Comparing Estimates and Item Nonresponse Rates of Interviewers Using Statistical Process Control Techniques[1]

Robyn Sirkis
U.S. Census Bureau
robyn.b.sirkis@census.gov

1. Introduction

Statistical process control (SPC) involves using statistical techniques to measure and analyze the variation in processes. Pioneered by Walter A. Shewhart in the early 1920's, SPC was later applied by W. Edwards Deming to successfully improve quality in manufacturing processes. SPC, however, has received only sporadic attention in survey practice and government data collection (Pierchala and Surti, 2009). There are challenges in applying SPC in a survey setting which includes complex sample designs and many moving data collection parts. For example, at the U.S. Census Bureau, management is a joint effort between headquarters staff, regional office (RO) staff, and interviewers who perform the data collection.

While data collection is more complex than a manufacturing assembly line, there is potential to use SPC in a survey setting. There is a vast amount of data available to guide survey managers in their efforts to improve interviewer performance and data quality. SPC is a powerful tool that can be used to analyze and summarize that information to optimize survey management decision-making. SPC can be used to establish quality standards and areas for process improvement, and can be an invaluable tool to examine data while it is being collected. This provides an opportunity to make adjustments and improve data quality in real time, prior to release of the data.

For many reasons SPC should be considered as a useful tool to evaluate interviewer performance. Flaws in the interview process can decrease the quality of the data. Interviewers contribute both to the bias and variance of survey estimates. An estimate can be biased if interviewers systematically influence their respondents to choose a particular response alternative, probe incorrectly for insufficient responses, fail to record information correctly provided by the respondent, and fail to read the question as worded. An interviewer may record more "no" responses in order to skip questions to shorten the interview or read the first response category to the question instead of asking the respondent the entire question and reading all the response categories. Item nonresponse rates can be an indication of problems with interviewer performance and data quality (Japec, 2002). High item nonresponse can introduce bias into estimates and reduce analyzable sample sizes. Lower item nonresponse rates are preferable, however an extremely low item nonresponse rate can indicate potential interviewer falsification. This can all introduce bias into the estimates.

A survey manager or survey practitioner can monitor quality indicators or measures such as item nonresponse rates or estimates from a survey with SPC charts. An example of an estimate from a survey is the proportion of "no" responses to a question. In this paper, we describe a set of SPC tools, namely control charts and analysis of means (ANOM) charts that can be utilized to monitor interviewer performance and data quality. An example of a quality indicator is the proportion of "not at all" responses to the question, "Do you NOW smoke cigarettes every day, some days or not at all?" in the National Health Interview Survey. In the ANOM charts the proportion of "not at all" responses for an interviewer would be compared to the proportion of "not at all" responses to the entire RO for a given month to determine if significant differences exist. In an interviewer level control chart, the proportion of "not at all" responses is calculated for each month over the course of a few years for an interviewer and the process would be examined for any unexpected or unpredicted variation. The other quality indicators described in the research include the proportion of responses in a question or estimate excluding don't know and refusals, the proportion of don't know and refusal responses for a question or estimate, and the average values for a question. An example of an estimate is the current asthma estimate, which is a combination of responses from several questions. The proportion of don't know and refusal responses is considered item nonresponse in this research and sometimes considered paradata. Paradata are any information that describes a data collection process (Couper, 1998; Couper

[1] Disclaimer: Any views expressed are those of the authors and not necessarily those of the U. S. Census Bureau.

and Lyberg, 2005; Laflamme, 2008). Recently this has been expanded to include information that may be recorded by interviewers.

One of the goals of this research is to note the importance of examining multiple types of quality indicators for each interviewer to make decisions about the impact interviewer performance has on data quality. For example, an interviewer can have a low proportion of don't know and refusal responses for income, but an average income that is significantly different from the overall average income for the RO.

2. Background

The SPC techniques demonstrated in this paper were applied to data collected from the 2008 to 2010 National Health Interview Survey (NHIS).

2.1 National Health Interview Survey

The National Health Interview Survey (NHIS) is an annual survey of the health of the civilian, noninstitutionalized household population of the United States, and is conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC). The survey produces nationally representative data on health insurance coverage, health care access and utilization, health status, health behaviors, and other health-related topics.

The NHIS is a cross-sectional household interview survey that is continuous throughout the year. The sampling plan consists of a multistage area probability design that allows for a representative sampling of households and noninstitutional group quarters. The current sampling plan was implemented in 2006. The first stage of the current sampling plan consists of a sample of 428 primary sampling units (PSU's) drawn from approximately 1,900 geographically defined PSU's that cover the 50 States and the District of Columbia. A PSU consists of a county, a small group of contiguous counties, or a metropolitan statistical area. The two types of second-stage units within a PSU are area segments and permit segments. Area segments are defined geographically and contain an expected eight, twelve, or sixteen addresses. Permit segments include housing units built after the 2000 census.

The current NHIS sample design includes an oversampling of Black, Hispanic, and Asian persons. The households and noninstitutional group quarters selected for interview each week in the NHIS are a probability sample representative of the target population. .

Roughly 650 trained interviewers with the U.S. Census Bureau, operating out of 12 ROs, administer in-person interviews on laptop computers using computer-assisted personal interviewing (CAPI), with some telephone follow-up. The survey instrument contains four main modules: household composition, family, sample child, and sample adult. In addition to the core survey modules, supplemental questions on special topics are added to the NHIS questionnaire each year. For more information on the NHIS see Centers for Disease Control and Prevention, 2009.

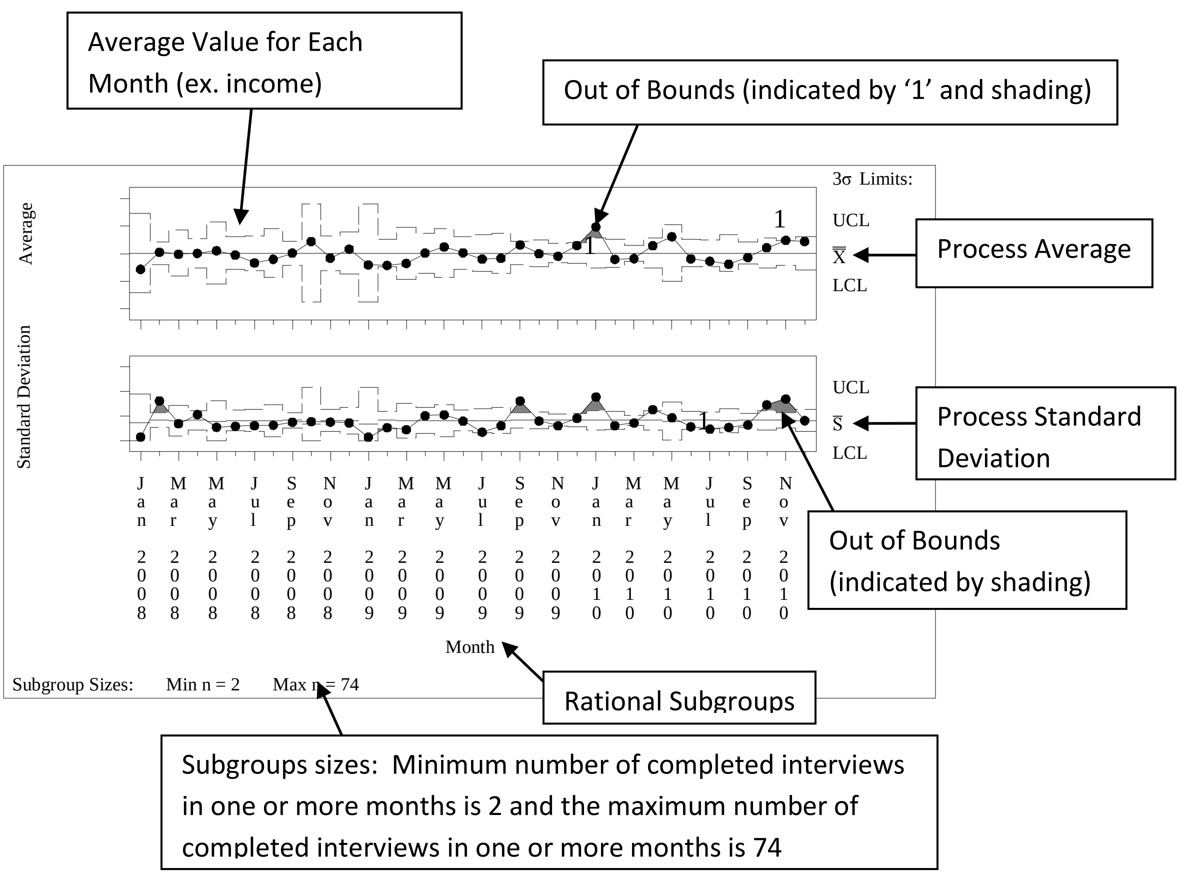
2.2 Statistical Process Control Terminology

Statistical process control (SPC) is a method used to analyze and measure the variation of a process. The fundamentals of SPC and control charts were developed for manufacturing processes by Dr. Walter A. Shewhart during the1920's (Ryan, 1989). In a survey setting such as the U.S. Census Bureau, communication occurs between the headquarters staff, RO staff, and interviewers working in the survey field environment. The equivalent of the production output from the machines are the interviews conducted by the interviewers in the field. It is not as directly modifiable like changing settings on a machine. However, this does not mean that SPC techniques cannot be applied to complex processes in a survey setting, especially one that is repeated over time under standard procedures in which statistical tolerance limits can be established (Jans, Sirkis, and Duffy, 2010).

There are a wide variety of SPC quality control tools including Pareto charts, fishbone diagrams, control charts, and ANOM charts. The focus of this paper is control charts, ANOM charts, and multivariate charts. A typical control chart or shewhart chart is a graphical display of a quality indicator or measure over time. The chart's center line is the average value of the quality indicator, and is called the process average, bounded by upper and lower control limits. The process is in a state of statistical control if all sample points are within the control limits and do not show any systematic patterns (Western Electric Co., Inc., 1956). There are numerous types of control charts such as average and standard deviation charts and proportion of nonconforming items charts. The SHEWHART procedure in SAS produces these charts.

The average and standard deviation control charts were used in the analysis for continuous quality indicators such as the NHIS total family income. The SHEWHART procedure was used to produce the average and standard deviation charts also known as an xschart. Figure 1 shows an example of the average and standard deviation chart. The process average, process standard deviation, rational subgroups, and control limits are shown in Figure 1. The top part of the chart shows the average total family income and the central tendency of the process. The bottom part of the chart shows the standard deviation and the variability of the process for the total family income. The standard deviation chart is used to determine if the distribution for the quality indicator is stable. The process average is the expected average for the total family income and is indicated by a constant straight line. The process standard deviation will vary since there are different sample sizes or number of completed interviews for this analysis for each month. The values on the y-axis and process average were removed for nondisclosure purposes. The chart uses 3-sigma limits. Months were used as the rational subgroups on the x-axis. Each point on the graph is either the average or standard deviation (depending on the chart) for the quality indicator for the month. One method of interpreting the two charts together is to first explore the standard deviation chart (the lower chart) and determine if it is in control. For points where the standard deviation exceeds the limits the corresponding points in the average chart should not be interpreted regardless of whether it is outside the control limits. The average total family income was outside the control limits indicated by the shading and test ‘1’ label. The test ‘1’ label was not used for the standard deviation part of the chart since there were many points outside the control limits. More details on these SPC concepts is described later in the section. For information on the average and standard deviation formulas see the SAS documentation http://www.okstate.edu/sas/v8/saspdf/qc/chap44.pdf.

Figure 1: Average and Standard Deviation Statistical Process Control Chart



The proportion of nonconforming items charts were used in the analysis for quality indicators involving proportions such as the NHIS total family income item nonresponse rate. The proportion of nonconforming items represents the fraction of the number of items that are unacceptable or defective. The SHEWHART procedure was used to produce the proportion of nonconforming item control charts also known as a pchart. Figure 2 shows an example of the proportion of nonconforming items chart. The process average, rational subgroups, and control limits are shown in Figure 2. The process average is the expected average for the total family income item nonresponse rate and is indicated by a constant straight line. The process average is the expected average for all data included in the chart. The process average and the values on the y-axis are proportions. The process average was 22 percent. The chart uses 3-sigma limits. Months were used as the rational subgroups on the x-axis. Each point on the graph is the total family income item nonresponse rate for the month. The process was not in control since the total family income item nonresponse rate was outside the control limits for at least one month indicated by the shading or test '1' label. More details on these SPC concepts is described later in the section. The proportion of nonconforming items formulas are shown below and can be found at http://www.okstate.edu/sas/v8/saspdf/qc/chap38.pdf.

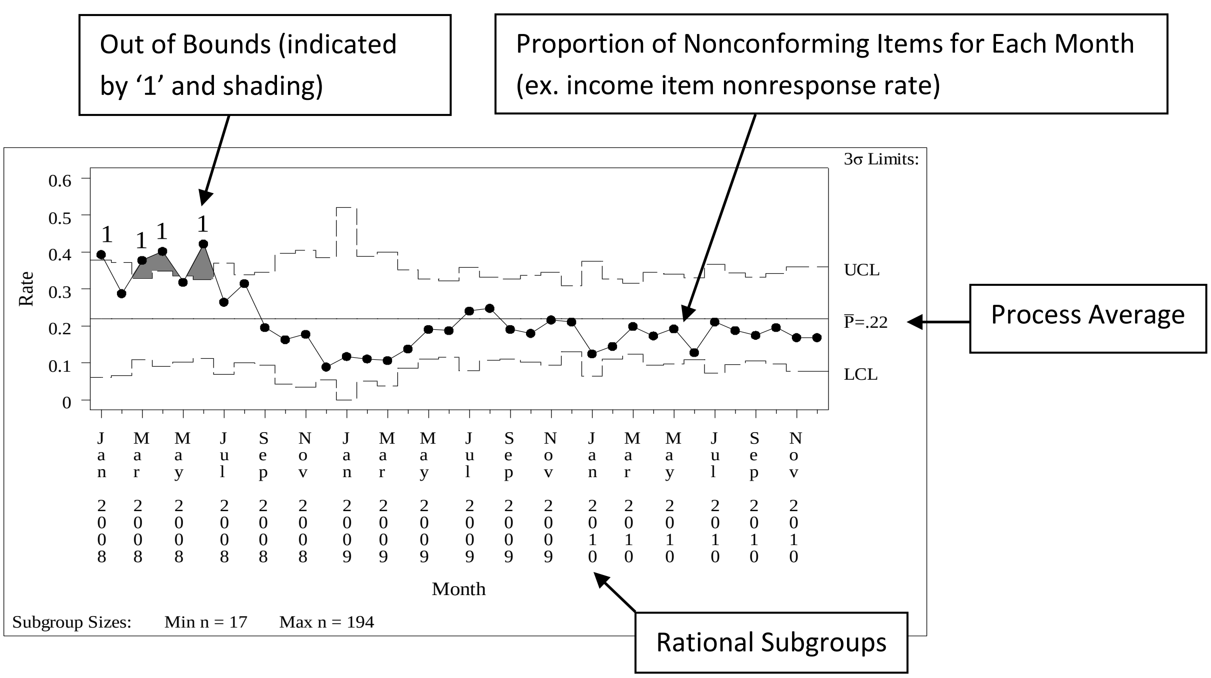
$\bar{p} = \frac{n_1 p_1 + \cdots + n_N p_N}{n_1 + \ldots + n_N}$, where

$\bar{p} =$ the average proportion of nonconforming items across the subgroups
$p_i =$ proportion of nonconforming items in the i^{th} group
$n_i =$ number of items in the i^{th} group
$N =$ number of subgroups

$$Upper\ Control\ Limit = \min(\bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}}, 1)$$

$$Lower\ Control\ Limit = \max(\bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}}, 0)$$

Figure 2: Proportion of Nonconforming Items Statistical Process Control Chart



A control chart is useful to the degree that it can identify variability in a process which can be changed to reduce the variability over time (Jans, Sirkis, and Duffy, 2010). A process should only be monitored or measured if the survey practitioner clearly defines the goals of the process. One of the primary goals is to not just maintain the process but to modify and improve the process in real time. Ultimately every process will display variation. Two types of variation are common cause variation and special cause variation (Gitlow, 1989). Common cause variation is random variation inherent in the process. It is stable and is seen when data points fall within the control limits and do not display any unique patterns. It is considered an integral part of the process, and typically requires more skill in reducing this variation. The system as a whole would have to be changed to improve the process. Special cause variation is nonrandom, unpredictable, and is seen when data either lie outside the control limits or display certain patterns. Survey managers should take immediate action to determine these sources of variation. The control charts should be used to identify when special causes occur so that one can eliminate them and bring the process into control without overreacting to normal variability. Special causes should not be ignored by assuming they are part of the process since they can be caused by outside factors (ex. holidays, seasonal effects). A survey manager, survey researcher, or survey methodologist should ensure that the control chart is not comparing more than one process and therefore, displaying false signals.

A core concept of the control chart is the rational subgroup. In a control chart the rational subgroups are time oriented and are displayed on the x-axis (Montgomery, 1985). Examples of rational subgroups include months and weeks. The goal of rational subgrouping is to produce groups that should, on average, have high between group variability and low within group variability when special causes of variation are present. There are other SPC charts that utilize rational subgroups other than time. It is imperative that the survey researcher adequately define the rational subgroups since this will ensure that reasonable expectations of variation are established before monitoring the process.

In addition to stating the rational subgroups the researcher should decide at what level to produce the control charts. For example, at the U.S. Census Bureau each RO and the interviewers within them may function differently. The control charts at the interviewer level would be helpful to managers who only have authority over the interviewers who work out of their own offices. The decision to produce the control charts at the interviewer level is to accurately identify special cause variation. Adding in additional sources of variation such as multiple ROs may increase the chance that special cause variation will be hidden within common cause variation reducing the likelihood of identifying problems (Jans, Sirkis, and Duffy, 2010). It may make sense to analyze ROs separately since geographic differences may exist between them. Therefore, producing individual interviewer level control charts within each RO may be more useful in terms of controlling the variation in the process.

The control charts can be used in real time in a production environment. SPC involves establishing practical control limits for a process and applying those limits in a real time setting. The ideal limits for a given process are guided by the survey management's quality standards. They need to decide how strict they want their quality standards and their acceptable level of variability in the process. The default control limits set by the software procedures may not be appropriate for the quality indicators being monitored (Jans, Sirkis, and Duffy, 2010). Also the control limits can be set differently for different quality indicators depending on one's requirements.

Creating control limits involves several steps. If this is the first time constructing a control chart for the quality indicator then it is necessary to determine trial control limits using past historical data. If there are points outside the control limits or special causes then these points should be investigated. If a cause can be assigned to these points then the points can be removed and the limits should be recomputed. This can result in new limits being tight and close together resulting in possibly more points that were barely inside the prior control limits but are now outside the new control limits (Ryan, 1989). Once the process is in a state of statistical control, the new data can be included with the past historical data to construct a control chart. However, the process average and control limits will be taken from the control chart which only included the past historical data and where the process was in a state of statistical control.

Customarily, the chart shows three sigma limits (Ryan, 1989). The use of three sigma limits was chosen by Walter A. Shewhart to provide the needed sensitivity without causing an unacceptable number of false alarms. These limits provide effective action limits when applied to real world data. A smaller sigma level denotes less variability, or close data points. Three sigma limits are also very robust and work with data that does not meet normality assumptions. Three sigma limits provide a balance between the Type I and Type II errors (Wheeler, 1995). There is no explicit rule that dictates that three sigma limits must be used in the control charts.

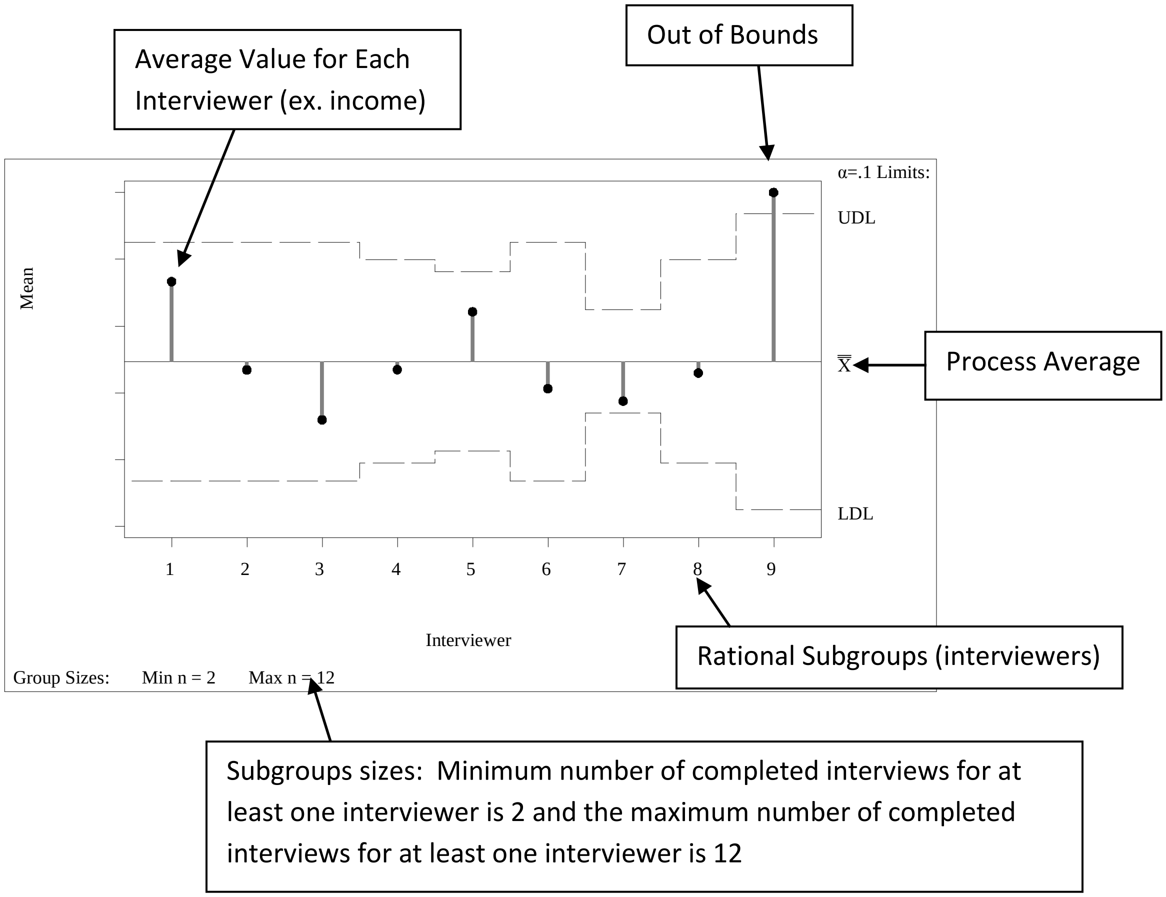
A couple of common misconceptions for using SPC charts are that the data used on a control chart must be normally distributed. Some authors have stated that three sigma limits work well even when the data are not perfectly normal (Wheeler, 1995). When using the data the manager or researcher should ensure that the data is in a form that is easy to interpret. It may be beneficial to avoid nonlinear transformations of the data such as logarithms, and probabilistic transformations (Wheeler, 1995). It is crucial to determine if it would be beneficial for making informative decisions about the variation of the process to transform data that does not follow normality assumptions.

The goal of using the SPC charts is to bring processes that are not into control into control by identifying and reducing special cause variation. A process is in control when all variation is due to common cause variation. If the quality indicator is outside the control limits indicating that the process is not in control there will be shading between the quality indicator and the control limits. There are also eight tests which are based on the Western Electric Rules that were developed for identifying special causes or patterns. These include: a) 1 point beyond 3 sigma limits, b) 9 points in a row on one side of average, c) 6 points in a row steadily increasing or decreasing, d) 14 points in a row alternating up and down, e) 2 out of 3 points in a row beyond 3 sigma, f) 4 out of 5 points in a row beyond 2 sigma, g) 15 points in a row beyond on either side of average, h) 8 points in a row on either or both side beyond 2 sigma and no points within 2 sigma (Western Electric Rules, 1956).

Other charts such as the ANOM chart, compare rational subgroups, but do not require a time element. The ANOM chart graphically compares the average or proportion of each rational subgroup to the overall average or proportion. The ANOM chart has an overall process average with upper and lower decision limits. If the quality indicator exceeds the decision limit for the subgrouping, than that group is considered to be statistically different from the overall mean or proportion. There are numerous types of ANOM charts such as average charts and proportion of nonconforming items charts. The ANOM procedure in SAS produces these types of charts.

The average ANOM charts were used in the analysis for continuous quality indicators such as the NHIS total family income. The ANOM procedure was used to produce the average charts also known as an xchart. Figure 3 shows an example of the average chart. The process average, rational subgroups, and decision limits are shown in Figure 3. The process average is the expected average for the total family income and is indicated by a constant straight line. The values on the y-axis and process average were removed for nondisclosure purposes. The significance level was set to 0.1. Interviewers were used as the rational subgroups on the x-axis. Each point on the graph is the average total family income for the interviewer. The average total family income was outside the decision limits for an interviewer indicating that there were significant differences between their average total family income and the overall average total family income of all the interviewers. For information on the average ANOM chart formulas see the SAS documentation http://support.sas.com/publishing/pubcat/chaps/59050.pdf.

Figure 3: Average Analysis of Means Chart



The proportion of nonconforming items ANOM charts were used in the analysis for quality indicators involving proportions such as the NHIS total family income item nonresponse rate. The ANOM procedure was used to produce the proportion of nonconforming items charts also known as a pchart. Figure 4 shows an example of the proportion of nonconforming items chart. The process average, rational subgroups, and decision limits are shown in Figure 4. The process average is the expected average for the total family income item nonresponse rate and is indicated by a constant straight line. The process average and the values on the y-axis are proportions. The process average was 19 percent. The significance level was set to 0.1. Interviewers were used as the rational subgroups on the x-axis. Each point on the graph is the total family income item nonresponse rate for the interviewer. The total family income item nonresponse rate was outside the decision limits for an interviewer indicating that there were significant differences between their total family income item nonresponse rate and the overall total family income item nonresponse rate of all the interviewers. The proportion of nonconforming items ANOM chart formulas are shown below and can be found at http://support.sas.com/publishing/pubcat/chaps/59050.pdf.

$\bar{p} = \frac{n_1 p_1 + \cdots + n_N p_N}{n_1 + \ldots + n_N}$, where

$\bar{p}$ = the average proportion of nonconforming items across the subgroups
p_i = proportion of nonconforming items in the i^{th} group
n_i = number of items in the i^{th} group
N = number of subgroups

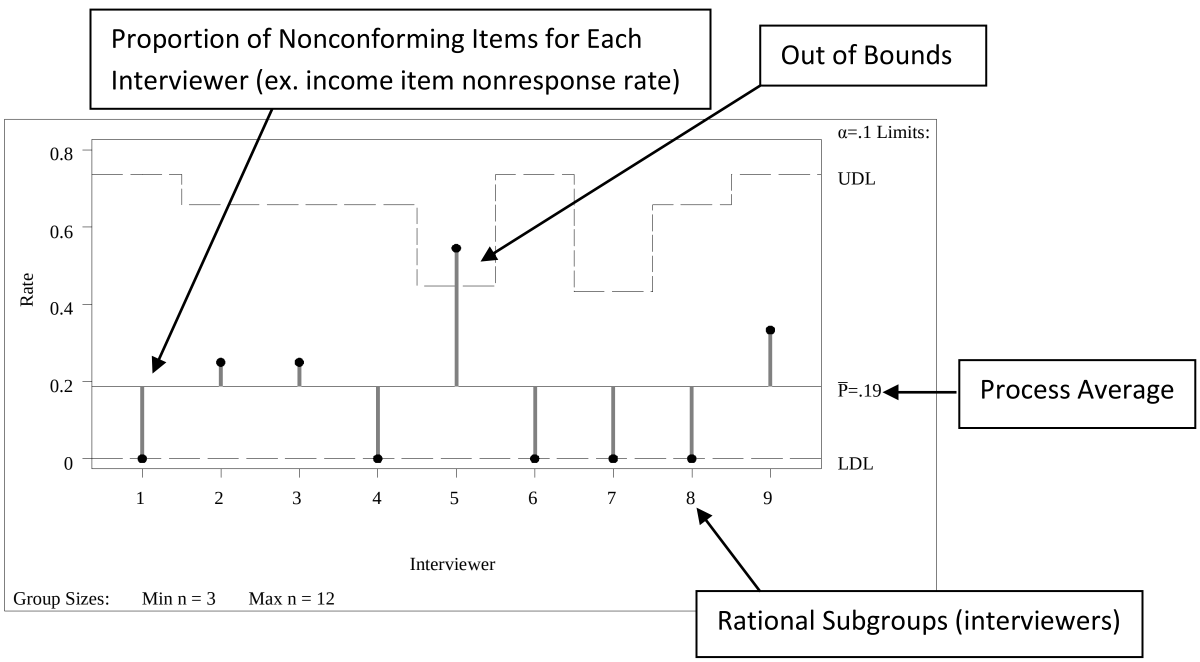
$$Upper\ Decision\ Limit = \min(\bar{p} + h(\alpha; k, n_1, \dots n_k, \infty)\sqrt{\bar{p}(1-\bar{p})}\ \sqrt{\frac{N-n_i}{Nn_i}}, 1)$$

$$Lower\ Decison\ Limit = \max(\bar{p} - h(\alpha; k, n_1, \dots n_k, \infty)\sqrt{\bar{p}(1-\bar{p})}\ \sqrt{\frac{N-n_i}{Nn_i}}, 0)$$

where

$h =$ critical value for ANOM
$\propto =$ significance level
$\mathrm{k} =$ number of subgroups
$n_i =$ sample size of i^{th} group
$N =$ total sample size

Figure 4: Proportion of Nonconforming Items Analysis of Means Chart



3. Statistical Analysis

In this study, we demonstrate how control charts and ANOM charts could be used to determine whether interviewers were collecting data in accordance with survey procedures in a real time setting. These tools could facilitate prompt corrective action that would need to be taken if the data are not being collected properly.

It is important to examine multiple quality indicators when evaluating interviewer performance. The two types of quality indicators in the analysis are those that include don't know and refusal responses, such as total family income item nonresponse rates, and those that exclude don't know and refusal responses such as the average total family income or proportion of "no" responses to a question asking "Have you ever had a pneumonia shot?" For example, if an interviewer has a low item nonresponse rate for the pneumonia shot question but the proportion of "no" responses is significantly different than the other interviewers than this could be an issue.

3.1 Analysis Data

The data used in this paper comes from the completed interviews from January 2008 through December 2010 NHIS. ANOM charts, control charts, and multivariate charts were developed to demonstrate how the charts could be used to improve interviewer performance and data quality in real time. Therefore, to demonstrate this in real time we focused on December 2010 as our hypothetical "current month". The quality indicators used in the analysis are based on asthma, smoking, pneumonia, and total family income questions. The asthma, smoking, and pneumonia questions are asked in the adult section of the data collection instrument. The questions are explained in greater detail when showing the charts for these quality indicators.

The estimates were not weighted in the analysis since we are developing SPC charts in a real time setting. The base weights are available prior to data collection. Subsequent weighting adjustments are not calculated until well after the data collection ends. The other NHIS weighting adjustments include the household nonresponse adjustment, first-stage ratio adjustment, and the second-stage ratio adjustment. The first-ratio adjustment corresponds to persons in non-self-representing PSUs and the second-ratio adjustment includes post-stratification to the U.S. population by age, sex, race/ethnicity (Shaw, Gonzalez, Khare, 2010). The initial value for the weights or inverse of the sampling rate was not used in the SPC charts but may be tested in the future. Although, it may be more beneficial not to include weights to see what is really occurring in the SPC charts without any adjustments to the data.

The quality indicators were calculated for each RO, cluster, and month. The combination RO, cluster, and month was used to create a control chart for each cluster and RO. The quality indicators were also calculated for each RO, cluster, month, and interviewer. The interviewer was included in the combination when creating ANOM charts and interviewer level control charts. The creation of the clusters and the types of control charts will be discussed in the next two sections.

3.2 Creation of the Clusters

Since interviewers are often assigned to work in a specific geographic area, it can be a challenge to separate geographic differences from differences among interviewers. Creating clusters helps to separate natural geographic variability in responses from interviewer effects. Interviewers were grouped into statistical clusters based on housing unit, socioeconomic, and demographic characteristics of the respondents. SPC techniques were used to measure the variation of the process within each regional office and cluster. See Sirkis and Jans, 2011 for similar information on the creation of the clusters.

To compare interviewers who worked on cases that were similar to each other, independent of where those cases were located, census tracts were grouped into clusters based on the characteristics of the tracts. The characteristics of the tracts come from the U.S. Census Bureau 2000 Planning Database (PDB). The PDB contains 100-percent and sample data from the Census 2000 long and short forms. See Bruce and Robinson, 2006 for a description on the variables on the database.

Since clusters were created at the tract level, interviewers can be in more than one cluster if they had cases in more than one tract. Clusters with fewer than five interviewers were combined with similar clusters based on geographic and respondent characteristics. Cases without a census tract identifier were placed into their own cluster.

A variable reduction technique was used to reduce the number of variables in the PDB to a manageable set. This was done to remove redundant statistical information. The procedure is closely related to principal component analysis and can be used as an alternative method for eliminating redundant variables. The procedure divides a set of numeric variables into either disjoint or hierarchical clusters. A large set of variables can be replaced by a single variable of each cluster often resulting in little loss of information. The variable reduction analysis finds groups of variables that are as correlated as possible among themselves and as uncorrelated as possible with variables in other clusters. The first phase is a nearest component sorting phase similar to the nearest centroid sorting algorithms. The second phase involves a search algorithm to determine if assigning each variable to a different cluster would increase the amount of variance explained. If a variable is reassigned during the second phase, the components of the two clusters involved are recomputed before testing the next variable (Nelson, 2001). A variable is selected from each cluster that has a high correlation with its own cluster and a low correlation with other clusters by using the $1-R^2$ ratio. R^2 is the proportion of variance accounted for by the clusters. The formula is below.

$$1\text{-}R^2\ Ratio = \frac{1 - R^2\ own\ cluster}{1 - R^2\ nearest\ cluster}$$

Nine variables were used to group the tracts into interviewer comparison clusters. An interviewer comparison cluster is a group of interviewers who have similar values for these nine variables within an RO. The nine variables were percent Hispanic origin, percent two or more housing units in the structure, percent of the population below poverty, percent Asian, percent vacant housing units, percent of the population under age 18, percent Native Hawaiian and Pacific Islander, percent White, and percent linguistically isolated other language households. Table 1 shows an example of how a variable was selected based on the $1\text{-}R^2$ ratio for cluster 2. Percent of two or more housing units in the structure had the smallest ratio so this variable would be included in grouping the census tracts.

Table 1: Results of Variable Reduction Analysis in Cluster 2

Variable	**Next Closest Cluster R^2**	**Own Cluster R^2**	**$1\text{-}R^2$ Ratio**
Percent 2 or More Housing Units in Structure	0.173976	0.917708	0.099624
Percent of Housing Units That are Not Single Detached or Attached Units	0.247495	0.894621	0.140038
Percent Single Detached or Attached Housing Units In Structure	0.247495	0.894621	0.140038
Percent Renter Occupied Housing Units	0.384327	0.878012	0.198138
Percent 10 or More Housing Units in Structure	0.110268	0.7251	0.30897
Percent Households That are Not Husband/Wife Families	0.332251	0.663047	0.504611
Percent Occupied Housing Units Where Householder Moved Into Unit in 1999-2000	0.061438	0.424589	0.613077

After reducing the number of variables, a second clustering technique was used to group the census tracts into clusters using the *k means* model method. This involved using Euclidean distances so the cluster centers are based on least squares estimation. The cluster centers are the means of the observations assigned to each cluster. The output statistics were the approximate expected overall R^2 and the cubic clustering criterion (CCC). The CCC is a comparative measure of the deviation of the clusters from the distribution expected if the data points were obtained from a uniform distribution (Sarle, 1983). The constant terms were chosen on the basis of extensive simulation results. A value greater than 2 indicates a good clustering algorithm.

$$CCC = \ln\left[\frac{1 - E(R^2)}{1 - R^2}\right] \times \frac{\left(\frac{np}{2}\right)^{0.5}}{(.001 + E(R^2))^{1.2}}$$

where $E(R^2)$ is the expected R^2, and p is an estimate of the dimensionality of the between cluster variation. The constant terms were chosen based on extensive simulation results (Sarle, 1983).

The number of clusters was increased until the R^2 was greater than 0.70, at which point the clustering algorithm stopped. Many additional clusters would need to be added in order to obtain a much higher R^2 than 0.70. The decision was also based on what would be practical for implementation in a survey field environment.

The number of interviewer comparison clusters ranged from three to seven. An additional cluster was added within each RO for cases that could not be assigned a tract. Interviewer comparison clusters were constrained to RO boundaries. Table 2 shows an example of the clustering statistics for five ROs. There was one cluster in a RO that had less than five interviewers and had to be collapsed with another cluster.

Table 2: Example Cluster Level Statistics By Regional Office

Regional Office	Number of Clusters	Number of Clusters Including New Construction	Cubic Clustering Criterion	R^2
1	3	4	14.92	0.70
2	4	5	25.92	0.73
3	6	7	2.37	0.74
4	3	4	9.45	0.70
5	4	5	21.13	0.73

3.3 Statistical Process Control Procedure

The primary objective of the control charts and ANOM charts is to determine whether interviewers are collecting data in accordance with survey procedures. Several steps were implemented to identify those interviewers whose quality indicator was significantly different from the overall quality indicator for the RO and the cluster. Control charts were then constructed for those interviewers. The five steps are as follows:

Step 1: Construct the cluster level control chart within RO including the data from January 2008 to November 2010.
Step 2: Remove the outliers from the cluster level control chart in Step 1.
Step 3: Construct the final cluster level control chart within RO including data from January 2008 to December 2010.
Step 4: Construct the ANOM charts within the RO and cluster for the current month, December 2010.
Step 5: Construct the interviewer level control charts if the quality indicator for the interviewer was outside the decision limits for the ANOM charts for December 2010.

A cluster level control chart is a chart that includes data for one RO and cluster combination. Months were used as the rational subgroup on the x-axis. The months were used as the rational group in the control charts in order to have a high between group variability and low within group variability when special cause variation is present. All interviewers within that RO and cluster combination were included in the charts. Control charts were created for each RO and cluster combination. An interviewer was in more than one cluster level control chart if the interviewer had completed interviews in more than one cluster within the RO. The quality indicator was calculated for each RO, cluster, and month combination. For example, the average total family income was calculated for each RO, cluster, and month combination. Cluster level control charts were produced to remove the outliers from all historical months excluding the hypothetical current month, December 2010. Only the months where there was more than one completed interview were included in the cluster level SPC charts. Removing all outliers except the current month, December 2010, allows us to see if the addition of the current month causes the process to not be in a state of statistical control. This is under the assumption that the interviewers have been retrained due to potential problems in the past months if the process was not in control. The cluster level control chart process average was compared to the interviewer level control chart process average. If the process average for the interviewer is much higher or lower than the process average for the cluster then that interviewer's data should be further explored.

Average and standard deviation control charts were used to plot estimates of total family income. In these charts outliers were removed that were beyond the control limits. If the average total family income for the month was above the control limits then the housing unit or completed interview with the highest total family income was removed. If the average total family income for the month was below the control limits then the housing unit or completed interview with the lowest average total family income was removed. The same method was used for the standard deviation control chart. This was done iteratively until the process was in control. If a housing unit or completed interview was removed for one quality indicator it does not mean that it was removed for another quality indicator since the charts are performing independent of one another.

Proportion of nonconforming items charts were used to plot the proportion of responses in a question or estimate excluding don't knows and refusals, and proportion of don't know and refusal responses for a question or estimate. An example of an estimate is the current asthma estimate, which is a combination of responses from several questions. The proportion of don't know and refusal responses is considered item nonresponse in this research. In these charts, outliers were removed that were beyond the control limits. The interviewers with the highest numerator (ex. "no" responses or item nonresponse) were removed if the proportion was above the control limits. The interviewers with the lowest numerator were removed if the proportion was below the control limits. This was done iteratively until the process was in control.

Typically assignable causes need to be made before removing outliers from the charts. However, for this analysis the purpose was to demonstrate how managers would use the control charts in a production real time setting. In the future a more defined outlier analysis will be developed as well as assigned reasons to the special causes. Once the process was in control in the cluster level control chart, the current month December 2010 was included along with the January 2008 to November 2010 historical data in the control chart. For the average and standard deviation charts, the process average, process standard deviation, and control limits were taken from the cluster level control chart using only the historical data where the process was in control. For the proportion of nonconforming items charts, the process average and control limits were taken from the cluster level control chart using only the historical data where the process was in control.

ANOM charts were constructed for December 2010 for each RO and cluster. An ANOM chart is a chart that includes data for one RO and cluster combination. Interviewers were used as the rational subgroup on the x-axis. All interviewers within that RO and cluster combination were included in the charts. An interviewer was in more than one ANOM chart if the interviewer had completed interviews in more than one cluster within the RO. The quality indicator was calculated for each RO, cluster, interviewer, and month combination. For example, the average total family income was calculated for each RO, cluster, month, and interviewer combination. The purpose was to determine if an interviewer had significant differences for a quality indicator from the overall quality indicator for the RO and cluster.

Finally, interviewer level control charts were constructed for those interviewers within each RO and cluster that had an average or proportion of responses in a category outside the decision limits in the ANOM charts for the current month, December 2010. An interviewer level control chart is a chart that includes data for one interviewer within an RO and cluster combination. Months were used as the rational subgroup on the x-axis. Multiple interviewer level control charts were constructed for an interviewer if they had completed interviews in more than one RO and cluster combination. The quality indicator was calculated for each RO, cluster, interviewer and month combination. For example, the average total family income was calculated for each RO, cluster, interviewer and month combination. Even though the quality indicator was calculated using the same combination as the ANOM charts only data for one interviewer was included in the interviewer level control chart while data for multiple interviewers within an RO and cluster combination were in the ANOM chart. The outliers that were removed from January 2008 to November 2010 in the cluster level control charts were also removed in the interviewer level control charts. Ultimately, this would mean that assignable causes have been established for the points that exceeded the limits for those months and were removed from the control chart. Then the process would be in control using just the historical data. The purpose was to show whether the inclusion of December 2010 impacted the process. The manager can determine if the process is in control once the interviewer has been retrained or whatever corrective action was made to improve the interviewer's performance. Interviewer level control charts were also constructed without removing the outliers. This showed what actually happened from January 2008 to December 2010.

3.4 Average and Standard Deviation Charts – Continuous Quality Indicators

SPC methods were used to evaluate interviewer performance with continuous questions or estimates such as the total family income. This involved using average and standard deviation control charts and ANOM charts displaying the average total family income. Proportion of nonconforming items control charts and ANOM charts were constructed using the total family income item nonresponse rate. It is necessary to examine several different quality indicators to examine an interviewer's performance. An interviewer may enter very few don't know and refusal responses which were included in the item nonresponse rate but may not enter actual total family income values correctly.

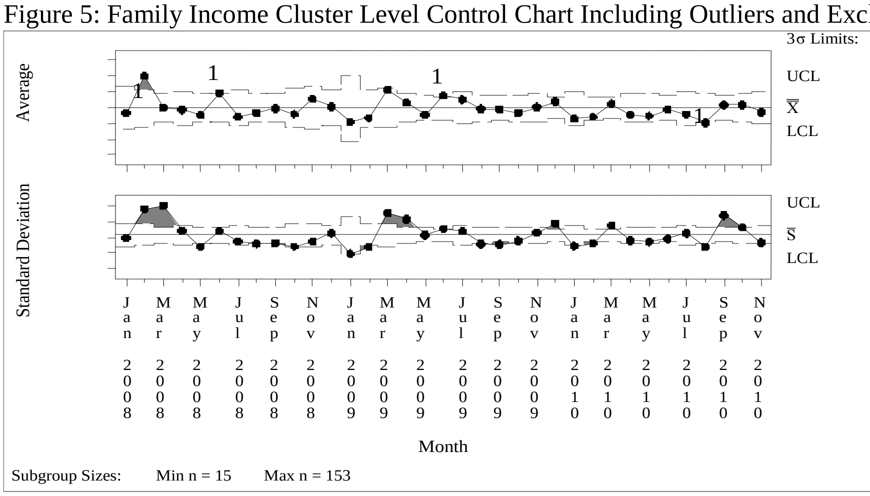
SPC techniques were applied to the following question in the family module: "What is your best estimate of [your total income/the total income of all family members] from all sources, before taxes, in [last calendar year in 4 digit format]?" A benefit to examining questions with continuous outcomes or estimates in general is that it requires the interviewer to key the entire answer from the respondent correctly. For example, an interviewer may leave out or insert extra zeros for an income related question. It may be easier to catch an interviewer who fabricates data with continuous questions or estimates covering a wide range of values over categorical questions or estimates with possibly only two response choices since the distributions may be significantly different from the RO and cluster for continuous quality indicators.

Ultimately the interest should be on catching both one time errors and systematic errors. There can be different types of out of control signals. These can include short term non-systematic errors that might occur as the result of an unusual high income for a housing unit in a given month. Persistent errors can occur that cause an increase in variability such as a wide range of total family income values possibly from fabricating the responses for several months.

The average and standard deviation charts will be shown first for the total family income and then the proportion of nonconforming items charts using the total family income item nonresponse rates will be discussed. The first step of the SPC procedure involved constructing the cluster level control chart within each RO only including the historical data, January 2008 to November 2010. All interviewers within the RO and cluster combination were included in the charts. Don’t know and refusal responses and those completed interviews not in the universe were excluded. Months with only one completed interview were excluded from the cluster level control chart. Figure 5 shows a cluster level control chart within RO for cluster 4 prior to the removal of the outliers. The top part of the chart shows the average total family income for each month and is used to analyze the central tendency of the process. The bottom chart shows the standard deviation for each month and is used to analyze the variability of the process. Each average total family income in the top part of the chart contains a standard deviation in the bottom part of the chart. Some survey practitioners may not want to interpret the average total family income for months such as February 2008 since the standard deviation was outside the control limits for the month. However, a manager may still want to investigate the interviewer’s performance even if the standard deviation is outside the limits but the average is within the limits because there still may be an issue.

At least one month in the chart had 15 completed interviews while at least one month had 153 completed interviews. This is shown at the bottom of the standard deviation chart. All of the months had between 15 and 153 completed interviews. The values on the y-axis and the process average were removed for nondisclosure purposes. There were several months where the average total family income or standard deviation was above the control limits throughout the three-year period shown by the shading and the label “1” in the control chart. The label “1” was not used in the standard deviation chart since there were many months where the standard deviation was outside the control limits. Many of the points were slightly outside the control limits even though they appear to be on the control limit line.

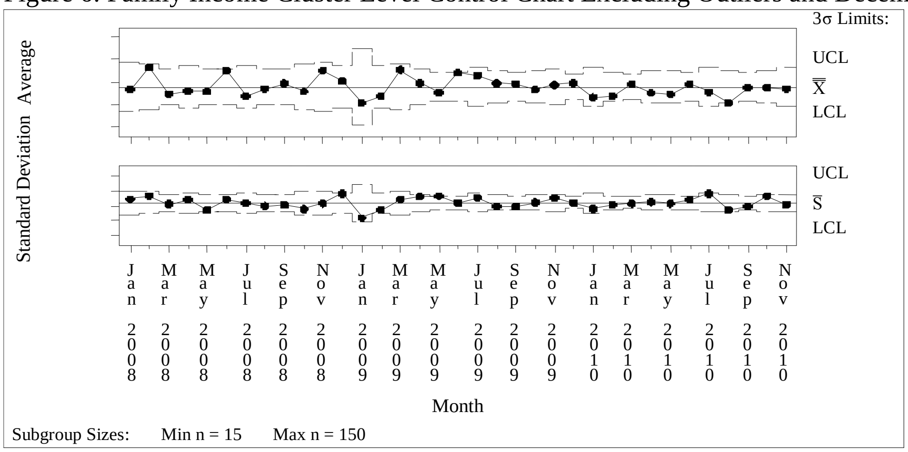
Figure 5: Family Income Cluster Level Control Chart Including Outliers and Excluding December 2010



The second step involved removing the outliers for the months where the average total family income or standard deviation was outside the limits for January 2008 through November 2010 in the cluster level control chart. If the average total family income for the month was above the control limits, then the housing unit or completed interview with the highest total family income was removed. If the average total family income for the month was below the control limits then the housing unit or completed interview with the lowest average total family income was removed. The same method was used for the standard deviation control chart. This was done iteratively until the process was in control. The cluster level control chart in Figure 6 has all outliers removed from January 2008 to November 2010. The process average and control limits have shifted. The month(s) with the maximum number of completed interviews was 150 instead of 153. Typically the survey manager or researcher would have to assign causes to each of the observations that were outside the control limits before removing them from the chart. After the outliers were removed the “current month,” December 2010, was included in the cluster level control chart. The process average, process standard deviation, and control limits was taken from the cluster level control chart with only the January 2008 to November 2010 data where the process was in control.

In the future, assignable causes will be established for any completed interviews removed from the chart as this was just for demonstration purposes. The goal would be to correct the process if it was out of control for the historical data which can include retraining interviewers on the procedures. Then the manager can determine if the interviewer’s performance has been improved when the current month, December 2010 is included in the chart.

Figure 6: Family Income Cluster Level Control Chart Excluding Outliers and December 2010



Next an ANOM chart was used to compare the average total family income for each interviewer to the overall average total family income for December 2010 as shown in Figure 7. The overall average total family income is calculated using the data from all interviewers in the RO and cluster combination for December 2010. The cluster level control chart in Figures 5 and 6 and the ANOM chart use data from the same RO and cluster. Don’t know and refusal responses and those completed interviews not in the universe were excluded. Interviewers with only one completed interview were excluded from the cluster level control chart. At least one interviewer in the chart had two completed interviews while at least one interviewer had ten completed interviews. This is shown at the bottom of the ANOM chart. All of the interviewers had between two and ten completed interviews. The average total family income was outside the decision limits for interviewers 2 and 3. Therefore, interviewer level control charts would be constructed for just these two interviewers for the average total family income for this cluster and RO. The number of completed interviews for interviewer 2 was closer to the minimum for the month while interviewer 3 had closer to the maximum number of completed interviews for the month. A survey researcher may be more concerned about an interviewer that had more completed interviews for the month than another interviewer that had just a few completed interviews. The decision limits are sensitive to sample size and subject to error, so it is important to consider the standard errors when making decision rules that lead to action. The interviewer level control chart will be shown for interviewer 3 since there were more completed interviews than interviewer 2.

Figure 7: Average Family Income ANOM Chart – December 2010

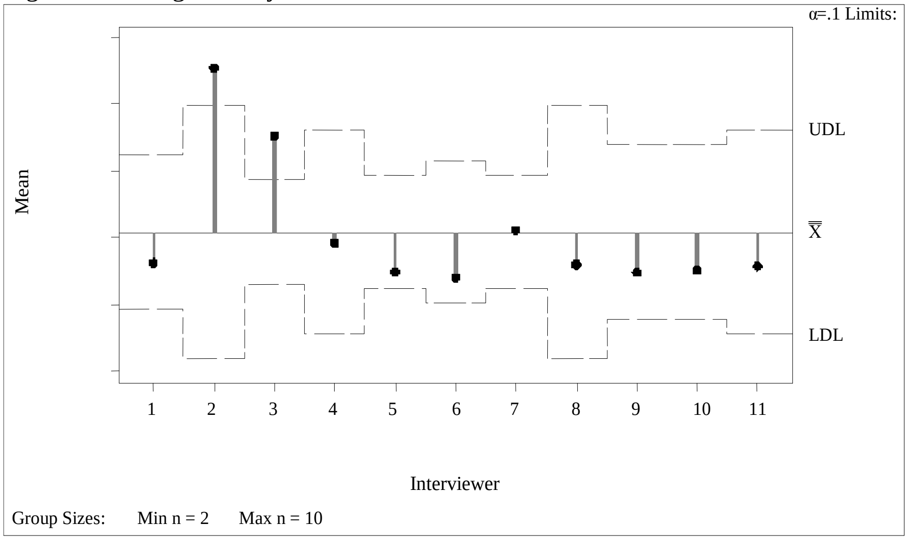


Figure 8 shows an example of an interviewer level control chart for the average total family income for interviewer 3 in cluster 4. The outliers identified in the cluster level control chart were not removed in the interviewer level control chart shown below. The reason to show the chart with the outliers is that a manager can see what actually happened throughout the three-year period for the interviewer. As stated previously, for points where the standard deviation exceeds the limits the corresponding points in the average chart should not be interpreted regardless of whether the mean is outside the control limits. Therefore, some survey practitioners may say not to interpret the average total family income for December 2008, December 2009, and December 2010 since the standard deviation was outside the control limits for these months. A manager may still want to investigate the interviewer's performance even if the standard deviation is outside the limits but the average is within the limits because there still may be an issue. Some of the confidence bounds on some of the time points are wide, suggesting a small sample size. A manager might not be concerned about a high average total family income for that month if it comes from a few completed interviews, and thus would not have a large impact on the overall average total family income for the interviewer or RO. Survey managers need to be cautious about when to intervene with potential problems to avoid identifying problems that are not really problems. However, the manager should still investigate the special causes particularly if they may be due to environmental or societal impacts (such as holidays, or weather related disasters).

Figure 8: Family Income Interviewer Level Control Chart Including Cluster Level Control Chart Outliers

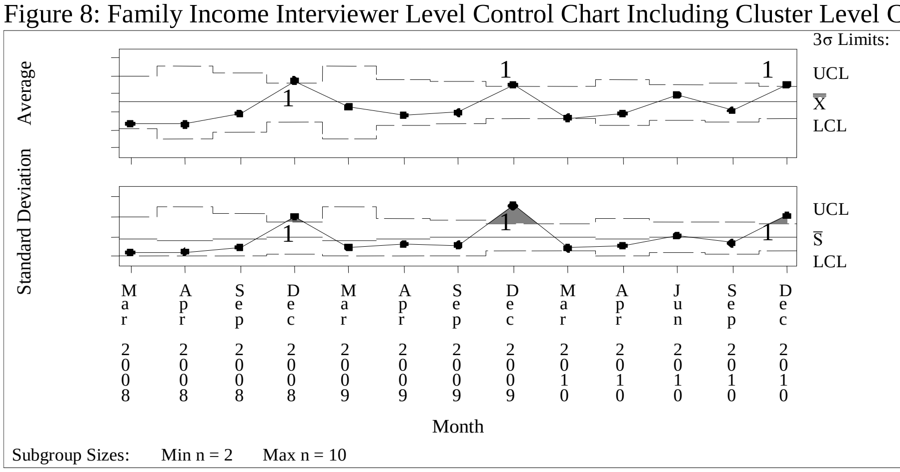
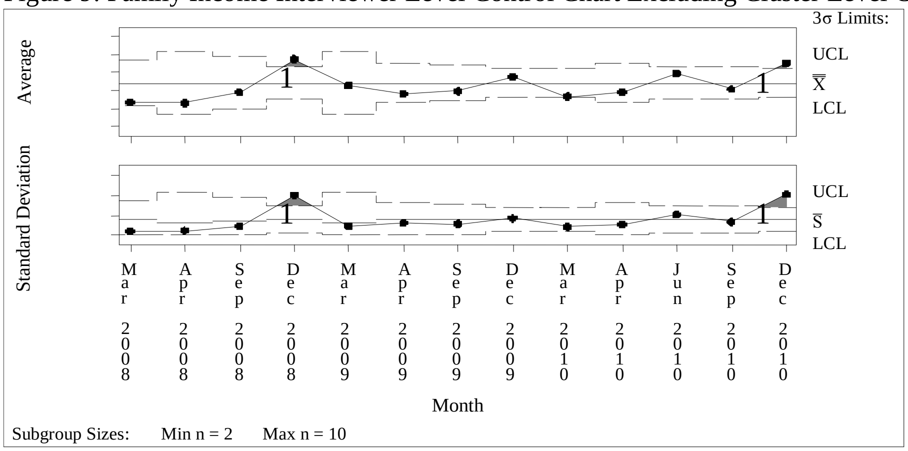


Figure 9 shows an interviewer level control chart for the average total family income for interviewer 3 where the outliers were removed for all months except the current month, December 2010, from the cluster level control chart.

This assumes that assignable causes have been established before removing any outliers. The goal of removing the outliers would be to correct the process when it is out of control for the historical data which can include retraining interviewers on the procedures. Then the December 2010 data would be included in the chart to see if it impacted the process and if the interviewer's performance has improved. The average total family income and standard deviation for December 2009 was no longer outside the control limits. To be cautious, the manager may still want to investigate the interviewer's performance in December 2008 and 2010 even though both the average total family income and standard deviation were outside the control limits.

Figure 9: Family Income Interviewer Level Control Chart Excluding Cluster Level Control Chart Outliers



SPC charts should also be developed showing the total family income item nonresponse rates for comparison purposes to the previous charts showing the average and standard deviation for the total family income. Therefore, ANOM charts were constructed for the total family income question, but this time the proportion of don't know and refusal responses was calculated instead of the average total family income. The numerator included don't know and refusal responses. The denominator excluded those completed interviews not in the universe. It is important to examine the total family income item nonresponse rate as well as the actual income values. The interviewer's estimate for average total family income could be within the limits, but the total family income item nonresponse rate could be much higher or lower than the overall total family income item nonresponse rate for the RO and cluster. There were no interviewers that had their total family income item nonresponse rate outside the decision limits for December 2010 in cluster 4. When the interviewer level control chart was examined for interviewer 3, the process was in control and the interviewer had a low total family income item nonresponse rate for most of the months.

The results from the ANOM charts indicate that the average total family income for interviewer 3 was significantly different from the average total family income for the cluster and RO. In the interviewer level control chart the average and standard deviation was outside the control limits for interviewer 3. However, interviewer 3 did not have their total family income item nonresponse rate significantly different from the overall total family income item nonresponse rate for the same RO and cluster for December 2010 in the ANOM chart. This illustrates that it is necessary to examine both quality indicators that include don't knows and refusals and those quality indicators that do not include those values. There needs to be additional exploration using the interviewer's data besides just examining the ANOM charts and control charts to determine what actions should be taken to possibly improve the interviewer's performance.

3.5 Proportion of Nonconforming Items Charts – Estimates Based on Proportions Quality Indicators

SPC methods were used to evaluate interviewer performance collecting data that is used to estimate proportions. This involves using proportion of nonconforming items charts and ANOM charts displaying quality indicators such as the proportion of "no" responses for the current asthma estimate and the proportion of "not at all" responses to a question about smoking. These questions will be described in more detail in the section. Proportion of nonconforming items control charts and ANOM charts will show item nonresponse rates for each of the quality indicators such as the current asthma item nonresponse rate. It is necessary to examine several different quality indicators to examine an interviewer's performance. An interviewer may enter very few don't know and refusal responses which are included in the item nonresponse rate but may have a distribution of responses for a category other than don't know and refusal responses that is very different from other interviewers in the same cluster for a RO.

There are many advantages to exploring the quality indicators that exclude don't know and refusal responses. For example, if an interviewer has a high proportion of "no" responses the interviewer may not be asking the question. However, if the interviewer has a high proportion of "yes" responses then it is possible that the interviewer consistently enters the first response to the question (if yes is the first response in order), or reads the question in a way that elicits "yes" responses from respondents, even if it is not the "true" response. Several questions may involve skip patterns meaning that if a respondent says "no" to a question the interviewer will not ask the respondent the next series of questions. The interviewer may save time by entering "no" to a question that involves asking additional questions. A benefit to analyzing proportions for questions that have multiple responses categories (ex. high, medium, low) is to ensure that the interviewers are reading all responses to the question and not just selecting the first value in the category. The SPC tools will not completely verify these assumptions but if the interviewer's proportion is significantly different from the overall proportion in the RO and cluster, it could be an indication of these occurrences or other possible issues that should be explored.

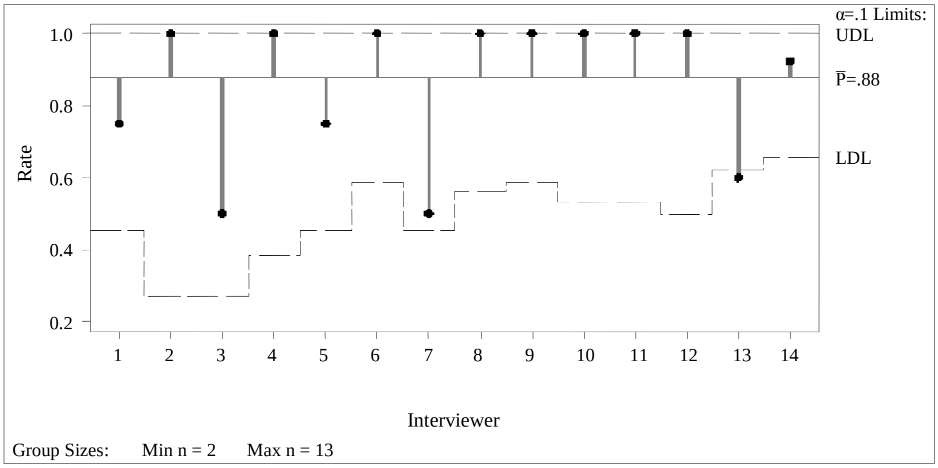
Item nonresponse rates can indicate problems with data quality. High item nonresponse can introduce bias into estimates, and reduce analyzable sample sizes. However, an extremely low item nonresponse rate can be indicative of potential interviewer falsification. An interviewer can influence the respondent to provide an answer that would not be included in the item nonresponse rate (ex. don't knows and refusals) although the respondent may really not know the answer or not feel comfortable answering the question. Examining the item nonresponse can help determine if the interviewer is asking the question correctly. Interviewers might not ask the question correctly to encourage the respondent to continue the interview and feel at ease with the questions, or simply to complete the interview quickly.

The proportion of nonconforming items charts will first be shown for quality indicators that exclude don't know and refusal responses and then the proportion of nonconforming items charts using the item nonresponse rates or quality indicators that include don't know and refusal responses will be discussed.

SPC techniques were applied to the proportion of "no" responses in the current asthma estimate. The current asthma estimate is based on responses from the following three questions: "Have you EVER been told by a doctor or other health professional that you had asthma?", "Do you still have Asthma", and "During the past 12 months, have you had an episode of asthma or an asthma attack?" These three questions are the first three questions in a series about asthma. A benefit to examining this estimate is that there are several questions (20 total on asthma in the sample adult section) that will not be asked if the respondent says "no" to any of these questions. Interviewers may save time if they enter "no" to any of these questions.

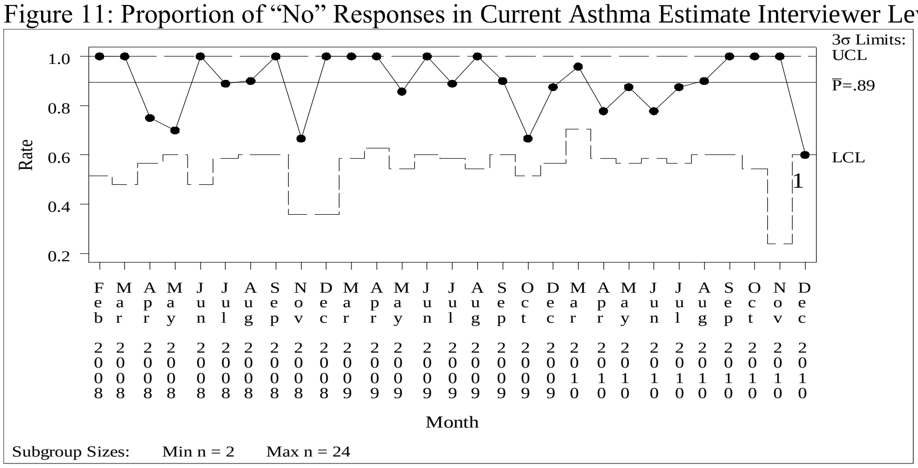
An ANOM chart compared the proportion of "no" responses in the "current month" (December 2010) for each interviewer to the overall proportion of "no" responses for all interviewers in the "current month". The denominator only included "yes" and "no" responses. Don't knows, refusals, and those completed interviews not in the estimate universe were excluded. Interviewers with only one completed interview were excluded from the chart. Figure 10 shows an example of an ANOM chart using the current asthma estimate for one RO and cluster. At least one interviewer in the chart had two completed interviews while at least one interviewer had thirteen completed interviews. This is shown the bottom of the ANOM chart. All of the interviewers had between two and thirteen completed interviews. It may not be as much of a concern for a manager if the percentage of "no" responses was 100% and the interviewer had a small number of completed interviews. The process average was 88 percent. Ultimately, this means that 88 percent of the respondents in the RO and cluster do not currently have asthma. The number of completed interviews for interviewer 13 that had their proportion of "no" responses outside the decision limits was closer to the maximum number of completed interviews for the month. Therefore, an interviewer level control chart would be constructed for this interviewer for the proportion of "no" responses for the current asthma estimate. In this case, interviewer 13 does not have a high proportion of "no" responses but the proportion of "no" responses was significantly different from the overall proportion of "no" responses for the RO and cluster. It seems as though a high percentage of “no” responses may be acceptable for this question. The decision limits are sensitive to sample size, so it is important to consider the decision rules that lead to action.

Figure 10: Cluster Level Proportion of “No” Responses for Current Asthma Estimate ANOM Chart – December 2010



When interviewers were isolated, we needed to examine their performance history separately to determine if corrective action is necessary. Figure 11 shows an example of the proportion of “no” responses for each month for cluster 3 and interviewer 13. The process was not in control since the proportion of “no” responses was outside the control limits for December 2010 shown by the label “1”. The process average was 89 percent for the interviewer, similar to the 92 percent in the cluster level control chart including all months. Months with only one completed interview were excluded from the interviewer level control charts. There were no outlier cases in the cluster level control chart for this interviewer. Some of the confidence bounds on some of the time points are wide, suggesting a small case load for some months. A manager might not be alarmed about a high proportion of “no” responses for months with few completed interviews. Survey managers need to be careful about when to intervene with potential problems to avoid misidentifying problems.

Figure 11: Proportion of “No” Responses in Current Asthma Estimate Interviewer Level Control Chart



SPC charts should also be developed showing the current asthma item nonresponse rates for comparison purposes to the previous charts showing the proportion of “no” responses to the current asthma estimate. Therefore, ANOM charts were constructed for the current asthma estimate, but this time the proportion of don’t know and refusal responses was calculated instead of the proportion of “no” responses to the current asthma estimate. Each interviewer in the RO and cluster shown in the ANOM charts for the proportion of “no” responses for the current asthma estimate had a zero item nonresponse rate for December 2010. An interviewer can have a low or zero item nonresponse rate for a question but have a proportion of responses in a category other than don't know and refusals significantly different from the overall proportion for the RO and cluster.

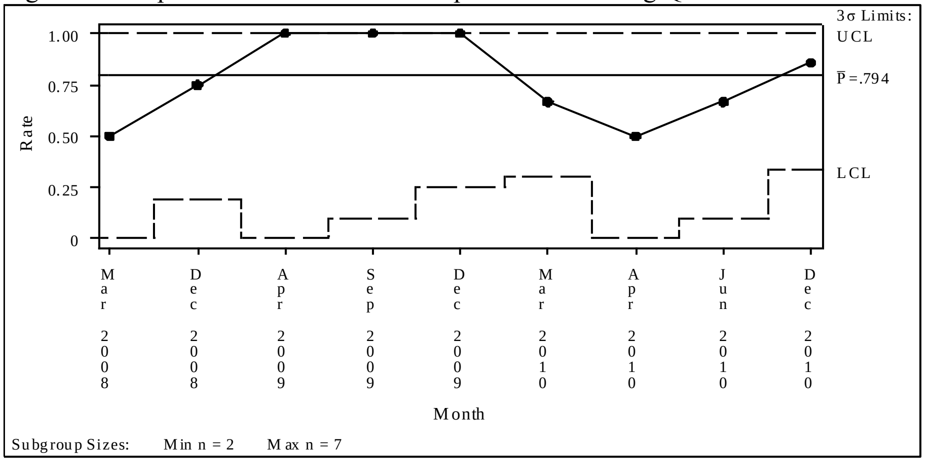
The results from the ANOM charts indicated that the proportion of "no" responses for the current asthma estimate for interviewer 13 were significantly different from the proportion of "no" responses for the current asthma estimate for the cluster and RO. In the interviewer level control chart the process was not in control for the proportion of "no" responses for the current asthma estimate for interviewer 13. However, interviewer 13 did not have their current asthma item nonresponse rate significantly different from the overall current asthma item nonresponse rate for the same RO and cluster for December 2010 in the ANOM chart. This demonstrates how ANOM charts for both estimates and item nonresponse rates can be used to explore interviewer performance. There needs to be additional exploration into the interviewer's data besides just looking at the ANOM and control charts to determine if any intervention is necessary.

A survey manager should also determine if there are identifiable patterns in the control charts besides just identifying quality indicators that are outside the control limits. Patterns in the process can be an indication that the interviewer is either not following interviewing procedures or there is potential for falsification. There are more formal methods for identifying patterns such as the Western Electric Rules. An example of the rules includes 6 points in a row steadily increasing or decreasing or 1 point beyond the 3 sigma limits. It is up to the manager to determine which Western Electric Rules to use. The following example using a question about smoking will show an interviewer level proportion of nonconforming items chart that does not have the quality indicator outside the control limits, but it seems there may be a possible trend that exists.

The SPC techniques were applied to the following question: "Do you NOW smoke cigarettes every day, some days or not at all?" The answer categories include "every day" (1), "some days" (2), "not at all" (3). Don't know and refusal responses and those completed interviews not in the universe were excluded. A benefit to examining this question, especially the last category is to ensure that interviewers are reading all responses to the question and not just selecting the first value in the category. The SPC tools will not completely verify this assumption but if the interviewer's proportion is significantly different from the overall proportion in the RO and cluster it could be an indication of this problem.

Figure 12 shows an example of a control chart for the proportion of "not at all" responses for each month for interviewer 3. The process is in control for cluster 4 and interviewer 3. However, there seems to be a pattern across the three years. The purpose of using SPC charts is to not just to determine if the quality indicator was outside the limits but to also see if there are any identifiable patterns in the process. The process average was 79 percent for the interviewer compared to 53 percent in the cluster level control charts. It is important to note that the interviewer is being compared to themselves in the chart so the process can be in control if they are consistent, even if their process average is much higher or lower than the RO and clusters process average. Therefore, this interviewer should probably be investigated to determine why there are identifiable patterns and a higher process average than the RO and cluster. As stated previously, the Western Electric Rules provides a set of guidelines for discovering systematic patterns. It is possible that the changes in the proportion of "not at all" responses in the smoking question or for that matter the other categories could be due to specific changes over time rather than interviewer errors. In the future this could be analyzed by adding time series models. A series or process can be influenced by seasonal factors that happen on a regular recurring basis. Therefore, a seasonal adjustment could be made to the data before implementing the SPC techniques. This could be accomplished with time series models. Methods can include developing Autoregressive Integrated Moving Average (ARIMA) models, which fit time series data to predict futures points in the series, and involves forecasting. Another method is to forecast time series by using exponential smoothing in which smoothing weights are optimized (Ramasubramanian, 2012). The application of time series models to SPC for this research will be explored in the future.

Figure 12: Proportion of "Not At All" Responses in Smoking Question Interviewer Level Control Chart



ANOM charts were constructed using the proportion of don't know and refusal responses instead of the proportion of "not at all" responses. Each interviewer in the RO and cluster shown in the chart above had a zero item nonresponse rate for December 2010. An interviewer can have a low or zero item nonresponse rate for a question but have a proportion of responses in a category other than don't know and refusals significantly different from the overall proportion for the RO and cluster.

The results from the ANOM charts indicate that the proportion of "not at all" responses for interviewer 3 was significantly different from the proportion of "not at all" responses for the cluster and RO. In the interviewer level control chart, the process was in control for the proportion of "not at all" responses for interviewer 3. However, interviewer 3's proportion of don't know and refusal responses was not significantly different from the overall proportion of don't know and refusal responses for the same RO and cluster for December 2010 in the ANOM chart. The survey manger should possibly explore the interviewer's data to make a determination of whether to follow-up with the interviewer since there may be identifiable patterns in their process even though the quality indicator was not outside the control limits for any month.

3.6 Estimates Based on Proportions – High Item Nonresponse Rates

High item nonresponse rates can be indicative of problems with interviewer performance and data quality. This can introduce bias into the estimates and reduce analyzable sample sizes. Lower item nonresponse rates are preferable. However, an extremely low item nonresponse rate can be indicative of potential interviewer falsification. The respondent may choose not to answer the question due to the tone of the interviewer or if the interviewer asked the question incorrectly. A respondent may not understand the question if the interviewer reads it incorrectly. Other reasons for a high item nonresponse rate not necessarily due to interviewer errors could be a lack of interest on the respondent's part or fatigue with the interview. The respondent may say don't know to get through with the topic or section. In addition, a respondent may say don't know if he or she feels it may be a sensitive question.

The previous examples had a small proportion of don't know and refusal responses for the ROs and clusters. Therefore, interviewer level control charts were not constructed for these quality indicators. However, there are interviewers that will have a high or abnormal item nonresponse compared to the overall proportion for the RO and cluster. It is still important to examine multiple types of quality indicators for an interviewer to determine if the interviewer may also have significant differences for quality indicators that do not include don't know and refusal responses from the RO and cluster. In this section we will focus on the total family income question and will start with the proportion of nonconforming items charts displaying the total family income item nonresponse rate.

An ANOM chart compared the proportion of don't knows and refusal responses for each interviewer to the overall proportion of don't knows and refusals for the current month, December 2010. The denominator excluded completed interviews that were not in the universe. Figure 13 shows an example of an ANOM chart for an RO and cluster for the total family income question. The process average was 38.5 percent. The proportion of don't know and refusal responses for interviewer 5 was significantly different from the overall proportion of don't know and refusal responses for the RO and cluster. Therefore, an interviewer level control chart would be constructed for this interviewer for the total family income item nonresponse rate. The ANOM chart shows that interviewer 5 had a much higher item nonresponse rate than the rest of the interviewers in the RO and cluster.

Figure 13: Family Income Item Nonresponse Rate ANOM Chart – December 2010

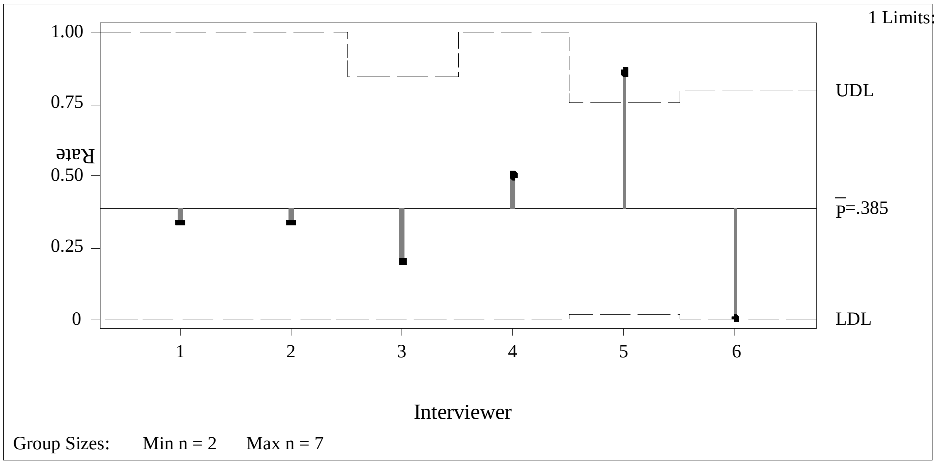
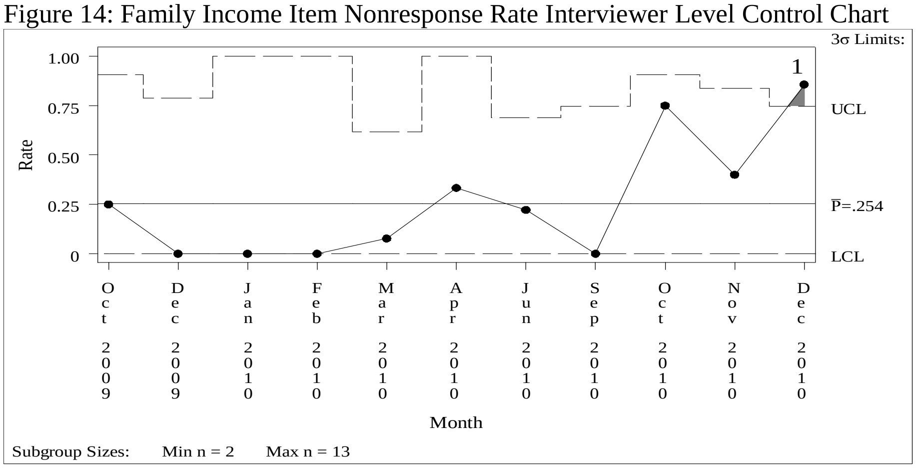


Figure 14 shows an example of a control chart for the proportion of don't knows and refusals for each month for cluster 4 and interviewer 5, with the outliers removed for all months excluding the current month in the cluster level control chart. The process was not in control since the proportion of don't know and refusal responses was outside the control limits for December 2010 (shown by the label "1" and the shading). The process average was 25 percent and slightly different than the cluster level control chart process average of 33 percent with outliers removed. The proportion seems to be somewhat inconsistent starting around March 2010, increases drastically in October 2010 and then increases more in December 2010. The confidence bounds for some of the points are wide, signifying a small number of completed interviews. A manager might not be concerned about a high total family income item nonresponse rate for that month if it comes from a few completed interviews and therefore would not have a large impact on the overall total family income item nonresponse rate for the interviewer or RO. Survey managers should be mindful of the sample size as to only intervene and inform the interviewer when it is necessary to take action. The manager should probably follow-up with the interviewer.

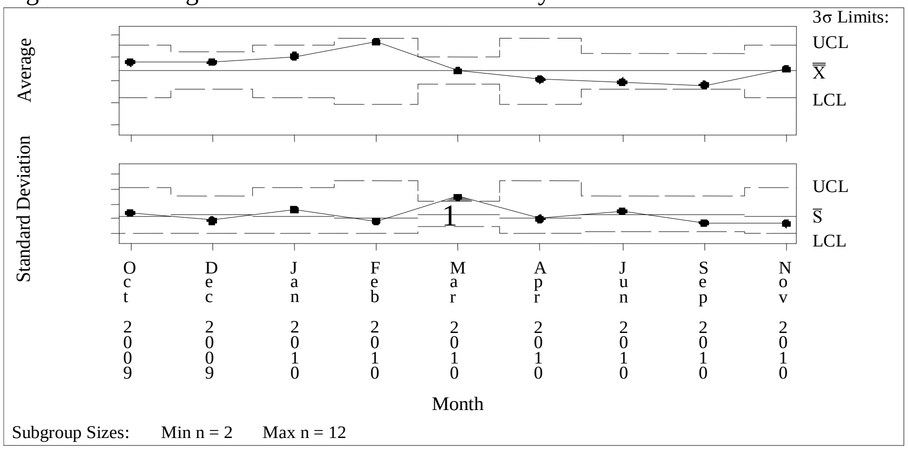
Figure 14: Family Income Item Nonresponse Rate Interviewer Level Control Chart



ANOM charts were constructed using the income values for the total family income question. Interviewer 5 was not in the ANOM chart for the RO and cluster for the average income since interviewer 5 did not have more than one completed interview that was not a don't know or refusal for December 2010. Don't knows, refusals, and those cases not in the universe were excluded.

Figure 15 shows an example of a control chart for the average total family income for interviewer 5. The interviewer did not have more than one completed interview that was not a don't know or refusal response for the total family income in December 2010 so the month was not included in the chart. There were no outliers removed for this interviewer in the cluster level control chart. There was only one month March 2010, where the standard deviation was outside the control limits.

Figure 15: Average and Standard Deviation Family Income Interviewer Level Control Chart



The results from the ANOM charts indicate that the total family income item nonresponse rate for interviewer 5 for the cluster and RO was significantly different from the total family income item nonresponse rate for the cluster and RO. In the interviewer level control chart the process was not in control for the total family income item nonresponse rate for interviewer 5. However, interviewer 5 only had one month where the standard deviation was outside the control limits in the interviewer level control chart. The interviewer did not have any months where the average was outside the control limits in the interviewer level control chart. There probably needs to be some corrective action here since there seems to be a large amount of don't know and refusals for the total family income question. However, additional exploration of data would be helpful for the managers in order to take the correct course of action.

4. Implementation and Challenges

Simply determining which data should be monitored is not a simple task. There are a vast number of questions on the NHIS questionnaire and hundreds of interviewers. Multivariate charts can help solve this issue by combining responses from two or more questions into one measure. Univariate charts would only need to be constructed for each question if the process was not in control in the univariate chart. There still can be plenty of interviewer level control charts to review even though the control charts are only constructed for those interviewers whose quality indicator was outside the decision limits in the ANOM charts. The manager may also not have the resources to follow-up with those interviewers to take corrective action to improve their interviewer performance. A large task will be to develop a method to determine interviewers who should be investigated given the available amount of resources for follow-up purposes. For example, a manager may want to follow-up with those interviewers who have the most quality indicators where the process was not in control in the interviewer level control chart.

It is extremely challenging to choose the appropriate quality indicators and those interviewers whose data should be further explored. This is one of the reasons that the early release program was developed. The early release program publishes selected estimates from the NHIS on an expedited schedule. Managers, researchers, and sponsors need to decide which quality indicators should be used in the SPC techniques to decide which interviewers require corrective action.

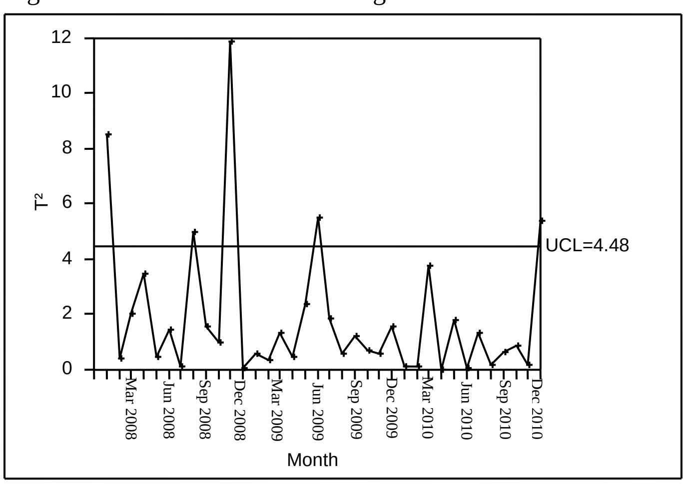
4.1 Choosing Quality Indicators

A survey manager, supervisor, or program analyst may not have time to examine an ANOM chart or control chart for every question in the questionnaire especially if charts are produced for the proportion of responses in each categorical question excluding don't know and refusals, proportion of don't know and refusal responses in each question, and average values for continuous questions. Multivariate charts can help solve this issue by reducing the number of charts that a survey researcher or practitioner would need to review each month or period of interest. These charts combine responses from two or more questions into one measure using the Hotelling's T distribution (Ryan, 1989) when the responses are correlated with one another. The benefit to using a multivariate chart is that if the process is in a state of statistical control then univariate charts would not need to be constructed for each quality indicator. A multivariate chart was constructed for the following questions since the responses were correlated with one another.

> "How often do you do VIGOROUS leisure-time physical activities for AT LEAST 10 MINUTES that cause HEAVY sweating or LARGE increases in breathing or heart rate?"
>
> "How often do you do LEISURE-TIME physical activities specifically designed to STRENGTHEN your muscles such as lifting weights or doing calisthenics? (Include all such activities even if you have mentioned them before.)"

The multivariate chart should only be constructed using questions that have the same total responses. To meet this requirement, a simple random sample without replacement was taken in each month to obtain the same number of completed interviews in each month. The month that had the smallest number of completed interviews determined the number of observations selected for each month. The number of completed interviews was around 400 for each month even though the maximum number of completed interviews for several months was over 2000. The multivariate chart only contains an upper control limit. The process was not in control since the combined measure was above the upper control limit. Therefore, the univariate charts for both the "vigorous leisure–time physical activity" question and "strengthen your muscles" question would need to be constructed. Figure 16 shows the multivariate chart for the combined measure for the vigorous leisure time and strengthen muscles activities questions.

Figure 16: Multivariate Chart – Vigorous Leisure Time and Strengthen Muscles Activities Questions



However, if the process were in control in the multivariate charts, then the survey practitioner or researcher would not have to produce as many charts. The multivariate charts can be very useful when having several (even more than two) questions that have responses that are correlated with one another. There are also other tools that can be used to summarize the multitude of data such as Chernoff charts that display multivariable data by in the shape of a human face (Dickinson, 2001).

4.2 Selecting Interviewers to Follow Up

A method or plan to prioritize interviewers should be developed since a survey manager, survey researcher, or survey methodologist may not have the time to follow-up with every interviewer whose process is not in control in the control chart in their RO and cluster. One method is to investigate the interviewers that have the most quality indicators where the process was not in control or were outside the decision limits in the ANOM chart. Another method could be to follow-up with the interviewers that are most outside the control limits for the current month in the control chart. An interviewer may only have one quality indicator where the process is not in control but it could be significantly outside the control limits in the control charts.

The survey manager, survey researcher, or survey methodologist should examine additional information to determine the appropriate action needed to improve interviewer performance. In the previous sections ANOM charts were constructed to compare the quality indicators for each interviewer to the overall quality indicator for the cluster and RO for December 2010. It may be useful to determine if the interviewer also had significant differences for previous months. Managers may prefer to follow-up with interviewers that had several months in a short period of time where their quality indicator was significantly different from the cluster and RO. A manager may want to hold off and not follow-up with an interviewer that only had one month with a quality indicator that was significantly different from the cluster and RO. A manager may also be concerned if the interviewer had significant differences for several clusters.

The example below uses the average total family income and total family income item nonresponse rate for interviewer 3 in section 3.4. Table 3 shows the number of months that the interviewer had completed interviews and the number of months that the average total family income for interviewer 3 was outside the decision limits in the ANOM chart by cluster and year. The interviewer had four months in 2008 and three months in 2010 where the average total family income was outside the decision limits in the five clusters with completed interviews. Surprisingly, the interviewer did not have any months in 2009 where the average total family income was outside the decision limits. The interviewer worked less months in 2010 than the other two years. This table indicates that the interviewer may not be following procedures for several months in 2008 and 2010. It is unclear why there are no significant differences in 2009 and why it was different from the other two years.

Table 3: Average Family Income - Number of Months Outside of Limits in ANOM Chart and Months Worked (Interviewer 3)

Year	**2008**		**2009**		**2010**	
Cluster	**Outside Decision Limits**	**Worked**	**Outside Decision Limits**	**Worked**	**Outside Decision Limits**	**Worked**
1	2	3	0	3	0	3
2	0	2	0	0	0	1
3	1	10	0	11	1	3
4	1	4	0	5	2	5
5	0	3	0	2	0	3

A manager may want to explore additional data and other quality indicators to determine whether to take corrective action for this interviewer since there were significant differences between their average total family income and the overall average total family income for the cluster and RO. Therefore, ANOM charts were constructed for several months using the total family income item nonresponse rate as opposed to the average total family income for interviewer 3. Table 4 shows the number of months that the average total family income and total family income item nonresponse rate was outside the decision limits by cluster and year for interviewer 3. This is similar to Table 3 but includes the total family income item nonresponse rate. Even though interviewer 3 did not have an item nonresponse rate outside the decision limits in December 2010, the item nonresponse rate was outside the decision limits for three months in 2008 and two months in 2009. Therefore, it seems as though interviewer 3 had a significantly different average total family income and total family income item nonresponse rate over the course of the three-year period from the RO and cluster. Both of these tables indicate that the manager should possibly follow-up with the interviewer to determine if interviewing procedures are being followed. This could involve retraining the interviewer or another course of action.

Table 4: Average Family Income and Family Income Item Nonresponse Rate
- Number of Months Outside of Limits in the ANOM Chart (Interviewer 3, NR= Nonresponse)

Year	**2008**		**2009**		**2010**	
Cluster	**Average Outside Decision Limits**	**Item NR Outside Decision Limits**	**Average Outside Decision Limits**	**Item NR Outside Decision Limits**	**Average Outside Decision Limits**	**Item NR Outside Decision Limits**
1	2	0	0	0	0	0
2	0	0	0	0	0	0
3	1	3	0	2	1	0
4	1	0	0	0	2	0
5	0	0	0	0	0	0

The survey manager or researcher should know approximately how many completed interviews are in each cluster for the interviewer. A control chart will not be as reliable and may have high a coefficient of variation (cv's) for the clusters with a small number of completed interviews. It may also be an issue if the majority of cases are in the cluster where the process is not in control for the interviewer. Although a manager may also be concerned if the process is not in control in a cluster with a small amount of cases. Table 5 shows the number of complete interviews and total cases for the interviewer discussed in the example using the total family income question in section 3.4. For nondisclosure purposes, the cells in the table do not contain the actual number of completed interviews values. The ANOM chart and interviewer level control charts that were shown were in cluster 4. The average and standard deviation for the total family income was outside the control limits for the interviewer in the interviewer level control chart. The interviewer had the majority of completed interviews in cluster 4 for December 2010 even though cluster 3 had the majority of completed interviews for January 2008 to November 2010. It may only make sense to produce an interviewer level control chart in cluster 4 since the other clusters do not have that many completed interviews and will not be as reliable.

Table 5: Family Income Item Nonresponse Rate – Number of Cases (Interviewer 3)

Cluster	**Number of Cases January 2008 to November 2010**		**Number of Cases December 2010**	
	Complete	**Total**	**Complete**	**Total**
1	10-49	50-99	0-9	0-9
2	0-9	10-49	0-9	0-9
3	100-299	100-299	0-9	0-9
4	50-99	100-299	10-49	50-99
5	10-49	10-49	0-9	0-9

It is informative to examine whether the interviewer had other quality indicators that were outside the decision limits in the ANOM chart and where the process was not in control in the interviewer level control chart. The SPC techniques were constructed using the proportion of either the "yes" or "no" responses for the current asthma estimate and pneumonia shot estimate for the same interviewer in a RO and cluster. The charts were also constructed using the proportion of "not at all" responses for the current smoking estimate for the same interviewer in a RO and cluster. This is using the same interviewer 13 found in the current asthma estimate example in section 3.5. The charts were also produced for the average total family income and standard deviation and the proportion of don't know and refusal responses for the family income question, current asthma estimate, pneumonia shot estimate and current smoking estimate for the same interviewer. Tables 6 and 7 show the quality indicators that were outside the decision limits in the ANOM chart or the control limits in the control chart. The tables also show the average differences between the quality indicator and the control limits if the quality indicator was outside the control limits in the interviewer level control chart. A "yes" entry in the ANOM chart column indicates that the quality indicator was outside the decision limits in the ANOM chart for December 2010. A "yes" entry in the control chart column indicates that the quality indicator was outside the control limits in the interviewer level control chart. The proportion of don't know and refusal responses for the total family income question, proportion of "yes" responses in the pneumonia shot question, proportion of don't know and refusal responses for the pneumonia shot question, and the proportion of "no" responses in the current asthma estimate were outside the decision limits in December 2010 or outside the control limits in the interviewer level control chart.

The current asthma estimate was the only quality indicator that was outside the decision limits for December 2010 and the interviewer's process was not in control. The proportion of don't know and refusal responses for the total family income question, proportion of "yes" responses for the pneumonia shot question and proportion of don't know and refusal responses for the pneumonia shot question were not outside the decision limits for December 2010 but the process was not in a state of statistical control in the interviewer level control charts for certain clusters. Other months could have caused the process not to be in control in the interviewer level control chart. The quality indicator could have been outside the decision limits for a month other than December 2010. Except for the proportion of "yes" responses in the pneumonia shot question for cluster 4 where the difference was 0.404, the difference between the control limits and the quality indicator when it was outside the control limits was extremely small.

The results from these tables show that the difference between the control limits and quality indicator was small when the quality indicator was outside the control limits in the interviewer level control charts for interviewer 13 indicating that this may not be a huge concern. In addition, the interviewer only had one quality indicator (so far) where there were significant differences from the overall quality indicator from the RO and cluster for December 2010. Even though there were other quality indicators besides the current asthma estimate where the process was not in control it does not seem to concern the current month, which is of primary interest. The manager or supervisor should examine additional quality indicators besides the ones in the table to determine if any corrective actions should be taken.

A future task would be to simply summarize all variables with an "out of control" rate. This would involve calculating the number of out of control questions divided by the total number of questions asked since the primary goal is to make a decision regarding the overall performance of each interviewer to maintain that certain level of data quality. Also a manager or supervisor would only have to examine one chart for an interviewer rather than examining several charts which would involve using significantly fewer resources.

Table 6: Current Asthma Estimate and Family Income Item Nonresponse Rate Quality Indicators – Summary of Chart Information

	Current Asthma (Binary "Yes")				Family Income Item Nonresponse Rate			
Cluster	**ANOM Chart**	**Control Chart**	**Number Months Outside Control Chart**	**Average Difference Outside Control Limits (Upper)**	**ANOM Chart**	**Control Chart**	**Number Months Outside Control Chart**	**Average Difference Outside Control Limits (Upper)**
2	No	No	0	0	No	No	0	0
3	Yes	Yes	1	0.001	No	Yes	1	0.074
4	No	No	0	0	No	No	0	0

Table 7: Pneumonia Shot Binary Question and Pneumonia Shot Item Nonresponse Rate Quality Indicators - Summary of Chart Information

	Pneumonia (Binary "Yes)				Pneumonia Item Nonresponse Rate			
Cluster	**ANOM Chart**	**Control Chart**	**Number Months Outside Control Chart**	**Average Difference Outside Control Limits (Upper)**	**ANOM Chart**	**Control Chart**	**Number Months Outside Control Chart**	**Average Difference Outside Control Limits (Upper)**
2	No	No	0	0	No	No	0	0
3	No	Yes	1	0.047	No	Yes	1	0.019
4	No	Yes	1	0.404	No	No	0	0

5. Limitations

Many of the interviewer level control charts contained a small number of completed interviews for each month resulting in a coefficient of variation greater than 30% for some of the estimates in the chart. Ultimately, this means that caution should be taken when interpreting these estimates. The control charts use three sigma limits while the ANOM charts and multivariate chart use confidence limits with a 0.1 significance level. The confidence bounds in the charts are subject to error when there are a small number of completed interviews for a given month for an interviewer in the control charts. The charts are useful in making inferences about the process of key estimates for interviewers but should be used with other quality control tools. It is important to understand that the interviewing process is dynamic and the result of complex interactions between the interviewer and respondent so some of the SPC assumptions that apply to a manufacturing setting may not hold true for the survey setting. Another limitation is that removing the outliers in the cluster level control charts may attenuate the standard errors and lead to more false positives. In addition, outliers may not represent a failure in the process, but rather the fact that a respondent with a different set of characteristics from the other respondents fell into the sample.

6. Conclusion

There is a wealth of data that can be used to improve interviewer performance in real time. SPC techniques allow managers to identify interviewers whose data quality may not meet a certain level or quality standard. A survey manager can monitor several different types of quality indicators with these SPC charts. It is crucial that the researcher or practitioner monitor both the proportion of don’t know and refusal responses and either the average value or standard deviation for a continuous question or estimate or the proportion of responses for a category other than don’t know and refusal responses for a categorical question or estimate. An example of why its vital to produce several different quality indicators for the same question or estimate is that some interviewers may have a high proportion of don’t know and refusal responses for total family income while other interviewers may be fabricating the data by making up total family income responses in which they may have an average significantly different from the overall average of the RO and cluster.

The next steps are to produce the SPC charts for all quality indicators for the NHIS. The final method needs to be developed to determine which interviewers should be investigated given that only a limited amount of resources may be available for follow-up purposes. This includes finalizing tables and charts that will summarize the information so that managers can decide which interviewers and quality indicators to explore further. Additional outlier detection methods need to be explored when creating the cluster level control charts. The clusters may also be modified so that the sample sizes can be increased to allow for additional reliability in using the SPC tools. The SPC techniques may appear to be complex, but they seem implementable and likely to be useful for improving interviewer performance to maintain high quality data.

7. Acknowledgements

I wish to thank James Dahlhamer at the National Center for Health Statistics, Timothy Kennel, and Candice Barnes at the U.S. Census Bureau for their useful comments and suggestions.

8. References

Bruce, A., and Robinson, G. (2006). Tract-level planning database. *Internal Census Bureau Report.* U.S. Census Bureau, Suitland, MD.

Centers for Disease Control and Prevention (2009). Retrieved May 20, 2011 from http://www.cdc.gov/nchs/nhis.htm.

Couper, M. and Lyberg, L. (2005). The use of paradata in survey research. *Proceedings of the 55th Session of the International Statistical Institute, Sydney, Australia.*

Dickinson, W. (2001). Escaping Flatland: Chernoff’s Faces Revisted. *SAS Users Group International Proceedings.*

Gitlow, H., Gitlow., Oppenheim, A., Oppenheim, R. (1989). *Tools and Methods for the Improvement of Quality*. Illinois: Richard D. Irwin.

Jans, M., Sirkis, R., and Duffy, B. (2010). Understanding Variability in Field Survey Structures*:* Using Statistical Process Control and Multilevel Modeling to Improve Collection Processes. Retrieved November 15, 2011 from http://www.southampton.ac.uk/s3ri/events/workshops/Files/JANS%20et%20al_Exploring_Field_Variability_PAPER.pdf.

Japec, L. (2002). The Interview Process Model and the Concept of Interviewer Burden. *Proceedings of the 2002 Joint Statistical Meetings, NewYork..*

Nelson, D. (2001). Variable reduction for modeling using PROC VARCLUS. *SAS Users Group International Proceedings.*

Pierchala, C., and Surti, J. (2009). Control Charts as a Tool for Data Quality Control. *Journal of Official Statistics.*

Ramasubramanian, V. (2010). Time-Series Analysis, Modelling and Forecasting Using SAS Software. Retrieved February 21, 2012 from http://web.iasri.res.in/nars/sas_manual/5-TS_SAS_lecture.pdf.

Ryan, T. P. (1989). *Statistical Methods for Quality Improvement*, 2nd edition. New York: John Wiley & Sons.

Sarle, W. S., (1983) *Cubic Clustering Criterion*, SAS Technical Report A-108, Cary, NC.

SAS Institute Inc., *SAS/QC® User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999.

Shaw, D., Gonzalez, J., Khare, M. (2010). Assessment of Alternative Weighting Methods for the National Health Interview Survey. *Proceedings of the 2010 Joint Statistical Meetings, Vancouver, Canada.*

Sirkis, R, and Jans, M.(2011). Using Statistical Process Control to Understand Variation in Computer Assisted Personal Interviewing Data. *Proceedings of the 2011 Joint Statistical Meetings, Miami Beach, FL.*

Western Electric Co., Inc (1956). *Statistical Quality Control Handbook,* 2nd ed. Easton: Mack Printing Company.

Wheeler, D. (2004). *Advanced Topics in Statistical Process Control*, 2nd edition. New York: SPC Press.