8-27-2009

Using control charts for computer-aided diagnosis of brain images

Sumner Williams IV

Sumner H. Williams IV

Candidate

Mathematics and Statistics

Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

,Chairperson

Using Control Charts for Computer-Aided Diagnosis of Brain Images

by

Sumner Williams

B.S., Mathematics, University of New Mexico, 2003

THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Statistics

The University of New Mexico

Albuquerque, New Mexico

March, 2009

Dedication

To my parents, family members, and professors that let me find my own way.

Acknowledgments

Aparna Huzurbazar, Phd. has been instrumental in the completion of this document and has painfully read, corrected, and reviewed multiple iterations. My mom has done the same.

Using Control Charts for Computer-Aided Diagnosis of Brain Images

by

Sumner Williams

B.S., Mathematics, University of New Mexico, 2003

M.S., Statistics, University of New Mexico, 2009

Abstract

We consider a novel application of quality control charts to brain scans performed using magnetic resonance imaging (MRI). Although our primary focus is on volume measures resulting from brain scans, issues related to the MRI scanner are also considered. The project evaluates a population of healthy control subjects using control charts to assess the MRIs obtained for the subjects. The results demonstrate our ability to automatically detect brain volumes that are statistical outliers, and can provide a potential cost savings of 10% for a moderately sized study. More importantly, our applied results will increase the sensitivity of tests from comparisons made between subject populations due to the fact that certain populations of subjects can be quickly, efficiently, and automatically identified based on outlier analysis.

Contents

List of Figures

List of Tables

Chapter 1

Computer-Aided Diagnosis

1.1 Computer's Role in Disease Detection

Computer-aided diagnosis of neuroimaging data is attracting an increasing amount of interest, funding, and research. The use of control charts with morphological measurements is a novel application in the neuroimaging field. Journals in the medical field are dedicating a portion of their space to statistics and quality in medicine. These articles are intended to promote better and less costly health-care delivery, and detecting disease more efficiently with fewer errors contributes to that goal. Using control charts with brain morphometrics is an as yet untouched field of research, but as this thesis will demonstrate, it is one which may be replete with opportunity for sampled populations as well as individuals. A sampled population can be more specifically defined to require both a diagnosed mental illness and the structural changes that are expected.

For schizophrenia, this would include criteria from the Diagnostic and Statistical Manual of Mental Disorders (DSM IV), the researcher's own tests, and expected morphological changes such as enlarged ventricles.[1] They normally increase in size with age, but are often seen to be enlarged in schizophrenic subjects. Figure 1.1

[1]The ventricles are located centrally in the brain and hold the cerebral spinal fluid.

gives more details about how the brain is labeled. With quantitative brain imaging, the minimum enlargement of the ventricles can be specified. Using control charts can help determine if the recruited population is meeting the criteria specified in the study design. Taking a series of scans while a patient is being treated can help to determine drug efficacy. If it is known that a mental illness affects a portion of the brain and that effect can be measured, then sampling over time can help to inform the researcher on whether the drug is working to counter the physical manifestations of the mental illness. With brain imaging and control charting, the effect would have to be neuromorphometric, which refers to measures performed on the anatomy of the brain.

Using a computer to diagnose patients has been of interest since the 1960s. An overzealous belief in the functionality of the 1960s computer coupled with the reality of then current computational abilities has made it so that the fervor which grows from reading about the possibilities dies with the realized difficulty of getting a machine to perform the functions of a doctor. Money does get seeded into the related field of computer-aided diagnostics. The current explosion of medical data can overwhelm medical personnel, but with computers the diagnostician can be facilitated by the algorithms developed for producing plausible differential diagnoses. The medical teams efforts can be used to treat possible scenarios rather than eliminate unlikely diagnoses. Computers role in medicine can facilitate diagnosticians, perform quality assurance, aide in training future physicians, and ultimately improve patient care Summers (2003).

For Magnetic Resonance Imaging (MRI), computers have been essential. Although the first MRI images had limited resolution, only a computer made the immense calculations possible. The images produced from the first MRI machines demonstrated the possibility of being able to look at the internal physiology without the use of radiation, but the images were also poor and unable to delineate the brain's anatomy. Paul C Lauterbur helped to develop the technology for two-dimensional imaging and Peter Mansfield helped to develop the mathematics to increase the speed by which an image could be acquired. Both received a Nobel prize in Physiology or

Medicine in 2003. The current state of MRI technology allows viewing the topology produced from a scan (Doi (2007)).

MRI imaging consists of stacking the two-dimensional images in order to produce the three-dimensional image. The thinner each two-dimensional image is, the greater the accuracy of the thee-dimensional image. Scanner sensitivity is increasing rapidly and eventually radiologists will not be able to scroll through the collection of two-dimensional images.

One method for getting around this problem of data explosion is to down sample the data: have the researcher look at every other planar image or every third, but this would unnecessarily decrease the amount of data available to the researcher. Computers and mathematical algorithms are an ideal alternative to down sampling because they are not constrained by the same factors as humans. As the algorithm's sensitivity to abnormalities in the scan continues to increase, the computer can direct the radiologist's attention to problem areas. This diagnostic assistance has already been in place for two decades, albeit only in mammograms.

Mammography has been using Computer-Aided Diagnostics (CAD) for years with great success (Kessler (2006)). The breast is a more homogenous tissue than the brain. We have good theories about the brain's function, but we are still baffled by it's plasticity.[2] We do not always know what causes brain cells to switch and become diseased, but sometimes the diseased tissue is nearly indistinguishable from normal healthy tissue. With lesions, researchers only know of the diseased tissue because of its location relative to other structures and its placement within a structure. Brain lesions can resemble other tissue types and can be difficult for a computer to distinguish from healthy tissue.

Lesions do not maintain their volumes over time; they can shrink and grow. This is also true of the brain and the structures of the brain. Over time, the relative proportion of gray matter and white matter changes (Tofts (2003)), but it does not

[2]Plasticity is the ability for one section of the brain to take over the functions of another ailing section.

occur too quickly in healthy brains until after age seventy (Scahill et al. (2003)). Knowing that brain structure volumes fluctuate over time lends itself to the use of quality control charts. Control charts are intended to use a time component in order to detect observations that are trending towards being out of control. A control chart should detect that the measurement of a product is trending to an out of control value before it actually does.

1.2 Software Used in MRI Studies

Dissecting the brain into areas of functionality and structure is integral to the exploration of human anatomy classes. Prior to imaging techniques, it was a destructive process - usually performed postmortem for humans, but sometimes done with live animals. Researchers also studied people that had known trauma to specific areas of the brain. Unfortunately, studying people that previously had brain trauma is not a well controlled experiment and is not repeatable. This creates an inference problem. X-rays and Computer Tomography (CT) scanners both give images of the brain, but they also irradiate the patient, which is a health hazard. As the MRI scanner became a common diagnostic tool, computers and programs were developed that could efficiently handle the administration, collection, and analysis of the data. MRI scanners do not suffer from causing known harm to a person, can be repeated, and are often used in controlled designed experiments.

Figure 1.1 is a graphic of a brain that has many of it's structural components labeled. Labeling the structural components of a brain is also known as segmenting. "Segmenting" or "Parsing" a brain was done by hand by a radiologist when MRI technology was first introduced. Eventually, technicians were trained to perform the task, but modern day scanners can produce a structural image in less than 15 minutes. Technicians could not manage that workload, so the computer and relevant software were introduced to automate the process. In fact, the low number of segmentation errors is leading to a general acceptance outside of research of the functionality of these software packages. We have most extensively worked with the

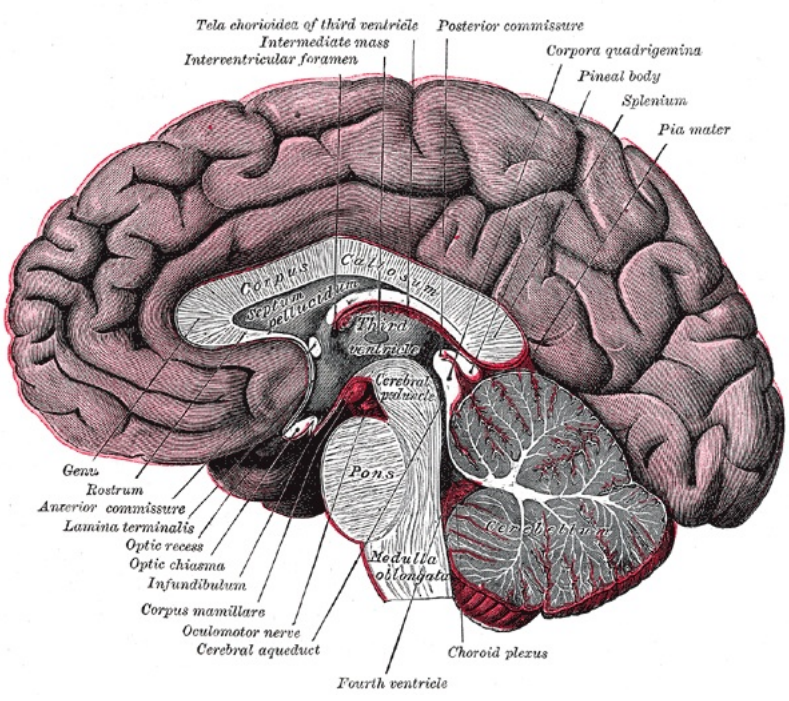

Figure 1.1: Segmented Brain

structural imaging packages *Brains2* (Magnotta et al. (2002)) and *Freesurfer* (Fischl et al. (2004)), but many other packages exist. *Freesurfer* was selected because of it's robustness, programmability for automation, and cortical thickness measurements.

A *Freesurfer* segmentation is shown in Figure 1.2. It does not look exactly like Figure 1.1 due to stripping away of the pial[3] surface. Each shaded area represents a different segmented area of the brain.

The software automatically generates a complete and detailed labeling of the entire cortex using prior information, cortical geometry, and a space-varying classification procedure (cf. Fischl et al. (2004)). Fischl et al. (2004) provided the details for section 1.2. Automating the software to label cortical gyri and sulci (folding patterns in the brain) is performed using the framework of Bayesian parameter estimation theory. **P** refers to a parcellation and **S** refers to a surface. Using Bayes' Theorem, the probability of a parcellation given an observed surface is related to the probability of a surface given a parcellation together with the prior probability of

[3]Pial refers to the outer surface of the brain

Figure 1.2: Freesurfer Rendering of a Brain

the parcellation, or:

$$p(P \mid S) \propto p(S \mid P)p(P) \tag{1.1}$$

The advantage of a Bayesian approach is the incorporation of prior information through the $p(\mathbf{P})$ term. An atlas space is a set of prior information that gives probabilities for segmentations occurring in a particular area of the brain, i.e. the frontal lobe has no chance of being classified as if it were in the back of the skull. The problem is more tractable when an atlas space is incorporated into the model, thus allowing both the priors on $\mathbf{P}$ and the probability of observing the surface given the classification to depend on the position in the cortical surface. In order for this process to work across subjects, a function $f(\mathbf{r})$ must be built which uses the raw image coordinate and maps it to the corresponding coordinate in the atlas. Subjects are therefore mapped to one atlas which allows comparability across subjects. The joint probability of the mapping function f and the parcellation $\mathbf{P}$, are given by:

$$p(P, f \mid S) \propto p(S \mid P, f)p(P \mid f)p(f) \tag{1.2}$$

The first term contains information on the class statistics which vary as a function of location, the second term expresses prior information regarding spatial structure of the parcellated labels, and the third term constrains the space of allowable atlas functions. The *Freesurfer* team used both a manual parcellation of subject brains along with a probabilistic brain created from the subjects to build the atlas. If a single subject had been used as the atlas, then the idiosyncrasies of the subject would appear in the atlas. Using an entirely probabilistic mapping may cause certain anatomical information to be lost. The team therefore went with both manually segmented brains and a statistically processed brain for obtaining the prior probabilities for a structure's location. The atlas that was generated provides two forms of prior information: the first gives the probability of a parcellation label occuring at a location in the atlas which is provided by f and the atlas; the second is an indicator for possible locations of parcellation labels. The second bit of information eliminates the possibility of extremely erroneous labels such as 'the right thalamus is on the left side of the brain.'

Given f, the atlas function, and $\mathbf{N}$ manually parcellated subjects, an estimate of the occurrence of $\mathbf{c}$, the prior probability of parcellation when independent of other locations is:

$$\Big(P(r) = c\Big) = \left(\frac{\text{number of times class c occurred at location} f(\mathbf{r})}{\text{number of times that map to } \mathbf{r} \text{ in the training set}}\right) \tag{1.3}$$

This equation finds the agreement between the manual parcellation of the brain with the statistical priors used to parcellate the brain. Two quantities representing the observed surface $\mathbf{S}$ are encoded as a vector $\mathbf{G}(\mathbf{r})$ at each point in the surface $\mathbf{S}$. ($\mathbf{r}$ refers to the spherical atlas coordinates, while $f(\mathbf{r})$ refers to the subject's spherical coordinates.) The two vector values are average convexity and mean curvature, but other quantities such as $\mathbf{T_1}$ and $\mathbf{T_2}$, Gaussian curvature, or cortical thickness can be incorporated. The likelihood of observing the surface geometry $\mathbf{G}$ given the parcellation label $\mathbf{P}(\mathbf{r})$ is a Gaussian model with parameters:

$$\mu_c(r) = \frac{1}{M}\sum_{i=1}^{M} G_i[f(r)] \tag{1.4}$$

that give the likelihood of observing the surface geometry **G** given the parcellation label **P**(**r**). A set of **M** surfaces occurs for label **c** at location *f*(**r**) in the manually labeled surface models $\mathbf{S_i}$. The $\mathbf{G_i}$ are the geometric parameters extracted from the set of M surfaces. Class **c** at location **r** has covariance matrix:

$$\Sigma_c(r) = \frac{1}{M-1}\sum_{i=1}^{M}\Big[G_i[f(r)] - \mu_c(r)\Big]\Big[G_i[f(r)] - \mu_c(r)\Big]^T \tag{1.5}$$

Geometric information across classes needs to be averaged because the information about the surface geometry is maintained separately for each parcellation label at every atlas location. An estimate of pairwise probabilities that two parcellation labels $\mathbf{c_1}$ and $\mathbf{c_2}$ occur when the labels are at locations $\mathbf{r_i}$ and **r**, respectively, with $\mathbf{r_i} \in \mathbf{N}(\mathbf{r})$. is expressed mathematically,

(1.6)

$$p[P(r) = c_1 \mid P(r) = c_2, r_i] = \frac{\#\text{ of times } c_2 \text{ occured at location } r_i \text{ when } c_1 \text{ occurred at } r}{\#\text{ of times class } c_1 \text{ occurred at location } r}$$

This information is stored separately for each atlas location. The probabilities are stored separately for each pair of classes and for each neighborhood location **r**, but due to the sparseness of the matrix, the computations remain tractable.

Freesurfer computes the maximum *a posteriori* (MAP) estimate of the parcellation **P** given the input surface geometry **G** and the nonlinear spherical transform *f*. Bayes' Thereom gives:

$$p(P \mid G, f) \propto p(G \mid P, f)p(P) \tag{1.7}$$

which can be re-written if the noise at each vertex is assumed to be independent from other vertices in the model as:

$$p(G \mid P, f) = \prod_{r \in S} p\Big[G[f(r)] \mid P(r)\Big] \tag{1.8}$$

The geometric properties of each class at each atlas location are distributed as a Gaussian with mean vector $\mu_{\mathbf{c}}(\mathbf{r})$ and covariance matrix $\mathbf{\Sigma_c}(\mathbf{r})$. The probability of $\mathbf{G}[f(\mathbf{r})]$ is expressed as:

(1.9)

$$p\Big[G[f(r)] \mid P(r) = c\Big] = \frac{1}{|\Sigma_c(r)|^{1/2}\sqrt{2\pi}} exp\Big[-.5\Big(G[f(r)] - \mu_c(r)\Big)\Big]^T \Sigma_c r^{-1}\Big[G[f(r)] - \mu_c(r)\Big]$$

Freesurfer assumes the prior probability of each parcellation $\mathbf{P}$ by assuming the spatial distribution of labels are well approximated by an anisotropic non-stationary Markov random field. Non-stationarity allows the prior information about the relationship between labels to be expressed as a function of location while anisotropy allows for the expression of direction. In essence, the location and direction are important and not interchangeable. The Markov assumption is written as:

$$p[P(r) \mid P(S - r)] = p\Big[P(r) \mid P(r_1), P(r_2), \ldots, P(r_k)\Big], \text{ where } r_i \in N(r) \tag{1.10}$$

Equation 1.10 means that the prior probability of a label at a vertex $\mathbf{r}$ is influenced only by it's neighbors. The log-likelihood function gives:

$$p(P) = \prod_{r \in S} p[P(r) \mid P(r_1), P(r_2), \ldots, P(r_k)], \text{ where } r_i \in N(r) \tag{1.11}$$

and Bayes' Thereom transforms the equation into:

$$p(P) \propto \prod_{r \in S} p[P(r_1), P(r_2), \ldots, P(r_k)) \mid P(r)], \text{ where } r_i \in N(r) \tag{1.12}$$

In order to make this a tractable problem, the first order conditional dependence is used. That is, only the local neighbors are used and the equation is rewritten as:

$$p\Big[P(r) \mid P(r_1), P(r_2), \ldots, P(r_k)\Big] = \prod_{r \in N(r)} p[P(r) \mid P(r_i), r_i] \tag{1.13}$$

The maintenance of r_i emphasizes that the probability densities are maintained separately for each neighbor position in $N(r)$. The full parcellation of the brain's prior probabilities can be expressed as:

$$p(P) \propto \prod_{r \in S} p[P(r)] \prod_{i=1}^{K} p[P(r) \mid P(r_i), r_i] \tag{1.14}$$

This technology enables researchers to automatically parcellate MRI scans so that limited research resources can be spent elsewhere. Some errors do occur in the processing and minor post processing is needed, but the advantages of a nearly perfect inter rater agreement, scalability, cost, and efficiency far outweigh the disadvantages.

1.3 Control Charts and Quality Assurance Software

Once the data were measured by *Freesurfer* it was placed in text files. The text files were aggregated using custom PERL scripts. The SAS language was used to create control charts. PROC SHEWHART was used for the three different control charts

developed for these data: a whole volume chart, a percent of total volume chart, and a regression control chart. More details of what each of these charts represent is given in Chapter 5.2, but briefly: a *Whole Volume Chart* is the raw measure of the brain structures, a *Percent of Total Volume Chart* is the proportional amount of volume that a structure in the brain occupies relative to the entire brain volume, and a *Regression Control Chart* uses the residuals after regressing the dependent structure on the covariates. There were 29 structures which were measured and validated, but the scale of their measurements are not all the same. Some structure's measurements differ by 1, 2, 3 or more scales of magnitude. PROC UNIVARIATE was used to find out the number of significant digits present in the mean of the structure's measure for scaling the axes. This was encoded in a macro variable, providing PROC SHEWHART the number of decimal places to use in chart construction. Macros eliminated much of the tedium of data preparation. The functionality of the macros that were created were: build and name the control charts with appropriately scaled metrics, regress structural variables on a single structure, and regress a structure with behavioral data. These are fairly powerful programs because they can be re-used, but added into this mix is some less-portable code which took the resulting data sets and generated tables of outliers. The development of this code took several months of work and even code execution can run up to 2 hours because of the number of charts and regressions performed. It has been made as efficient as possible and for a single *Whole Volume Chart* takes less than a second to run.

Scanning several subjects several times over multiple years is the ideal test of control chart usefulness in detecting changes in morphology. However, this is also the most impractical because the inevitable changes in brain composition can take years to occur in healthy subjects. MRI scanning is extremely costly (discussed in Chapter 5.3) and so are longitudinal studies. Scanners are very profitable multi-million dollar pieces of equipment and are therefore upgraded as frequently as possible. In general, study attrition rates depend on the study complexity, subject motivation, time commitment, and many other factors. All attrition rate factors are tested with MRI scans and so a large starting sample is essential. The cost and the time com-

ponent are not only the key pieces to a study on using control charts to assess brain morphology, they are also the barriers.

Our data set has seventy-one healthy subjects scanned once in the same year on the same scanner. This removes the inter-scanner differences and most of the variability from changes in a scanner. The study design also eliminates the ability to show that a single subject can be used on a control chart to detect when their brain scans become abnormal. The study demonstrates the feasibility of using control charts with morphometrics and encourages the exploration of other methods for detection of abnormal scans.

Chapter 2

Quality Control Charts

2.1 Introduction

Dr. Walter A. Shewhart (1891-1967) published a memo in the 1920's that lead eventually to his publication of the first book on statistical quality control (Shewhart (1931)). His ideas were revolutionary and have since proven to be instrumental in the technological developments of this last century. Statistical quality control ideas are used in manufacturing, medicine, computers, programming, and many other fields. He also influenced Dr. W. Edwards Deming (1900-1993), a man instrumental in process improvement of manufactured items during World War II and in post-war Japan.

There are a number of control charts used in industry and they are seeping into medical practices as well. Researchers are finding methods to monitor the quality of healthcare delivery. Recently, Quality Engineering (Volume 20, Issue 4, 2008) did a special issue dedicated to quality control and healthcare. The area of quality control charts in medicine has a set of challenges not found in more traditional industrial technologies. In the medical field data are finicky and sensitive to violations of assumptions. Astronomically increasing costs of medicine are focusing funding on both the problems and solutions that may be the drivers of costs. We review a few

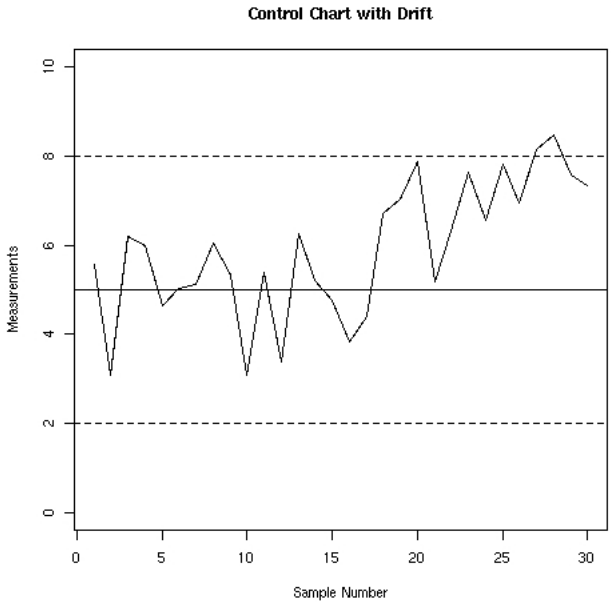

Figure 2.1: Sample Control Chart

of the more commonly used control charts, including: $\bar{X}$-charts, R-charts, S-charts, attributes charts, multivariate charts, and Individuals charts. Montgomery (2005) and Ryan and Ryan (2000) both offer excellent introductory quality control chart textbooks.

2.2 Quality Control Charts

A sample control chart is given in Figure 2.1. The key feature of a control chart is the time component which allows for detecting a subtle trend in a process. Figure 2.1 has had drift introduced at point 16, but depending on the rules used by the engineer, the drift may not be caught until point 25.[1] Sudden or dramatic changes in the process can be caught just as easily with an outlier analysis, but by the time a production process is that far out of range, several items will have likely been produced that were outside specifications.

[1] 8 samples in a row above center line

The ideal way to catch out of control points is to test everything frequently, but this is not usually economical. Over testing displaces the equity maintained between production costs and profitability. The manner in which a process can be tested determines how frequently it will be tested. Destructive, costly, and slow production rates all require infrequent quality testing. Automatic, inexpensive, and fast production rates are areas where frequent quality testing can be performed. The control charts used in each of these situations depends on what methods are available to the engineers.

An $\bar{X}$-chart averages the measurements made on a small sample (subgroup) of manufactured product and assigns that value to the lot that was produced. It is assumed the small sample is representative of the lot. Over time each of these samples are plotted and a lot by lot time series is produced. The time series would look similar to Figure 2.1. $\bar{X}$-charts indicate if the mean is in control, but the variability is not monitored and could be changing.

If the sample sizes are between 2 and 10 units, an R-chart is used in conjunction with the $\bar{X}$-chart. R-charts use the range of values in order to estimate the variability in the data. The range of a sample from a normal distribution is related to the standard deviation by $W = R/\sigma$ where W depends on the sample size n. The mean of W is d_2, a quantity that can be found in a text such as Montgomery (2005), which gives an estimate $\hat{\sigma} = R/d_2$.

If the sample size is greater than 10, R becomes a less accurate measure of the standard deviation. An S chart is used instead where the standard deviation is computed from the sample. Historically, prior to the availability of computers, R-charts were popular because of their ease of calculation. S-charts are always a better estimate of variability.

An attribute chart can be useful for MRI data. Attribute charts find the proportion of acceptable product produced in each lot and therefore a binomial distribution can model the observations. They are often used when items are assembled and produced or when it is difficult or too costly to develop an objective metric for the

system. For example, a company manufacturing dolls could set up control charts for whether the dolls had two eyes, two hands, two feet, etc. However, any one of those defects means the doll needs to be scrapped or reworked and so using an attributes chart makes more sense than modeling every component with it's own control chart. If the probability of an event is too small (or too large), the binomial distribution's control limits start misbehaving and taking on values outside the range possible for the binomial. Essentially, this means if we are using an attributes chart and the defect rate is low, the chart is not useful because it can not detect shifts or trends that may lead to a process being out of control. An attribute chart does not consider what components failed and lead to rejection. If we wanted to know how many defects occurred in a product, we might consider a Poisson distribution, which is a good approximation for phenomena that occur on a per unit basis. A right-skewed Poisson distribution corresponds to a low defect rate.

A multivariate chart looks at several attributes at the same time and considers those attributes as a whole to determine if a product or batch is to be rejected. If we constructed a box, a multivariate chart could be used to ascertain if the height, width, and depth of the box were within pre-specified limits. All three attributes could be within specifications, but the sum total of the error of the three could add up so the box is rejected. A multivariate chart is able to detect the case in which the sum of the parts produces a defective product, but the individual parts pass inspection.

For these data, Individuals Charts will be used. Individuals charts lend themselves to situations in which every product can be sampled inexpensively and non-destructively. Weight measurements performed daily is an example data set that would use an Individuals Chart. One sample weight taken at the same time daily would be a useful metric to determine weight stability over time. It is an inexpensive and quick method for assessing weight control. Relative to the cost of an MRI study and all of the processing that will already be put into assessing the scans, the creation of control charts on morphological data is also inexpensive. The Individuals Chart will be used on the raw data, a normalization of the data, and the residuals

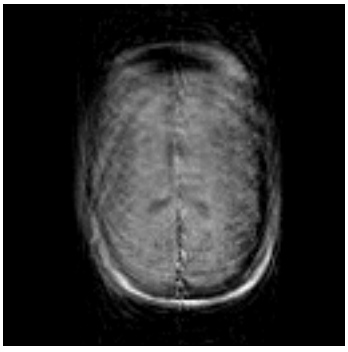

Figure 2.2: Extreme case of Motion in an MRI scan

after regressing the data on covariates.

We will assess healthy normal subjects 18-25 years of age which have had their structural MRI brain scans regionally segmented and measured by the *Freesurfer* software. Control charts allow for a quantitative evaluation of subject scans. Assessing brains with control charts has to-date not been published in the quality control or radiology literature. Individuals Charts were used for the example analysis in section 2.1, but other charts can be used under different sampling schemes.

The MRI scanner is our measurement tool and is expected to be in control because of quality assurance performed by the MRI technicians. The production cycle of a brain is somewhat obvious: conception, pregnancy, birth, and life up to the point of scanning. Events that may affect brain development include: trauma, disease, mental illness, alcohol and drug use, or physical abuse. A nurture argument may also include events like happiness, encouragement, or verbal abuse as having a significant affect on the production process (Johnson et al. (2002)) Knowledge of events such as trauma, disease, or mental illness can help the person examining quality control charts on brain images assign cause to out-of-control signals and can also be used in regression control charts. Unlike in industry, the subject can not be "scrapped" or re-worked, but the subject can be excluded from future analyses based on their morphometric profile. The rigor requisite for classifying a subject based solely on morphometrics is not present and social factors should not be ignored (Johnson et al. (2002)).

The other measurement the control charts will be considering is the quality of the

scan. Most software packages will fail if a brain scan has too much motion artifact. If there is little motion artifact, the segmentation can move forward, but erroneous measures are likely to occur. The control charts should be able to flag those errors.

2.3 Individuals Charts

The paradigm being used is that each brain is identically and independently sampled. The subjects and their brains were randomly selected. Each brain has undergone a "manufacturing process" about which we know little and are unable to control. The only option available to the researcher is to reject a person from the study. Rejecting a data point without assignable cause is acceptable, but has no analytical justification (Montgomery (2005)).

The Individuals Chart is considered to be a Shewhart Chart because it falls under the paradigm developed by Shewhart. A control chart can accomplish three things (Duncan (1965)):

1. It defines the goal or standard for a process that management might strive to attain.

2. It is used as an instrument of attaining that goal.

3. It serves as a means of judging whether the goal has been reached.

Our purpose is to define a goal or standard. The data we have are being used to estimate the control limits. This means that the control charts are in Phase I. Phase II would mean that prior limits have already been established and are set for a future data set (Montgomery (2005)).

Montgomery (2005) gives five reasons for using an Individuals control chart:

1. Automated inspection and measurement technology is used, and every unit manufactured is analyzed so there is no basis for rational subgrouping.

2. The production rate is very slow, and it is inconvenient to allow sample sizes of $n > 1$ to accumulate before analysis

3. Repeat measurements on the process differ only because of laboratory or analysis error

4. Multiple measurements are taken on the same unit of product.

5. Measurements on some parameter will differ very little and produce a standard deviation that is too small to be significant.

Reasons 1, 2, and 3 apply to quality control charts for brain measures.

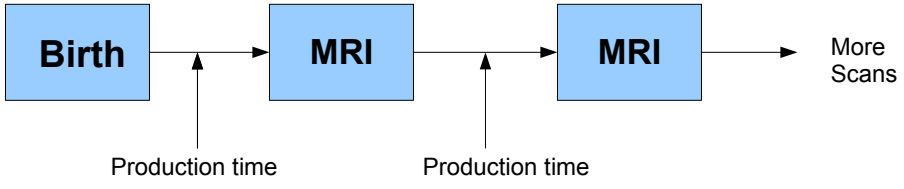

Figure 2.3: Diagram of production time between MRI scans

The segmentation of the brain is automated as was noted in chapter 1.2. Using the *Freesurfer* software it can take up to 34 hours to segment a brain and a few more minutes to perform measurements. The 34 hours is analogous to weighing an object

on a scale; it simply takes a long time to "weigh" an MRI scan. Production time would be the time it takes to scan the brain again. Steen et al. (2007) re-scanned his subjects after 12 weeks which gives a production rate of once every 12 weeks. The authors were looking for changes in morphology after a short period of time which they did not find in subjects that had not suffered a serious event affecting their brains (such as stroke, vehicular accident, or other catastrophe). Since Steen et al. (2007) did not see appreciable changes in morphology after 12 weeks, it is unlikely any would be seen after a few minutes. The few minutes corresponds to the time it would take to rescan a subject. The algorithms used in the *Freesurfer* software are somewhat robust and the process will differ very little. *Freesurfer* will produce a segmented image that is then measured by region.

Our primary interest is in three preliminary research questions.

1. Whether or not a subject is structurally within normal limits.
2. The quality of structural scans.
3. The image analysis pipeline's performance in correct brain parcellation.

Although the three questions are inter-related, each is considered separately.

For this study, the subject was entered into the study based on the profile provided by a battery of psychological tests. In other studies, a researcher will try to use a set of healthy controls for comparison against mentally ill subjects. The number of psychological tests and clinical interviews necessary to rule out all diseases and declare a subject as a healthy normal control is prohibitive; therefore a subset of tests is used. Declaring someone healthy when they are not is a type II error. The control chart helps to mitigate the effects of type II errors by determining whether the subject's morphology based on the MRI is also considered to be within normal limits.

Quantitatively valid scans are currently not a clinical requirement (Tofts (2003)). However, in the future, comparability of scans may be essential which would require

quantitative validity (Tofts (2003)). Our scans meet that requirement and allow us to objectively state whether a region is structurally within normal limits. That is, the subject is within the control limits set by the study data. This is Phase I because we do not use reliable *a priori* limits.

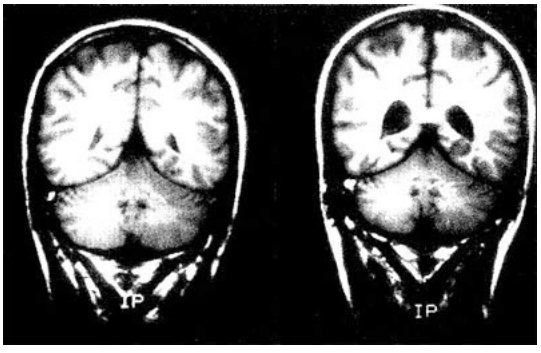

Figure 2.4: The healthy brain is on the left and the schizophrenic brain is on the right. Courtesy of kevinturnquist.org/images/housing21.jpg

Statistically significant structural volume differences between subject populations exist, for example between a non-schizophrenic subject's and a schizophrenic subject's regional brain volumes.

Differences are found between brain structures of subjects with disease versus those considered normal (Miller et al. (2002)).

Although there are a number of research studies that involve "brain", "volume", and "magnetic resonance" in the medical literature (Worth et al. (2001); Tofts (2003)), currently none in statistical quality assurance literature have addressed this area. In recent years, quality control techniques are finding applications in the healthcare industry (cf. Woodall (2006); Manos et al. (2006); Benneyan et al. (2003)). Studies are able to discern volume differences between healthy controls and subjects, but their assessment of health status is based on psychometrics rather than a fusion of psychometrics and morphometrics (cf. Castellanos et al. (2001); Grachev et al. (1998); Plessen et al. (2006); Strakowski et al. (1999); Szeszko et al. (1999)).

The second question considers the quality of the structural scans. Jovicich et al. (2006) informs us that many different artifacts can affect image quality such as oper-

ator skill level, subject motion, scanner inhomogeneity, magnetic field nonuniformity, and so on. For example, when placing a subject in the scanner, the static magnetic field is adjusted and becomes inhomogeneous. Several saturation scans are performed to reduce the inhomogeneity; however, the effects of air pockets (e.g. sinuses), cannot be eliminated in full.

The precision of a subject's volume measures can change with cardiac and breathing patterns (Tofts (2003)). Motion artifacts can cause severe distortions, but as scanners improve and better head restraint methods are used, the motion artifact becomes less of an issue. We recommend that quality assurance (QA) procedures immediately follow a structural scan to minimize the loss of collected data due to poor image quality, thus allowing for a repeat scan.

Mitigating the rejection of data due to bad scans will increase the potential number of desirable subjects in a study and decrease variability within subject type. If QA procedures are implemented, we would know (1) whether the subject is structurally within normal limits, i.e. should the subject keep the same classification as assigned upon entering the study, and (2) whether the scan is poor. The results of the quantitative analysis would not change a subject from the status "healthy normal control" to being in the comparison group based on brain volumes. The subject may be removed from later analyses due to structural measures that are outside of normal limits. If the scan is poor, the registration of images becomes more tedious, region of interest analyses cannot be performed, or the data may need to be deleted from the analysis.

The third question concerns the performance of software on the images. Note it is possible the software is unable to process certain brains; a scenario which would alert the technician to a bad scan. An example scan which would be impossible to process due to the extreme motion artifact was given in Figure 2.2. This error, when caught during a scan, makes the final data cleaner and more accessible.

Chapter 3

Magnetic Resonance Imaging and Quality

In the information technology age, statistics and computer science are making headway in reducing and summarizing the vast amounts of data procured from medical imaging modalities for medical personnel to digest. Structural MRI (sMR) data is a large interdisciplinary medical data set that benefits from statistical summarization, but has yet to be quality assured with control charts. Incidental findings, such as a neoplasm, have been found in asymptomatic volunteers (cf. Katzman et al. (1999); Vernooij et al. (2007)). A person can be asymptomatic for a mental illness, but the physiologically correlated signs of mental illness might be present. Our method helps flag an abnormal scan quickly so that appropriate medical personnel can be alerted. We hope for this to be a first step towards automation and routine screenings in clinical MRI such as already occurs in breast cancer screening (Garvican and Field (2001)).

3.1 Physics of Magnetic Resonance Imaging

3.1.1 Introduction

Magnetic resonance imaging (MRI) for the human body is a recent development in medicine which has overcome many of the pitfalls of computed tomography and x-ray scans. Magnetic resonance gives well-contrasted images of soft tissue, which is something neither X-rays nor CT scans do without the use of contrast agents. MRI does not need radioactive substances in order to produce a structural image delineating tissue types because differing tissue types have their own characteristic resonance. This makes it ideal for use in at risk populations such as pregnant women. The aim of this section is to introduce a brief history of the development of MRI.

3.1.2 Magnetic Resonance Imaging Physics

The history of developing MRI technology is pulled mostly from Mattson and Simon (1996) while the physics of MRI technology is pulled from Hashemi et al. (1997) and Tofts (2003). The scientific foundation necessary for brain imaging began at the turn of the century with Max Planck proposing quantized energy packets for electromagnetic radiation. America was slow to come to terms with quantum mechanics, but with several European physicists immigrating to the U.S., academic research in the "new science" blossomed. World War II provided the necessary impetus for America to commit extensive resources into exploring the atomic level of matter from which would come an atomic bomb, the lesser known, but now ubiquitous, science of radar technology, and the Diagnostic and Statistical Manual of Mental Disorders (DSM). All three technologies have a role in this thesis. Radar came to the US from Britain and at the time, the researchers were using a 10 cm wavelength. Developing the technology to a 5 cm and 2.5 cm wavelength improved the ability of the radar to detect aircraft. The 1.25 cm radar was an abysmal failure, especially on humid or cloudy days. Water absorbs energy around 1.33 cm which was close enough to absorb much of the signal. The reasons behind the loss of signal presented another opportunity

for research after the war.

MRI works with the magnetism of protons. Spinning charged particles create electromagnetic fields causing the nucleus to act like a bar magnet. Placing a bar magnet into an electro-magnetic field will orient the magnet, but at the atomic level, the field must be maintained or the atoms will re-orient to their preferred space. The randomization of the atoms after the magnetic field has been turned off was explained by equations developed by Dr. Felix Bloch. Later he would publish the theory behind solid state nuclear magnetic resonance, thereby earning a Nobel Prize in physics in 1952. Edward Purcell and his team initially discovered solid-state resonance for which Purcell also received a share of the 1952 Nobel Prize in physics.

MRI scans predominantly use the hydrogen nucleus, though other elements with an odd number of protons can be used. Hydrogen is particularly useful in that it exists in only two states, known as spin (S), and is very abundant in the body. The spin of a proton relates to the number of energy states a proton can be in.

Energy states = 2*S+1

Hydrogen's proton has a spin of $\frac{1}{2}$ and therefore has two energy states, one energy state being higher than the other.

When spinning, unpaired protons are placed in a magnetic field, some will line up parallel to the field and some antiparallel. Slightly more, a million to a million minus one, will be in the lower energy state of being parallel to the field.

The sheer number of protons in the body enables such a small difference to be very significant. The rate at which the protons align themselves follows a growth curve similar to Figure 3.2.

All electromagnetic waves travel at $c = 2.99792458 * 10^8$ meters/second and are composed of an electric wave and a magnetic wave. Both waves are sinusoidal, out of phase by 90 degrees, and propagate each other. The magnetic field is what is used in magnetic resonance while the electric field produces heat, an unwanted side effect.

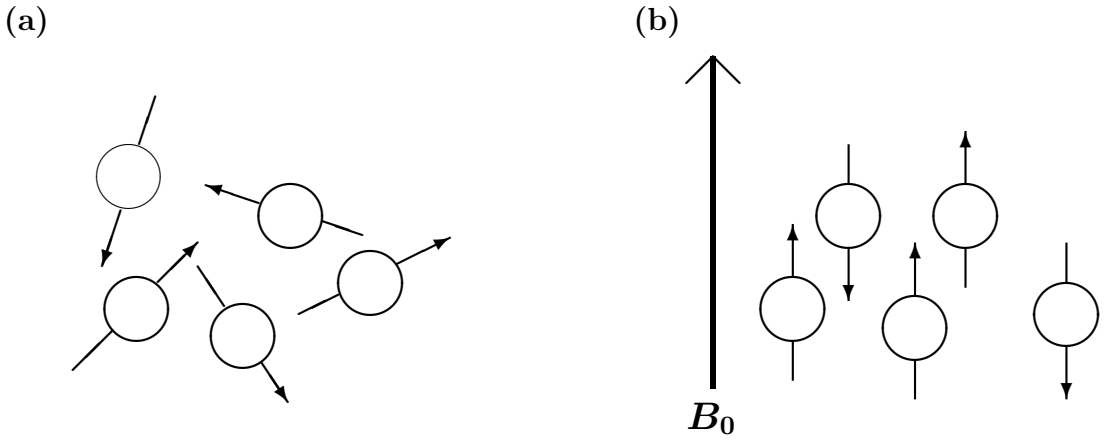

Figure 3.1: a) Protons, the circles, are arranged randomly with no external magnetic field applied b) With the external field (B_0) applied, the protons align themselves with or against the external field

In order to perturb a proton, energy must be added to the system, and, in order to be absorbed and induce a flip, it must be added at the Larmor frequency which is in the range of the radio frequencies[1] and is the rate at which protons are precessing

[1]Radio frequencies occur between 3-100 MHz (megahertz).

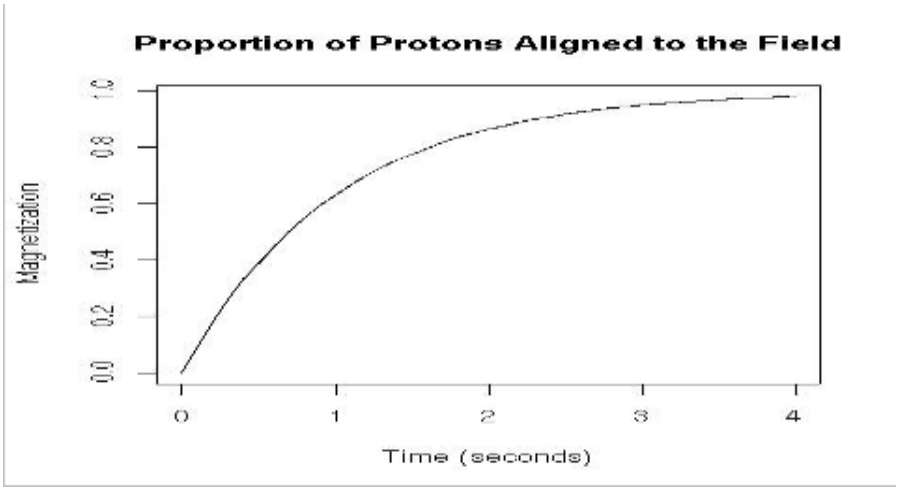

Figure 3.2: The growth curve of the number of molecules aligning with the external magnetic field over time

Source	Frequency
AM radio	.54-1.6 MHz
TV	64 MHz
FM radio	88.8-108.8 MHz
MRI pulse (RF)	2-100 MHz

Table 3.1: Frequencies of use for common devices

about the main magnetic field. Table 3.1.2 below gives the various radio frequencies we use everyday.

The magnetization is described by $M = 1 - e^{-\frac{t}{T}}$. For tissue, the rate of net magnetization is directly proportional to the strength of the external field. Note that the above equation in no way accounts for tissue type. Tissue type must affect the rate of magnetization, otherwise brain scans would look uniform with no contrast between structures.

The reason that field strength affects the T_1 relaxation is believed to be due to the particulates in the tissue. An accepted but unproven explanation is that particulates modify sub-atomic forces. Therefore it is believed that the particulates affect the relaxation because they are not made up of protons that are as mobile. The density of mobile protons within the tissue affects net magnetization and so the above equation should be $M = N(H)(1 - e^{-\frac{t}{T_1}})$, where N(H) represents the density of mobile protons.

To induce a signal from the tissue, the scanner must input another weaker radio frequency pulse at an angle to the main magnetic field. This new magnetic pulse must also be at the Larmor frequency for the energy to be efficiently absorbed. A 90° pulse will cause the protons to flip into the transverse plain, creating a transverse magnetization. The pulse will also cause some of the protons to absorb energy and boost themselves into the higher energy state causing a decrease in net magnetization. As the magnetization decays, a signal (energy) is released from the system and is detected by coils within the scanner.

The rate at which the protons begin realigning themselves after the weak radio frequency pulse is turned off, is determined by the T_1 constant. The equation is $M_z(t) = M_0(1 - e^{-\frac{t}{T_1}})$. T_2 is the decay of the signal in the transverse plane. $T_2{}^*$ is used instead to account for external field inhomogeneities. The overall signal intensity that a receiver coil detects follows $S = M_0(1 - e^{-\frac{TR}{T_1}})(e^{-\frac{TR}{T_2*}})$

Scanners are unable to measure the signal immediately after the radio frequency pulse has been added to the system. The gradient coils, also known as receivers, wait a small interval of time before they acquire the signal. This is the Echo Time or TE.

In order to acquire a three-dimensional image, a radio frequency pulse must be repeated for different planes of the brain. This repetition is known as a TR (repetition time). Each repetition flips the magnetization vector back into the transverse plane. As the proton recovers, it will give off a signal. The signal does not have dimension.

The signal received is made up of the T_1 and the T_2 signals. As the TR increases, the weighting due to the T_1 decreases. As the TE decreases, the weighting due to the T_2 decreases.

3.1.3 Imaging and Contrast

T_1 and T_2 images emphasize tissues differently and each is used to detect different things. The cerebral spinal fluid is very dark in a T_2 image while it is bright in a T_1 image. This is expected because the images are inversely proportional to each other.

The T_2 image is affected by external magnetic field inhomogeneities and by spin-spin interaction. Water is not usually isolated in a human and therefore spin-spin[2] interactions cause the T_2 relaxation time to be long. Solids are compact and the dephasing caused by spin-spin interactions is prevalent. Fat has a relaxation time between that of solids and water.

The T_2 properties are related to the efficiency of the system to absorb or give

[2]This is the interaction between hydrogen protons

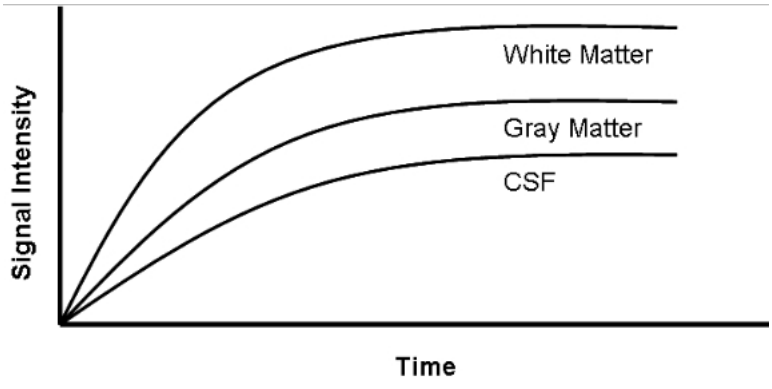

Figure 3.3: Relative relaxation times for CSF, White matter, and Grey Matter

off energy to the surroundings. If the translation, rotation, or vibrations are at the Larmor frequency, the system absorbs or releases energy at the most efficient rate. Figure 4.1 shows the signal intensity of Cerebral Spinal Fluid (CSF), White Matter, and Gray Matter changes with respect to time.

Chapter 4

Application

4.1 Introduction

Neuropsychology examines the relationship between the structure of the brain and psychological manifestations. It is a field that has rapidly expanded over the last three decades, and now is so replete with research that no one could possibly hope to stay abreast of all current practices. Most of our background information was contributed by Lezak (1995). Similar to the science for MRI and the statistics of quality control, neuropsychology found it's beginnings in the early 20th century with a rapid increase in research during the 1970's. The mental and emotional manifestations of physical brain damage are not well understood, except to say there are generalizations which neuropsychological literature reports. Currently research is being conducted in developing tools that would allow the MRI scanner to be used as a diagnostic tool for mental health. This is one reason why comparisons are performed between healthy normal controls and subjects with mental health problems. In using control charts to reject certain subjects from a study based on volume measurements, we are accepting that physical brain abnormalities do inherently result in psychological abnormalities.

4.2 Neuropsychology

Neuropsychologists are involved in diagnosis, patient care and planning of treatment, and rehabilitation and treatment evaluation.

Diagnosis: Neuropsychology distinguishes between psychiatric and neurological symptoms; neuropsychologists can find neurological disorders with no psychiatric presentations, distinguish something as being a purely psychiatric manifestation with no neurological deficit or vice versa, or localize the sight of a neurological deficiency using behavioral data. Though advances in *in vivo* techniques (MRI, CT) may help with diagnostics, neuropsychological techniques will still be necessary in diseases like toxic encephalopathies because the quantitative equipment is not and will not be sensitive enough.

The other arms of diagnosis are prediction and screening. Coma patients should be assessed for prediction of outcomes using CT and MRI (van der Naalt (2001)). This would allow for more accurate predictions in outome expectations. Blacker et al. (2007) found that screenings could predict future cognitive decline.

Patient care and planning of treatment: A neuropsychologist is often used in assessing patients after a medical event which can affect their cognition. The mental requirements of a 55 year old stroke patient who is the CEO of a company differ from those of a 27 year old construction worker involved in a vehicular accident. It may be true that the CEO has to percept more than the construction worker, but if the construction worker slips or drops equipment, an immediate danger is imposed on the lives of others. The expectations of a patient affects the determination made by the neuropsychologist.

Rehabilitation and treatment evaluation: A neuropsychologist may only give recommendations on what needs improvement based on tests (MRI, CT, or other paper based exams), but will not administer the tests. For treatment evaluation, multiple tests over a period of time would be formed and evaluated. In rare cases, a baseline assessment might have been performed before the brain assault occured (perhaps in

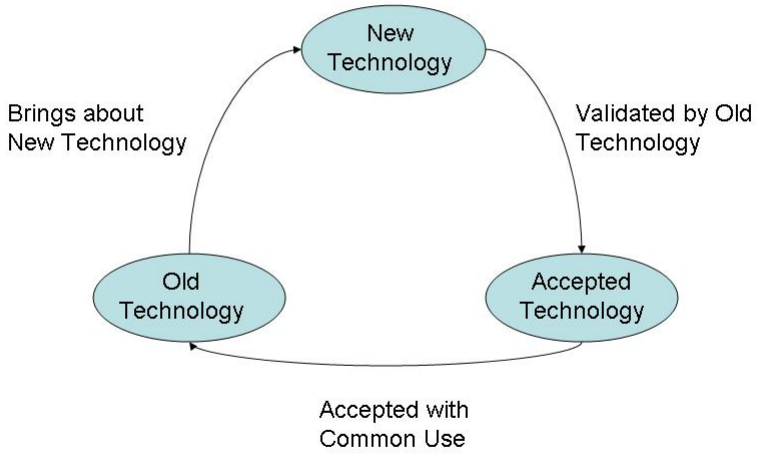

Figure 4.1: A concept of how technology matures

a screening).

Research: Research into new testing methods and combining other testing methods is a part of the neuropsychological field. Research allows for the validation of old methods and the impetus to explore new ones. In 1983, MRI machines became more prevalent in clinical medicine (Mattson and Simon (1996)). With the advent of a new assessment technology came the need to validate the technology from the neuropsychological perspective. Once MRI had been validated from the viewpoint of neurologists, it was time to begin validating neurological exams from the perspective of MRI. Once new technologies such as *functional* MRI are now in the process of validating other theories.

The use of quality control charts will hopefully follow a similar pattern. The technology will be added to the structural measurements to assess morphometrics. When the control charts have been validated as a method of assessment, then they can be used in screenings by neuropsychologists. It would be useful if a neuropsychologist

could accurately inform a patient that based on previous MRI measures, they can expect a certain physiological outcome. The neuropsychologist may be capable of discussing preventive measures as well.

4.3 Neuropsychology and Modeling

A neuropsychologist attempts to incorporate all the known data into a tapestry which will give potential future outcomes and eliminate unlikely scenarios. They build a model based on prior models that they have developed or researched. They have before them the predictor variables, behavioral data from psychological tests and MRI/CT scans, and are asked to determine future outcomes or proper treatment. The patient's condition is not static. As the neuropsychologist gets to know the patient better, the neuropsychologist gets a picture of the patient's variability. Adding quantitatively valid scans to the repertoire of tools could help the neuropsychologist assess potential outcomes earlier. A quantitatively valid scan is similar to the other tools diagnosticians in other fields use to assess a patient's health. When a patient sees a family practice doctor for high cholesterol, the doctor knows that the patient is at risk for coronary artery disease, stroke, or blood clots. Diagnosticians have even started to classify the risks associated with different cholesterol levels. Quantitive scanning attempts to allow the neuropsychologist the same insight into a patient's future risk. Quantitative scanning over time on a patient should also allow a neuropsychologist to chart disease progression. If an older patient arrives with enlarged ventricles and those ventricles slowly increase over time, quantitative assessments will detect the increase. The brain's morphology is not static. This is another reason why control charts are good for patient care. The control chart allows for the quantitative variability to be assessed as being normal over time or out of control.

Chapter 5

Study Design and Implementation

The study was developed at the MIND Institute at the University of New Mexico. The principal investigator, Dr. Rex Jung (PI), is a trained psychologist. The primary goal of the study was to investigate the contribution of white matter integrity to broad measures of cognition and personality in normal subjects, but this thesis is using only the sMR data. All scans were done on a 1.5 Tesla scanner at the Mind Research Network. All imaging sequences used standard imaging parameters, well established in clinical and research settings. This study was University of New Mexico Institutional Review Board approved and met all necessary standards for human protections. One protection is radiological review, results of which are compared against the control charts.

- **Subjects and Screening:**

  Subjects were recruited from the UNM campus via informational postings. Only subjects in the age range of 18-25 were considered in order to minimize age effects on brain development. To ensure that the subjects were in good physical and mental health, an interview and physical history was conducted by a licensed psychologist. Exclusion criteria included: 1) pregnancy, 2) presence of a potentially dangerous metallic device, implanted or otherwise, 3) age less than 18 or greater than 25, 4) presence of a medical, neurological, or psychi-

atric condition that makes MRI scanning dangerous or confounds experimental hypotheses. 71 subjects were collected.

- **Procedure:**

  The study took place at the MIND Institute over one experimental session consisting of five hours in total. The Principal Investigator (PI), (REJ), conducted all the screening and interview procedures. The PI or his assistants conducted MRI scanning and administration of cognitive and personality measures. The experimental session consisted of consent and screening (.75 hours), MRI studies including morphological, diffusion tensor, and spectroscopic imaging (allowed for 2 hours) and cognitive/personality/creativity evaluation (2.25 hours) for a total of 5 hours. Participants were paid $50 for their time.

- **Magnetic Resonance Imaging (MRI) Studies:**

  MRI examinations were performed on a 1.5 Tesla Siemens Sonata scanner using an 8-channel phased array head coil. T1-weighted images were obtained with a 3D transversal gradient echo sequence (TR/TE=12/4.76ms, field of view (FOV) = 200x200mm, flip angle 200, resolution 192x192, slice thickness=1.5mm, structural scan time = 6 minutes 47 seconds). The scan time was 1.25 hours, but up to two hours were available if needed. *FreeSurfer* software was used to segment brain regions into regional brain volumes (i.e., cortical thickness x gyral area) (Fischl and Dale (2000) and Fischl et al. (2004)).

5.1 Implementing Control Charts for QA

Whole volume and normalized volume control limits are calculated by averaging all regional segmented volumes across subjects and finding the standard deviation. The Upper Control Limit and Lower Control Limit were calculated from the data by looking at 3 standard deviations from the sample mean. Since this is considered Phase I, the use of these control limits in Phase II would be changed by deleting outliers, and recalculating the mean and standard deviation. Many structures do

not follow a normal distribution, but Montgomery (2005) notes that for univariate control charts this assumption is not necessary.

For these data, we take the individual brain structure volumes along with the hemispherically summed and differenced volumes and divide them by the intracranial volume. This deflects egregious site, gender, volume, and age differences (Tofts (2003)). Standardized control charts should not be used on the normalized data because the normalization makes the brains (or segmentations) very similar. The data was normalized to make each brain as similar as possible which means standardizing the original data is not useful.

5.2 Overview of Charts

Subject 19 appears in all three quality control charts: Whole Volume Charts, Normalized charts, and Regression control charts and therefore can be used to illustrate each chart's features and interrelationships. Subject 19 is a 19-year-old right-handed female with an IQ of 116. Subjects were recruited mainly from the UNM campus which means that our sample is not a good representation of the general population of healthy 18-25 year olds. After contacting the researcher or his assistant, she was given a psychological review and interview. She then was given the MRI scan. A radiologist reviewed her scan to look for abnormalities. The data was later processed using the *Freesurfer* software, the measurements aggregated, and SAS was used to create quality control charts.

Whole volume charts flagged her as outside the upper bounds for Right Cerebellum White Matter and Summed Cerebellum White Matter. The Right Cerebellum White Matter is obviously large enough to place the summed structure out of control. She was not flagged as being quantitatively asymmetrical for this structure. Her normalized Right Cerebellum White Matter was also found to be out of control. IQ was found to be the only significant regressor for the Right and Summed Cerebellum White Matter. Gender, age, and handedness were not significant regressors and

were back selected out from the model. Note that each of the different methodologies flagged the same structure.

Whole Volume Charts

One example "whole volume control chart" is given in Figure 5.1. The volumes are measured in cubic millimeters. Charts for all volumes were created using the SAS macro programming language.

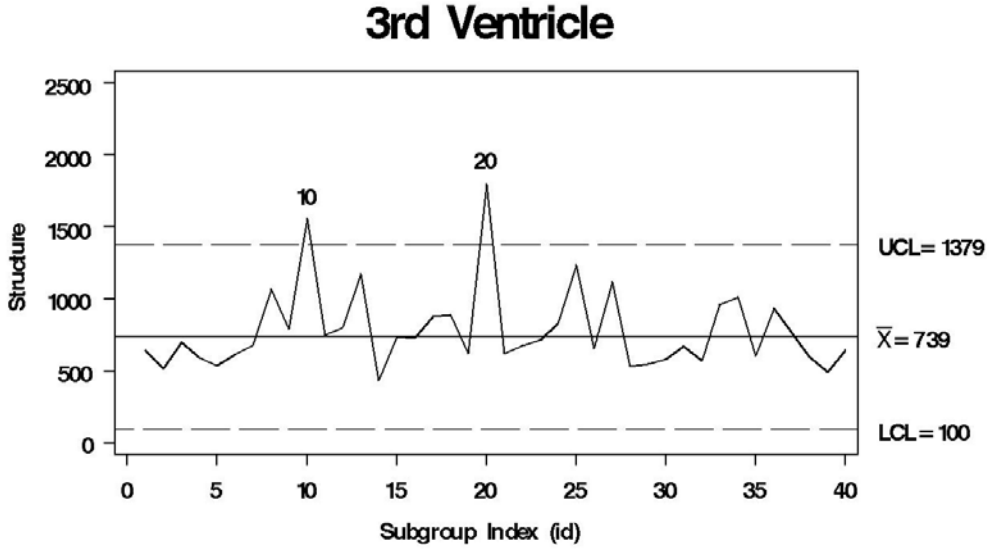

Figure 5.1: A whole volume control chart indicating two subject's ventricles are larger than expected. Enlarged ventricles are implicated in schizophrenia, Alzheimer's, alcoholism, and other mental disorders. Not all 71 subjects are shown, but the rest of the subjects were in control.

Whole volume charts are hindered because the tools being used to measure a product, that is not itself homogenous, have their own variability. Steen et al. (2007) also looked at healthy normal controls, but his subject's brain volumes were larger. A comparison of studies will show the fact that site has an affect on the volumet-

rics. Using Steen et al. (2007) to establish our control limits would render a good proportion of our subjects as being out of control. The *Freesurfer* volumes that are readily available for comparison came from multiple sites. We used our data for the control charts limits to avoid adding the variability across sites to charts. Biological differences, such as age or gender, affect brain volumes (Tofts (2003)). Because of the multitude of factors affecting brain volumetrics, several control limits may need to be established for different populations.

Table 5.1 is a summary from the whole volume control charts of the subjects' structures found to be outside of control limits. *ID* is the subject's unique identifier. *Volume* refers to the cubic millimeters of the structure, and *Structure (Whole Volume)* refers to which structure was found to be out of control. Structures with "Differenced" in the name occur when the left hemisphere has been subtracted from the right hemisphere. "Summed" means the hemispheres were summed. Differencing was performed to look for asymmetries across hemispheres which the radiologist is able to do qualitatively and summing was performed to see if the hemispheres balanced each other out. Most subjects found out of control break the upper control limit (a result that will be further explored with a larger data set). Again SAS Macros were used to generate this table and only subjects' structural volumes outside of the control limits are presented. This summary allows for the reader to see the number of subjects and structures that were tested and found to be outside normal limits. Notice that there are repeat outliers in subject 34. Our subjects were expected to be healthy normal controls, but (Vernooij et al. (2007); Katzman et al. (1999) also reported incidental findings in asymptomatic volunteers.

Percent Volumes

The subject flagged in the percent volume chart (Figure 5.2) was not flagged in the whole volume chart.

Subject 31 was considered healthy by the psychological reviews. Notice the subject is barely outside the control limits, but physiological data (i.e. regional brain volume measures) has a great amount of variability (Tofts (2003)).

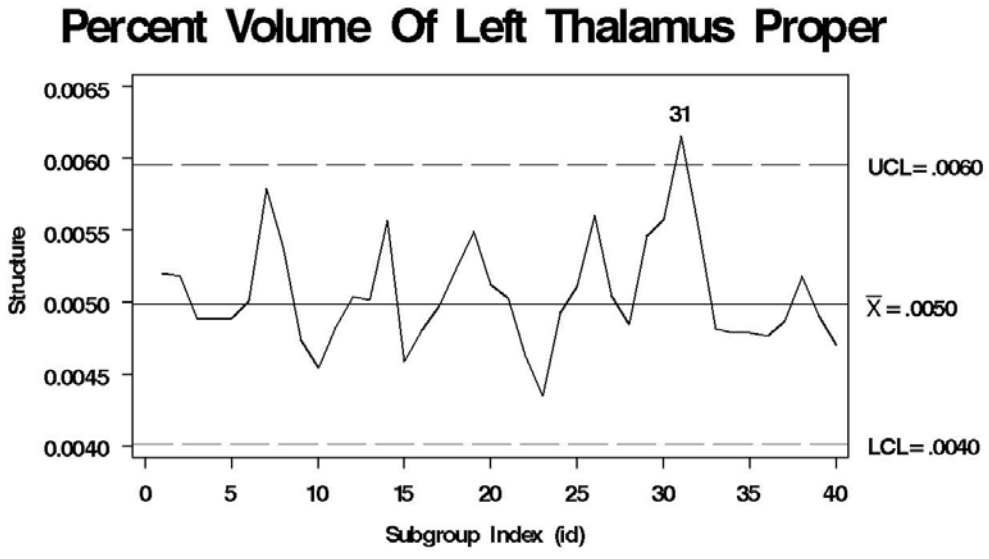

Figure 5.2: Control chart based on the percentage of total brain volume the brain structure occupies. These control charts are better suited for multiple site/scanner acquisitions.

Table 5.2 gives a summary of the structures found to be out of control for normalized volumes. *ID* refers to the subject's unique identifier. *Volume* refers to the percentage of the volume that the structure occupies relative to whole brain volume. *Structure (Normalized Volume)* refers to the structure that was outside of normal limits. "Differenced" and "Summed" have the same meaning as above, except for the normalization factor. SAS Macros were used to generate the table. Notice that more outliers were detected across more subjects and that percent volume charts are more sensitive. A type I error may occur here, i.e. flagging someone as out of control when they are not. Type I errors in health data are preferable to Type II errors because follow up will be performed if someone is considered outside normal limits. Note the number of repeat outliers (19, 20, and 34) within subjects and that subject 34 disappears as an outlier in the percent volume charts (Table 5.2). Subject

ID	Structure (Whole Volume)	Volume (mm^3)
2	Right Putamen	3987
8	Fourth Ventricle	4094
10	Third Ventricle	1556
15	Right Amygdala	2236
19	Right Cerebellum White Matter	20299
	Summed Cerebellum White Matter	44108
20	Left Lateral Ventricle	18670
	Third Ventricle	1798
22	Differenced Caudate	-701
31	Right Inferior Lateral Ventricle	1164
33	Summed Accumbens area	1683
34	Left Cerebral Cortex	377543
	Right Caudate	5597
	Right Cerebral Cortex	369986
	Summed Cerebral Cortex	747529
	Fourth Ventricle	4020
35	Differenced Cerebellum Cortex	-5439
52	Left Inferior Lateral Ventricle	1264
56	Right Cerebellum White Matter	19138
65	Left Lateral Ventricle	16808
68	Differenced Putamen	-953

Table 5.1: Outliers from control charts using original brain volume measures

34 does not have the largest intracranial volume measures of the study which makes the disappearance of the flag surprising.

Regression Control Charts

Hawkins (2003) and Mandel (1969) guided the use of regression control charts. It was intended that a multivariate control chart be used, and (Mason et al. (1995) suggested a method for decomposing Hotelling's T^2 statistic so that the actual cause of an out of control flag could be determined, but multivariate control charts did not prove to be an appropriate method for including behavioral data. Therefore, a regression control chart was more appropriate and simpler both in use and implementation.

The advantages to usng regressions are: (1) The number of variables included in

the regression model is limited only by the number of variables that are collected, but selection methods exist for decreasing the number of variables; (2) behavioral as well as other structural data can be incorporated into the model; (3) the residuals are a standardization of the dependent variable. The regressions highlight correlations between structures, and with behavioral data included, this can highlight well-correlated structures involved in disease processes and suggest new avenues of research.

Regression control charts (Mandel (1969)) are the next step to "normalize" a brain across subejcts since normalization by the intracranial volume does not handle gender differences. Nor is it likely to account for mental illness. Regression control charts can be very powerful and much more sensitive because they use the residuals as the plotting point rather than the original measure. Instead of accounting for intracranial volume, regression control charts can account for IQ, spatially-related structures, and psychiatric diagnosis. Unfortunately a researcher might be tempted to over-parameterize the control chart.

In order to prevent over-parameterization, the initial regressors were picked because of structural relationships, i.e. one structure in the left hemisphere was regressed by the other remaining left hemisphere structures. We did not use across hemisphere regressors due to the fact that structures across hemispheres were highly correlated and we would have issues with collinearity. We used PROC REG in SAS with selection "backward" to select variables.

The structural regressions performed were: Left against left, right against right, normalized left against normalized left, normalized right against normalized right. Since this study had demographic data, we also did regressions on age, gender, dominant hand use, and IQ. Since IQ is such a polarizing statistic, we also regressed with only IQ.

Women have smaller brain volumes than men, which would justify using intracranial normalization to account for gender effects, but the distribution of grey and white matter is also different between genders (Tofts (2003)). Women have

smaller white matter volumes and men have larger grey matter volumes Tofts (2003) which complicates the usefulness of intracranial normalization. Regression on many of the structures confirms that gender remains a significant regressor for normalized volumes in some regions. These structures were the right and left hemispheres of the thalamus, putamen, pallidum, hippocampus, cerebral white matter, cerebral cortex, cerebellum white matter, cerebellum cortex, and caudate. Gender was also significant when right and left hemispheres had been differenced which means that brain asymmetries are correlated with gender. The purpose of this study was control charts and so exploring gender as it relates to morphology was best left to other researchers, but these data suggested that gender may be more deterministic of the composition of brain morphology. Women and men not only have differences in the size of their brains, but also in the integration of structures. Perhaps rather than comparing raw measures from MR modalities or even normalized measures, studies might need to focus on comparing residuals after controlling for gender (or age).

Tables 5.3 and 5.4 represent the outliers found using regression control charts. As can be expected, many structures are listed as out of control multiple times. Subject 55 has a right accumbens that is flagged as out of control after regression on the structures in the right hemisphere, after normalization and regression on IQ, after normalization and regression on structures in the right hemisphere, after normalization and regression on Age, Gender, and IQ. This structure is obviously out of control and the multiple flags only support that. More structures are flagged with regression control charts, but some structures can be flagged multiple times.

The radiologist and psychologist flagged the subjects listed below, but for reasons our control charts were not able to detect. The *Freesurfer* software does not find cysts or demyelinating disease.[1] Myelin allows for faster transfer of information across axons; demyelination is a deficiency of myelin that causes the axonal connections to "short-circuit." The radiologist flagged someone for being asymmetrical, but this is a qualitative assessment. Our control charts can quantitatively look for volume

[1]There is software being developed through the National Alliance for Medical Image Computing (www.na-mic.org) project to automatically detect demyelination.

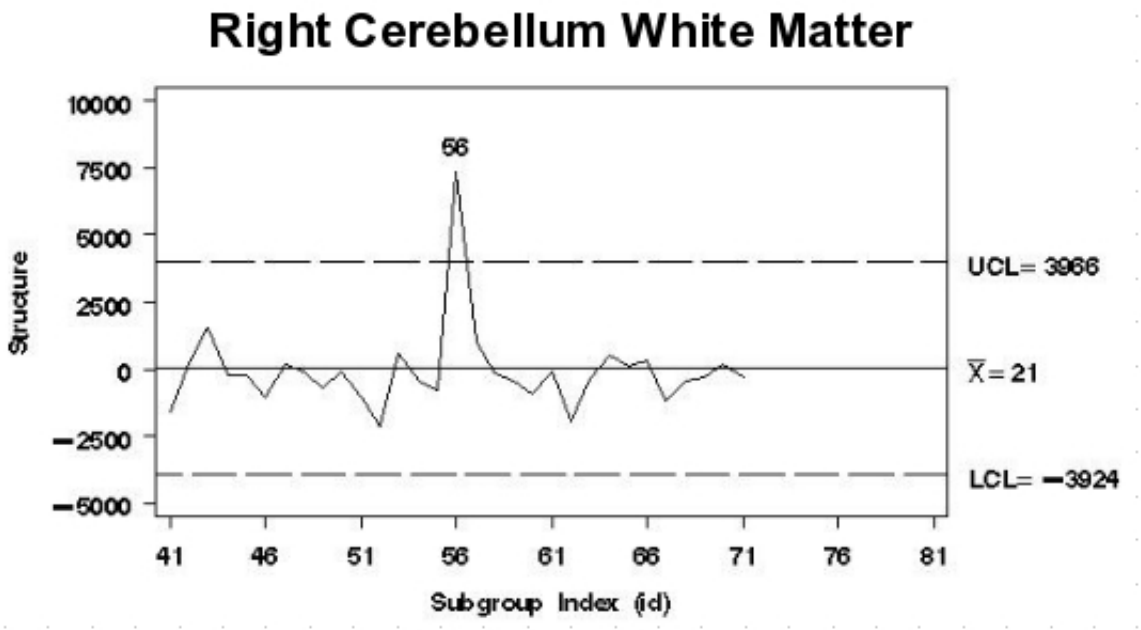

Figure 5.3: Regression control chart using other right hemisphere structures as predictors. The pallidum and cerebellum cortex were significant regressors after backwards selection.

asymmetry.

It is surprising to see that the *Summed Accumbens Area* for subject 33 in Table 5.1 is outside normal limits while the two separate hemispheres are not outside of normal limits. Such anomalies lead us to think multivariate control charts will need to be explored as well as other summations and differences. The radiologist's findings also hint at other research projects. If software could parcellate cysts from normal brain matter, the computer could make the measurement performed on subject 5 and inform the doctor of their presence in subject 55. Demyelinating disease is currently being studied in other projects. Asymmetries can also be assessed if that particular structure was measured. Quantitative scans and the software that parcellates the results can add numeric credence to the radiologists professional judgment.

5.3 The Costs of Scanning

The cost of an MRI scan at the Mind Research Network is ≈$650/hr. This study required 1.25 hours of scan time, but studies involving functional MRI can be longer (2-3 hours). Poor structural images that cannot be processed can ruin a region of interest analysis. Quality control of the structural images would help to mitigate this loss of data because it is much easier to repeat a 7 minute structural scan while a subject is already in the scanner than it is to get a subject to return, especially if they have traveled for the scan. Moderately-sized studies often have 50-100 subjects which translates to 50-300 hours of scan time or ≈$32,500-$195,000. A low 1% attrition rate would mean as little as ≈$325-$1,950 in lost scan time. Most studies we have been associated with lose more than 1% of their data due to bad scans. There are costs associated with recruitment and analysis which are not factored in. Those costs increase significantly for harder to find populations such as lupus patients (personal experience with other projects). From a cost perspective, performing automated QA on scans is prudent.

5.4 Conclusions and Future Work

Control Charts are not a substitute for radiological review, but can quantify the radiologists findings. They are intended to assure that the subject is in the same range expected for their health status and thereby keep study populations as close to expectations as possible. They can give automated feedback about a patient, and alert the technicians to bad scans sooner than a one week review would. If a patient is outside of normal limits, they can signal a need for a re-evaluation. In the future, control charts may aid in rapid detection of brain abnormalities, bad scan runs, and flag subjects who may be outliers in study cohorts. Control charts can help in the assessment of drugs being developed, especially when a series of images are taken over a period of time (cf Tofts (2003)). As an example, a brain infarct causes a loss of blood flow to a region and thereby increases atrophy in that region. A drug

that increases blood flow within the brain can be statistically assessed, based on morphological measures taken over several weeks, to see if the drug actually reduces atrophy and the rate of deterioration.

Future imaging technology will likely be faster and increase both in sensitivity and specificity allowing for cheaper scans that have greater resolution. While this would make them a valuable routine diagnostic tool, as the resolution increases, the amount of time a radiologist spends on each scan (as well as cost) will increase. Using automated procedures can assist in flagging outliers and inconsistencies that might occur and be of great benefit in other imaging modalities. With greater scan resolution, the human eye will not be able to capture every detail and a down sampling of the data will be a likely work around. Software and computers do not require down sampling. Both technologies can be used to parse out brain structures, look for tumors or lesions, and with repeated scans, the technologies can begin to discern changes in pathology.

Neuropsychologists look at covariates along with images which suggests a regression control chart. Regressing the volume measures by the behavioral data allows the neuropsychologist to standardize the volume measures across patient type and demographics. Using the resulting residuals would allow for one set of control limits to be used across several differing populations. A neuropsychologist would then have another quantitative measure to compliment the patient assessments.

In the future, we would like to explore multivariate control charts because many disease processes affect more than one region. Difficulties lie in transforming the data and checking for normality, relation of structures in disease, and which multivariate control chart(s) should be implemented for most efficient detection. Decomposition of Hotelling's T^2 statistic can facilitate detecting which structures are contributing most to an out of control flag (Mason et al. (1995)).

Other imaging modalities were involved in this study, and we plan on developing quality control charts for the Diffusion Tensor imaging data. Multivariate control charts and regression control charts can be used on these other modalities as well.

The different image modalities can be used as parameters to the multivariate control chart.

ID	Structure (Normalized Volume)	Percent of Total Brain Volume
8	Fourth Ventricle	0.00267
10	Third Ventricle	0.00092
17	Left Cerebellum Cortex	0.04572
	Right Cerebellum Cortex	0.04661
	Summed Cerebellum Cortex	0.09233
18	Fourth Ventricle	0.00265
19	Right Cerebellum White Matter	0.01146
20	Third Ventricle	0.00094
22	Differenced Caudate	-0.00046
31	Left Thalamus Proper	0.00615
	Right Inferior Lateral Ventricle	0.00083
	Summed Inferior Lateral Ventricle	0.00136
	Summed Thalamus Proper	0.01164
32	Right Thalamus Proper	0.00578
35	Differenced Cerebellum Cortex	-0.00303
51	Differenced Cerebellum Cortex	0.00450
	Differenced Cerebral Cortex	0.01217
52	Left Cerebral White Matter	0.13609
	Left Inferior Lateral Ventricle	0.00075
53	Left Amygdala	0.00150
	Right Amygdala	0.00144
	Summed Amygdala	0.00295
55	Right Accumbens area	0.00056
56	Right Cerebellum White Matter	0.01283
	Summed Cerebellum White Matter	0.02683
65	Left Lateral Ventricle	0.00987
68	Differenced Putamen	-0.00056

Table 5.2: Control charts using normalized brain volumes.

ID	Structure
12	Regressed IQ on Left Accumbens area
	Regressed Behavioral Data on Left Accumbens area
	Regressed Behavioral Data on Differenced Accumbens area
13	Right Structures Regressed on Right Caudate
	Regressed Behavioral Data on Right Caudate
	Regressed Behavioral Data on Summed Caudate
16	Regressed Behavioral Data on Differenced Inferior Lateral Ventricle
17	Left Structures Regressed on Left Cerebellum Cortex
	Right Structures Regressed on Right Cerebellum Cortex
19	Regressed IQ on Right Cerebellum White Matter
	Regressed IQ on Summed Cerebellum White Matter
	Regressed Behavioral Data on Right Cerebellum White Matter
	Regressed Behavioral Data on Summed Cerebellum White Matter
20	Regressed IQ on Left Lateral Ventricle
	Regressed Behavioral Data on Left Lateral Ventricle
22	Regressed IQ on Differenced Caudate
	Regressed Behavioral Data on Differenced Caudate
26	Regressed IQ on Left Thalamus Proper
	Regressed IQ on Summed Pallidum
	Regressed IQ on Summed Thalamus Proper
	Regressed Behavioral Data on Left Pallidum
	Regressed Behavioral Data on Left Thalamus Proper
	Regressed Behavioral Data on Right Pallidum
	Regressed Behavioral Data on Right Thalamus Proper
	Regressed Behavioral Data on Summed Pallidum
	Regressed Behavioral Data on Summed Thalamus Proper
31	Right Structures Regressed on Right Inferior Lateral Ventricle
	Regressed IQ on Right Inferior Lateral Ventricle
	Regressed IQ on Left Thalamus Proper
	Regressed IQ on Right Inferior Lateral Ventricle
	Regressed IQ on Summed Inferior Lateral Ventricle
	Regressed Behavioral Data on Right Inferior Lateral Ventricle
	Regressed Behavioral Data on Summed Inferior Lateral Ventricle

Table 5.3: Outliers from control charts using residuals from regression

ID	Structure
32	Right Structures Regressed on Right Thalamus Proper
35	Regressed IQ on Differenced Cerebellum Cortex
	Regressed Behavioral Data on Differenced Cerebellum Cortex
36	Regressed IQ on Left Putamen
	Regressed IQ on Right Putamen
	Regressed IQ on Summed Putamen
	Regressed Behavioral Data on Left Putamen
	Regressed Behavioral Data on Right Putamen
	Regressed Behavioral Data on Summed Putamen
38	Regressed Behavioral Data on Left Pallidum
	Regressed Behavioral Data on Summed Pallidum
52	Regressed IQ on Left Inferior Lateral Ventricle
53	Regressed IQ on Left Amygdala
	Regressed IQ on Summed Amygdala
	Regressed Behavioral Data on Left Accumbens area
	Regressed Behavioral Data on Left Amygdala
54	Right Structures Regressed on Right Lateral Ventricle
55	Right Structures Regressed on Right Accumbens area
	Regressed IQ on Right Accumbens area
	Regressed Behavioral Data on Right Accumbens area
56	Right Structures Regressed on Right Cerebellum White Matter
63	Left Structures Regressed on Left Cerebral Cortex
65	Regressed IQ on Left Lateral Ventricle
68	Regressed IQ on Differenced Putamen
	Regressed Behavioral Data on Differenced Putamen

Table 5.4: Outliers from control charts using residuals from regression, continued

Table 5.5: Findings by the radiologist

ID	RECOMMENDATION	Keep in Study
5	Benign finding. Cyst of the par intermedia measuring 4mm.	YES
42	Ventricular asymmetry, normal variant. No action needed.	YES
44	Findings consistent with demyelinating disease; MS.	NO
55	Small bilateral mucous retention cysts in the maxillary sinuses are of little clinical significance. No specific action required.	YES
61	None. Moderate to severe chronic paranasal sinus disease.	YES
64	Asymmetry of the temporal horns with the right larger than the left. May be normal anatomic variant. If subject has history of seizure, this may be an early finding of mesial temporal sclerosis.	NO
67	Correlate to subject history	NO

References

Benneyan, J., Lloyd, R., and Plsek, P. (2003). Statistical process control as a tool for research and healthcare improvement. *British Medical Journal*, 12(6):458–464.

Blacker, D., Lee, H., Muzikansky, A., Martin, E., Tanzi, R., McArdle, J., Moss, M., and Albert, M. (2007). Neuropsychological measures in normal individuals that predict subsequent cognitive decline. *Archives of Neurology*, 64(6):862.

Castellanos, F., Giedd, J., Berquin, P., Walter, J., Sharp, W., Tran, T., Vaituzis, A., Blumenthal, J., Nelson, J., Bastain, T., et al. (2001). Quantitative brain magnetic resonance imaging in girls with attention-deficit/hyperactivity disorder. *Archives of General Psychiatry*, 58(3):289.

Doi, K. (2007). Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, 31(4-5):198–211.

Duncan, A. (1965). *Quality Control and Industrial Statistics. 3d ed.* Richard D. Irwin, Homewood, Illinois.

Fischl, B. and Dale, A. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D., Busa, E., Seidman, L., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., and Dale, A. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, 14(1):11–22.

Garvican, L. and Field, S. (2001). A pilot evaluation of the R2 image checker system and users' response in the detection of interval breast cancers on previous screening films. *Clinical Radiology*, 56(10):833–837.

Grachev, I., Breiter, H., Rauch, S., Savage, C., Baer, L., Shera, D., Kennedy, D., Makris, N., Caviness, V., and Jenike, M. (1998). Structural abnormalities of frontal neocortex in obsessive-compulsive disorder. *Archives of General Psychiatry*, 55(2):181–182.

Hashemi, R., Bradley, W., and Lisanti, C. (1997). *MRI: The Basics*. Lippincott Williams & Wilkins.

Hawkins, D. (2003). Regression adjustment for variables in multivariate quality control. *Journal of Quality Technology*, 25(3):170–182.

Johnson, M., Munakata, Y., and Gilmore, R. (2002). *Brain Development and Cognition: A Reader*. Blackwell Publishers.

Jovicich, J., Czanner, S., Greve, D., Haley, E., van der Kouwe, A., Gollub, R., Kennedy, D., Schmitt, F., Brown, G., MacFall, J., Fischl, B., and Dale, A. (2006). Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data. *Neuroimage*, 30(2):436–443.

Katzman, G. L., Dagher, A. P., and Patronas, N. J. (1999). Incidental findings on brain magnetic resonance imaging from 1000 asymptomatic volunteers. *Magnetic Resonance Imaging of the Brain*, 281(1):36–39.

Kessler, A. (2006). *The End of Medicine: How Silicone Valley (and naked mice) will Reboot Your Doctor*. HarperCollins.

Lezak, M. (1995). *Neuropsychological Assessment*. Oxford University Press, New York, 3d ed. edition.

Magnotta, V., Harris, G., Andreasen, N., O'Leary, D., Yuh, W., and Heckel, D. (2002). Structural MR image processing using the BRAINS2 toolbox. *Computerized Medical Imaging and Graphics*, 26(4):251–64.

Mandel, B. (1969). The regression control chart. *Journal of Quality Technology*, 1(1):1–9.

Manos, A., Sattler, M., and Alukal, G. (2006). Make healthcare lean. *Quality progress*, 39(7):24–30.

Mason, R., Tracy, N., and Young, J. (1995). Decomposition of T^2 for multivariate control chart interpretation. *Journal of quality technology*, 27(2):109–119.

Mattson, J. and Simon, M. (1996). *The pioneers of NMR and Magnetic resonance in medicine: the story of MRI*. Bar-Ilan Univeristy Press.

Miller, D., Barkhof, F., Frank, J., Parker, G., and Thompson, A. (2002). Measurement of atrophy in multiple sclerosis: pathological basis, methodological aspects and clinical relevance. *Brain*, 125(8):1676–1695.

Montgomery, D. C. (2005). *Introduction to Statistical Quality Control*. John Wiley and Sons, 5th Ed. edition.

Plessen, K., Bansal, R., Zhu, H., Whiteman, R., Amat, J., Quackenbush, G., Martin, L., Durkin, K., Blair, C., Royal, J., Hugdahl, K., and Peterson, B. (2006). Hippocampus and amygdala morphology in attention-deficit/hyperactivity disorder. *Archives of General Psychiatry*, 63(7):795.

Ryan, T. and Ryan, A. (2000). *Statistical methods for quality improvement*. Wiley New York.

Scahill, R., Frost, C., Jenkins, R., Whitwell, J., Rossor, M., and Fox, N. (2003). A Longitudinal Study of Brain Volume Changes in Normal Aging Using Serial Registered Magnetic Resonance Imaging. *Archives of Neurology*, 60(7):989–994.

Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. Princeton.

Steen, R., Hamer, R., and Lieberman, J. (2007). Measuring brain volume by MR imaging: Impact of measurement precision and natural variation on sample size requirements. *American Journal of Neuroradiology*, 28(6):1119.

Strakowski, S., DelBello, M., Sax, K., Zimmerman, M., Shear, P., Hawkins, J., and Larson, E. (1999). Brain magnetic resonance imaging of structural abnormalities in bipolar disorder. *Archives of General Psychiatry*, 56(3):254.

Summers, R. (2003). Road Maps for Advancement of Radiologic Computer-aided Detection in the 21st Century 1. *Radiology*, 229(1):11–13.

Szeszko, P., Robinson, D., Alvir, J., Bilder, R., Lencz, T., Ashtari, M., Wu, H., and Bogerts, B. (1999). Orbital frontal and amygdala volume reductions in obsessive-compulsive disorder. *Archives of General Psychiatry*, 56(10):913.

Tofts, P., editor (2003). *Quantitative MRI of the Brain: measuring changes caused by disease*. Jon Wiley and Sons Ltd., The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England.

van der Naalt, J. (2001). Prediction of outcome in mild to moderate head injury: A review. *Journal of Clinical and Experimental Neuropsychology*, 23(6):837–851.

Vernooij, M., Ikram, M., Tanghe, H., Vincent, A., Hofman, A., Krestin, G., Niessen, W., Breteler, M., and Van der Luyt, A. (2007). Incidental findings on brain MRI in the general population. *New England Journal of Medicine*, 357(18):1821–1828.

Woodall, W. (2006). The use of control charts in health-care and public-health surveillance. *Journal of Quality Technology*, 38(2):89–104.

Worth, A., Makris, N., Kennedy, D., and Caviness Jr, V. (2001). Accountability in methodology and analysis for clinical trials involving quantitative measurements of MR brain images. Technical report, Technical Report TR20011117, Neuromorphometrics, Inc.