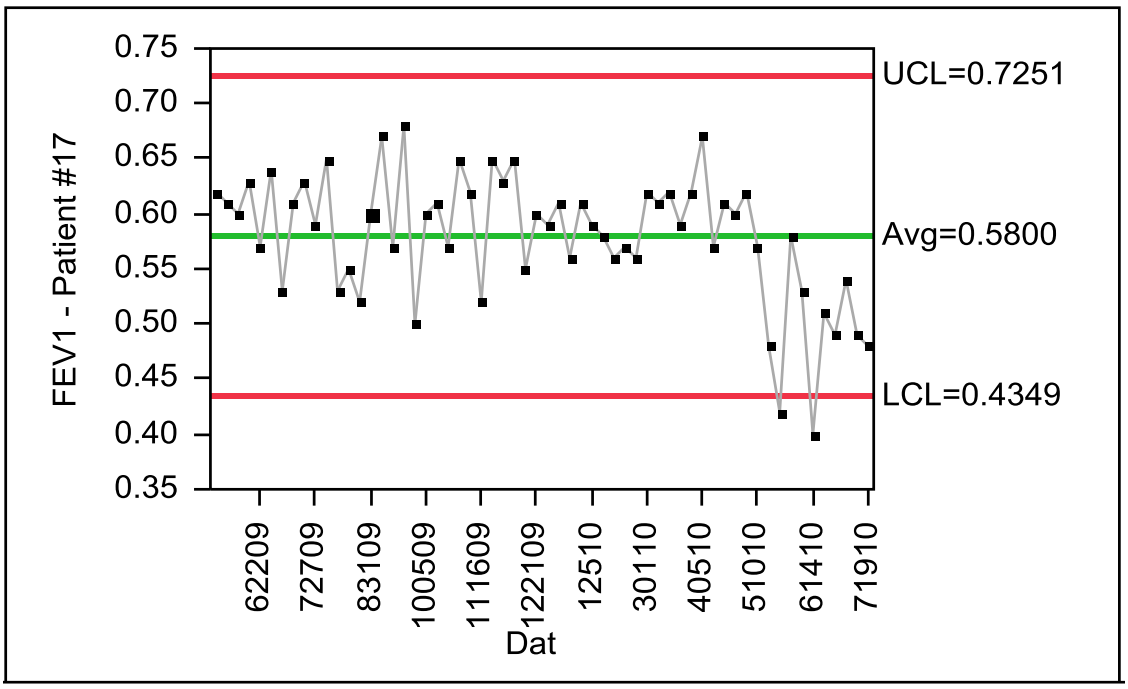


Chart 6. Patient #17 (Clinician Threshold = .52)

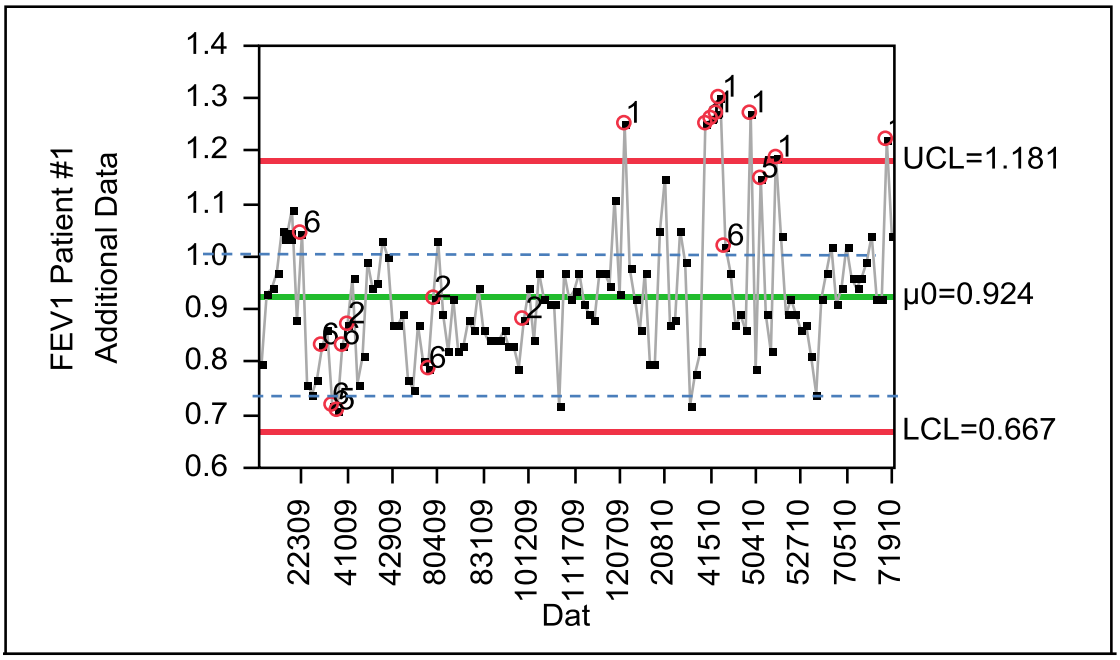


Chart 7. Patient #1 (Clinician Threshold = .75)

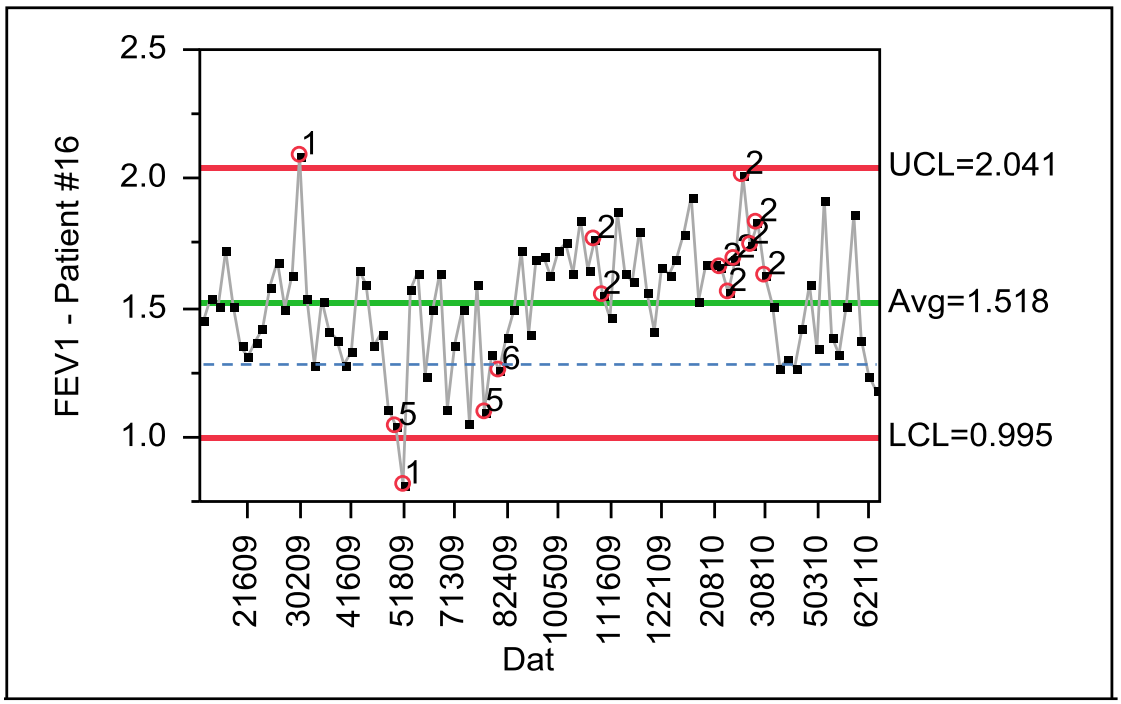


Chart 8. Patient #16 (Clinician Threshold = 1.25)

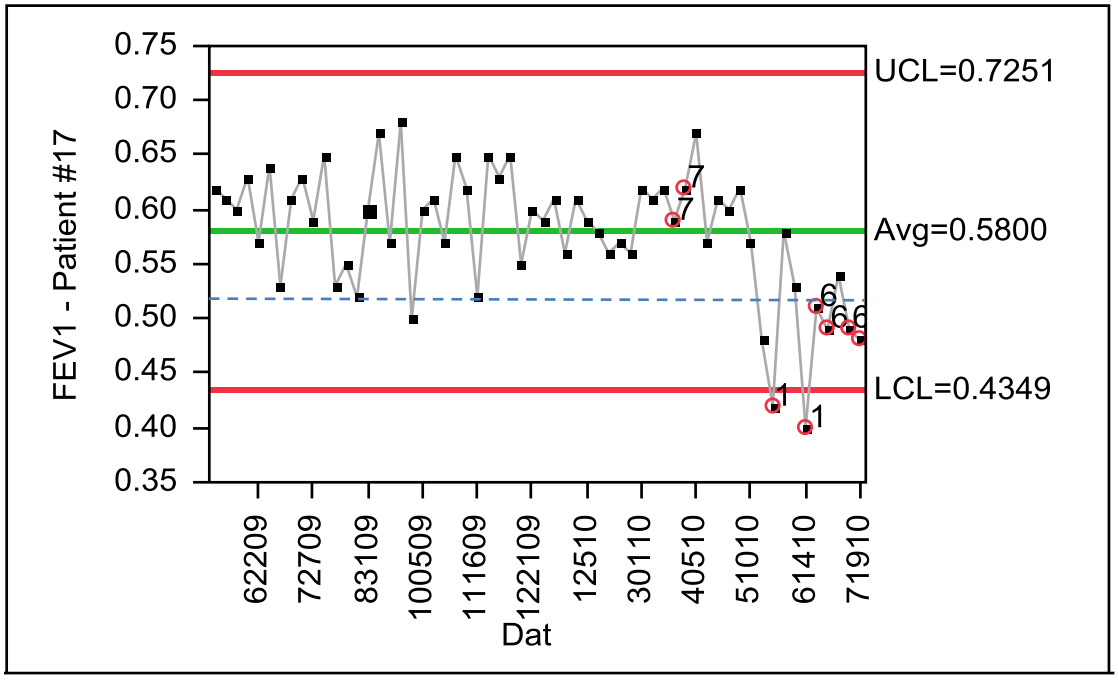


Chart 9. Patient #17 (Clinician Threshold = .52)

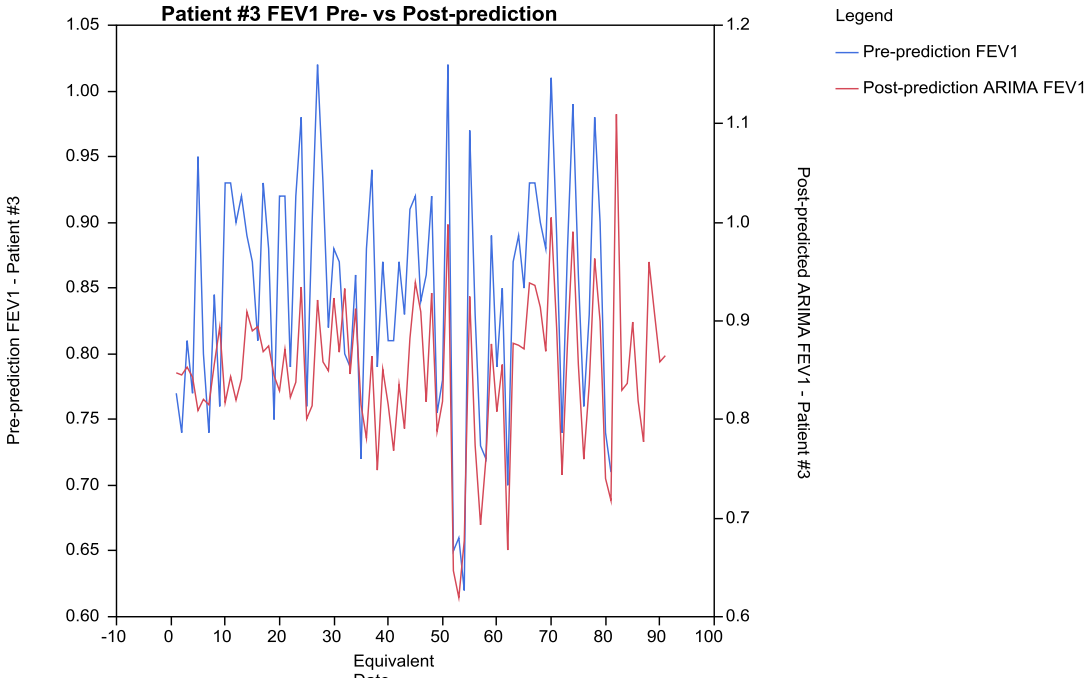


Chart 10. Patient #3 Predicted FEV1 Values

Rule	Point Location	Detection
Rule 1	One point beyond Zone A	Detects a shift in the mean, an increase in the standard deviation, or a single aberration in the process.
Rule 2	Nine points in a row in a single (upper or lower) side of Zone C or beyond	Detects a shift in the process mean.
Rule 3	Six points in a row steadily increasing or decreasing	Detects a trend or drift in the process mean. Small trends will be signaled by this test before Test 1.
Rule 4	Fourteen points in a row alternating up and down	Detects systematic effects such as two alternately used machines, vendors, or operators.
Rule 5	Two out of three points in a row in Zone A or beyond and the point itself is in Zone A or beyond.	Detects a shift in the process average or increase in the standard deviation. Any two out of three points provide a positive test.
Rule 6	Four out of five points in a row in Zone B or beyond and the point itself is in Zone B or beyond.	Detects a shift in the process mean. Any four out of five points provide a positive test.
Rule 7	Fifteen points in a row in Zone C, above and below the center line	Detects stratification of subgroups when the observations in a single subgroup come from various sources with different means.
Rule 8	Eight points in a row on both sides of the center line with none in Zones C	Detects stratification of subgroups when the observations in one subgroup come from a single source, but subgroups come from different sources with different means.

BIOGRAPHICAL SKETCH

While pursuing her PhD at ASU, Celeste Fralick has been a Staff Architect/Principal Engineer in the Intelligent Systems Group and the Digital Health Group at Intel Corporation. She currently insures embedded (intelligent) products are connected, managed, and secured in the broad market and medical, industrial, energy, military, aviation, and government market segments. Celeste has been a technical leader in Intel's biotech strategies and development of a Class II regulated medical device, product qualification, and life cycle programs. Active in various industry boards, journal editorial staffs, and consortiums, her experience spans semiconductor process and product development, regulatory, medical device, quality and reliability technology over the past 33 years. She will be using her PhD to complement her future work in industry, particularly in predictive analytics.