

Leveraging SHEWHART Procedure Options to Monitor and Evaluate Improvements in Healthcare

Roderick Jones, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL
Lynn (Xiaohong) Liu, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL

ABSTRACT

In healthcare, the purpose of statistical process control (SPC) is often to quantify improvements and identify unintended consequences resulting from an intentional change in an environment, policy, treatment protocol, or decision-support tool. SAS/QC facilitates the production of statistics and their visualization through the SHEWHART procedure. Unlike in manufacturing, process change - rather than stability - is commonly sought, and interventions might be frequent and staggered over time. Using a PCHART example for a proportion metric, we describe approaches to 1) defining a baseline period, and extending its center line prospectively to apply special cause variation tests against it; 2) removing (or "ghosting") from baseline calculations any subgroups that are considered the result of special cause variation; 3) using TESTNMETHOD=STANDARDIZE when subgroup sample sizes are not constant; 4) using the TESTACROSS option to detect variation spanning phases; 5) leveraging the contents of output datasets.

INTRODUCTION

The Institute for Healthcare Improvement works with health systems, countries, and other organizations on improving quality, safety, and value in health care (Institute for Healthcare Improvement 2018). One aspect of its endorsement of the science of improvement is the use of statistical process control (SPC) to quantify improvements and identify unintended consequences resulting from an intentional change in an environment, policy, treatment protocol, or decision-support tool. SAS/QC® facilitates the production of statistics and their visualization through the SHEWHART procedure. Unlike in other contexts such as manufacturing, process change – rather than stability – is sought, and interventions might be frequent and staggered over time.

The Health Care Data Guide: Learning from Data for Improvement provides guidance to healthcare institutions to develop and operationalize the knowledge and tools needed for “Learning from Variation in Data” (the title of chapter 4) through the use of SPC (Provost and Murray 2011). The section titled “Establishing and Revising Limits for Shewhart Charts” highlights several nuances operational leaders and their statisticians should understand and consider in analyzing the impact of intervention. These include defining a baseline period; “ghosting” data points in the baseline period that are considered special cause variation; and detecting improvements after freezing and extending an initial mean, or center line.

This paper describes the foundational syntax for creating a p chart for the proportions of nonconforming units in a healthcare context using PROC SHEWHART's PCHART statement, as well as the prerequisite summary data preparation. We share our experiences with defining baseline periods, taking into account both statistical and operational factors. And we highlight the problems and solutions involved in freezing and extending p chart central lines across phases, conducting special cause tests, and identifying and displaying shifts using a variety of PROC SHEWHART options. With this orientation, healthcare statisticians have a framework for applying these measurement tools to quality improvement, patient safety and cost-effectiveness initiatives.

GETTING STARTED WITH PCHART

The foundation for SPC in SAS/QC is PROC SHEWHART. A common scenario for its use in healthcare is the measurement of non-compliance with a protocol in a specified group of patients. For example, the proportion of hospital visits by patients with a particular condition (the denominator, or “population”) whose care does not include the recommended diagnostic testing and/or treatment (the numerator, or “nonconforming” units). The SPC chart appropriate for a proportion (or percentage) is the p chart. The

core syntax used to produce the *p* chart is:

```
PROC SHEWHART data=NumDenom ;
  PCHART Nonconforming*&TemporalUnit / SUBGROUPN = population
                                     DISCRETE SKIPHLABELS=1
                                     TOTPANELS=1;
  LABEL Nonconforming ="Proportion nonconforming";
RUN;
```

DISCRETE and SKIPHLABELS are used to manage the labeling and sparseness of the X-axis. TOTPANELS=1 ensures output is contained within a single graphic image. The data set NumDenom contains at least three variables – the temporal unit identifier (e.g., week, month), the number of nonconforming units in that period, and the population (or “subgroup”) size in the same period.

PREPARING SUMMARY DATA

Our simulated detail-level data set of visits that are eligible for the intervention, Eligible, contains a unique identifier for the visit, the visit date, the week-ending date, month, and whether the recommended care was not provided. Output 1 shows five representative records.

ID	Visit	Week	Month	Nonconforming
9627	09/14/16	09/17/16	2016-09	Yes
9541	09/14/16	09/17/16	2016-09	No
9469	09/12/16	09/17/16	2016-09	No
9368	09/12/16	09/17/16	2016-09	No
8918	09/09/16	09/10/16	2016-09	Yes

Output 1. Excerpt of the detail-level ELIGIBLE data set from a PROC PRINT Statement

These steps are required to convert the detail-level data set into the summary one, NumDenom, before executing PROC SHEWHART:

```
%LET Dataset=ELIGIBLE;
%LET TemporalUnit=Week;
%LET StartEvaluation=10SEP16;
%LET EndEvaluation=29JAN17;

ODS OUTPUT CROSSLIST=CROSSTABS;
PROC FREQ DATA=&dataset;
  TABLES &TemporalUnit*Nonconforming/CROSSLIST ;
  WHERE VISIT GE "&StartEvaluation"D AND VISIT LT
    "&EndEvaluation"D;
RUN;

DATA Numerator;
  set CROSSTABS;
  if Nonconforming="Yes";
  if &TemporalUnit ne '';*To eliminate the Summary row;
RUN;
```

```

DATA Denominator;
  set CROSSTABS;
    if F_Nonconforming="Total";
    if &TemporalUnit ne '';*To eliminate the Summary row;
RUN;

PROC SQL; CREATE TABLE NumDenom as
  SELECT Numerator.&TemporalUnit,
    Numerator.frequency as Nonconforming label="Nonconforming (N)",
    Denominator.frequency as Population label="Population",
    Numerator.rowpercent as Percentage label="Proportion
nonconforming"
  from Numerator RIGHT JOIN Denominator on
    Numerator.&TemporalUnit=Denominator.&TemporalUnit;
QUIT;

```

Output 2 shows five representative records of the data set prepared for PROC SHEWHART processing.

Obs	Week	Nonconforming	Population	Percentage
1	09/17/16	3	13	23.08
2	09/24/16	8	20	40.00
3	10/01/16	1	10	10.00
4	10/08/16	3	12	25.00
5	10/15/16	5	12	41.67

Output 2. Excerpt of the summary-level NumDenom data set from a PROC PRINT Statement

DEFINING A BASELINE

Provost and Murray propose establishing a baseline mean and control limits based upon at least 20 data points during which the process of interest was stable, presumably counting backwards from the date of implementation of an improvement strategy (Provost and Murray 2011). These points could pertain to any time span – days, weeks, months, quarters, etc.

The primary statistical consideration in defining a baseline period is the subgroup size. For p charts, an accepted rule is that the population size in each temporal unit should be greater than 300 divided by the proportion of non-conforming units ($n > 300/p$). In some cases, the desired temporal unit is too granular. In our simulation ELIGIBLE data set, an analysis of the 20 weeks prior to an initial intervention (which occurred the first week of February 2017) revealed the overall proportion of visits not resulting in the recommended care was 0.22, and a calculated minimal subgroup size of 14. But only 3 of the 20 weeks had a population of at least 14. By changing the temporal unit to month and extending the time period into the past such that 20 monthly data points were available, the proportion was similar at 0.21 and the smallest subgroup size was 26.

A comparison of the p charts produced for week (Figure 1) and month-based (Figure 2) baseline illustrates the drawback of smaller subgroup sizes: wider control limits.

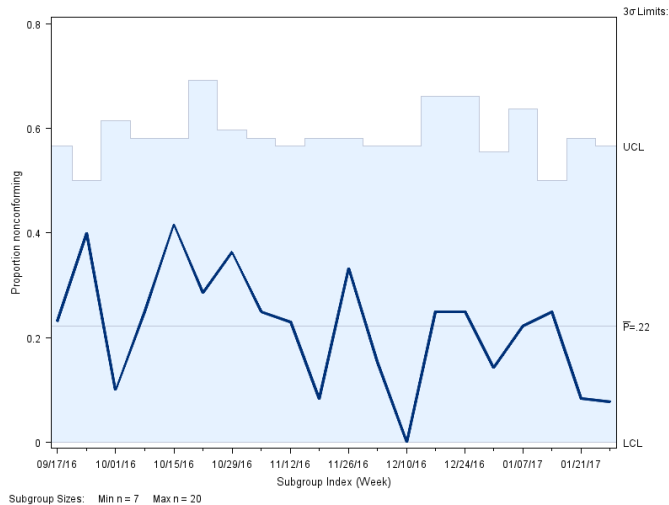


Figure 1. *p* Chart of Nonconforming Units over a 20-Week Baseline Period

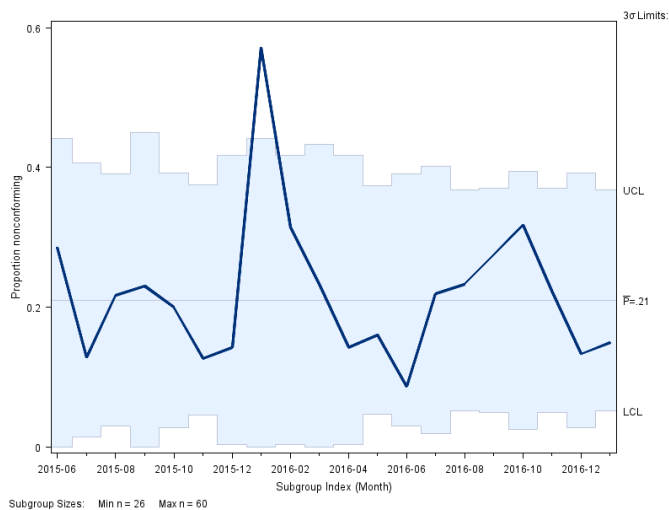


Figure 2. *p* Chart of Nonconforming Units over a 20-Month Baseline Period

An important – and in some cases, contradictory – operational consideration related to baseline definition is that a longer baseline is not necessarily better. This might be counter-intuitive at first: a longer period of observation of a process ought to give a more stable picture of reality. Perhaps not in a healthcare setting, where the ecosystem of environment, processes, personnel, clinical decision support (i.e. electronic medical record features), and patient characteristics is dynamic. To detect change associated with an intervention, a baseline that is temporally compact, and less likely to be impacted by variables from the distant past that are no longer relevant, is preferable.

TESTING FOR SPECIAL CAUSE VARIATION

The month-based figure displays one point, 2015-11, outside the control limits. PROC SHEWHART provides four standard tests for special causes that are recommended for *p* charts. Among these, we expect the statistical detection of one point beyond the control limits (Test 1) to be positive at 2015-11.

To apply these tests to the simulation, it is necessary to add two options to the PCHART statement:

```
TESTS=1 to 4
TESTMETHOD=STANDARDIZE
```

The TESTS= option establishes which tests are sought, but will yield no results because of the unequal subgroup sizes. TESTNMETHOD=STANDARDIZE allows the tests to be applied to data with unequal subgroup sizes (Nelson 1989). As expected, Test 1 is positive, and will need to be accounted for later in the process.

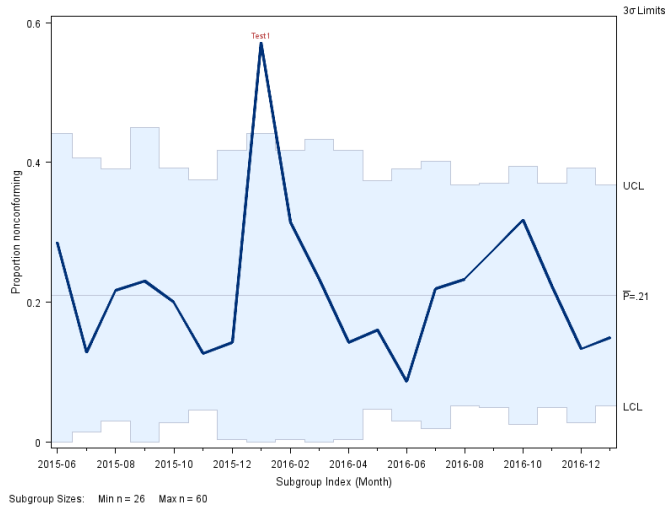


Figure 3. *p* Chart of Nonconforming Units over a 20-Month Baseline Period with Test 1 Positive

FREEZING AND EXTENDING THE CENTER LINE

The main purpose for defining the center line and control limits for the baseline period is the generation of a comparison period against which improvements or unintended consequences of the intervention can be detected and statistically quantified. Once observations are available for a time period in which the improvement interventions are occurring, the foundational PROC SHEWHART code is no longer adequate. In the absence of enhancements, the *p* chart generated for the baseline period plus 10 months of intervention period, shown in Figure 4, is incorrect. The center line calculation makes use of the entire time series rather than distinguishing between points in the baseline and intervention period.

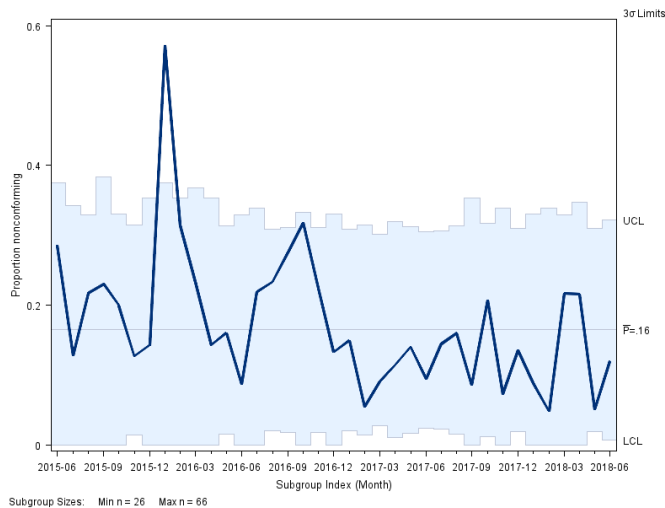


Figure 4. *p* Chart of Nonconforming Units over a 20-Month Baseline and 10-Month Intervention Period

ADDING PHASES TO THE PCHART

Revising the code to reflect baseline and intervention phases is necessary. First, the phases are defined within the summary data set:

```

data NumDenom;
  SET NumDenom;
  format _phase_ $20.;
  If SUBSTR(Month,1,4) IN ('2015', '2016') OR MONTH='2017-01'
    THEN _phase_='Baseline';
    Else _phase_='Intervention';

RUN;

```

Second, a preliminary PROC SHEWHART is needed to generate a data set containing center line values for each phase in future runs. OUTLIMITS= is the option used to create it. Graphic output is not needed so the NOCHART option is used:

```

PROC SHEWHART data=NumDenom;
  BY _phase_ notsorted;
  PCHART Nonconforming*&TemporalUnit / SUBGROUPN = population
    NOCHART
    OUTLIMITS=Limits
    (RENAME=( _phase_=_index_ ));

RUN;

```

Third, the center lines and phase annotations are applied in the PROC SHEWHART used to generate the *p* chart using the added statements READPHASE=, READINDEXES=, PHASEREF and PHASELEGEND:

```

PROC SHEWHART data=NumDenom
  LIMITS= Limits;
  PCHART Nonconforming*Month / SUBGROUPN = population
    READPHASE = ALL
    READINDEXES = ALL
    PHASEREF
    PHASELEGEND
    TOTPANELS=1
    DISCRETE
    SKIPHLABELS=1;
  label Nonconforming ="Proportion nonconforming";

RUN;

```

The resulting *p* chart shown in Figure 5 remains incorrect in that the interruption in the center line graphically suggests a statistical shift beginning with the first month of the intervention.

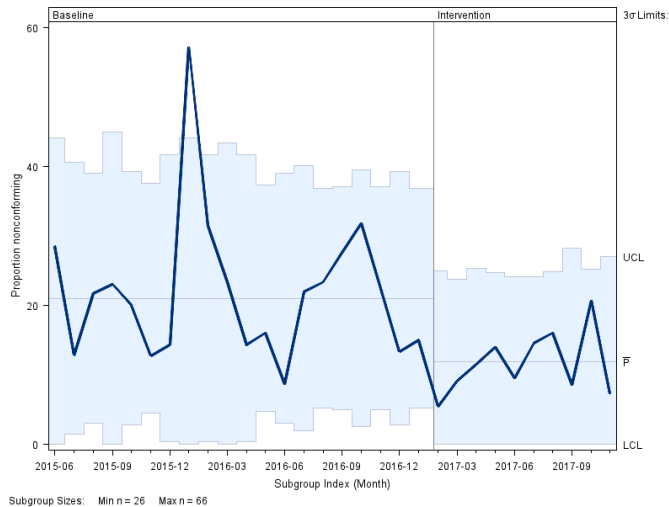


Figure 5. p Chart of Nonconforming Units over a 20-Month Baseline and 10-Month Intervention Period with Phases defined

Manufacturing uses SPC differently than healthcare. In many manufacturing examples of SPC, there are system changes that produce discrete “before” and “after” periods. For example, the replacement of a machine that produces widgets. In healthcare, the approach is different:

“A good technique in using a Shewhart chart to detect improvement is to create initial limits . . . and freeze and extend these initial limits and center line into the future. Special cause in the new data being added to the chart will be detected more rapidly because the new data are not influencing the center line or the limits.” (Provost and Murray 2011, p. 122.)

Additionally, while a typical manufacturing scenario might involve sampling observations, and therefore generate data sets with identical subgroup sizes over time, healthcare cases more often make use of the entire population meeting the inclusion criteria. This leads to uneven subgroup sizes and means that the center line, but not the control limits, should be extended.

GHOSTING SPECIAL CAUSE POINTS

But even before implementing the steps to extend the center line, there is another adjustment needed. The Test-1 positive result for 2015-11 ought to be removed or “ghosted” from the calculation of the center line upon verification by operational stakeholders that the conditions that caused this outlier were unique to that time period and irrelevant to the measurement of the intervention’s impact, this data point must be removed from the calculation of the center line value in the baseline period, or “ghosted”.

“Ghosting” here means leaving the data visible on the graph but excluding it from the calculation of the mean and limits. The limits revised by this technique will reflect only common cause variation in the process and may now be extended into the future.

To ghost the special cause point and extend the center line value of the baseline period now requires a series of enhancements. The ghosting and recalculation of the center line is done on the months from the baseline period using output data sets from the OUTTABLE and OUTHISTORY options. The special cause point is identified using the value of `_EXLIM_`, merged with the output summary statistics, and then the merged data set is used in a run of the SHEWHART procedure where `data=NumDenom` is replaced by `history=HISTORY`:

```

PROC SHEWHART data=NumDenom (where=( _phase_='Baseline' ));
    PCHART Nonconforming*&TemporalUnit / SUBGROUPN = population
        DISCRETE
        SKIPHLABELS=1
        TOTPANELS=1
        TESTS=1 to 4
        TESTNMETHOD=STANDARDIZE
        OUTTABLE=OUTTABLE
        OUTHISTORY=HISTORY;
    label Nonconforming ="Proportion nonconforming";

```

```
RUN;
```

```

DATA OUTTABLE;
    SET OUTTABLE;
        if _EXLIM_ ne '' then _COMP_='N';
        else _COMP_='Y';
    KEEP &TemporalUnit _SUBN_ _COMP_;

```

```
RUN;
```

```

DATA HISTORY;
    MERGE HISTORY OUTTABLE;

```

```
RUN;
```

```

PROC SHEWHART HISTORY=HISTORY;
    PCHART Nonconforming*&TemporalUnit / SUBGROUPN = population
        DISCRETE
        SKIPHLABELS=1
        TOTPANELS=1
        TESTS=1 to 4
        TESTNMETHOD=STANDARDIZE
        OUTTABLE=OUTTABLE
        OUTHISTORY=HISTORY;
    LABEL Nonconforming ="Proportion nonconforming";

```

```
RUN;
```

The refreshed center line value is saved for later use as the macro variable &CenterLine:

```

PROC SQL NOPRINT;
    SELECT _p_
        INTO: CenterLine
        from OUTTABLE (obs=1);
quit;

```

TESTING FOR IMPROVEMENT

To freeze and extend the center line, the LIMITS output data set must be updated. First it is recreated using the OUTLIMITS= option.


```

PROC SHEWHART data=NumDenom ;
    BY _phase_ notsorted;
    PCHART Nonconforming*&TemporalUnit / SUBGROUPN = population
        NOCHART OUTLIMITS=Limits
        (RENAME=( _phase_ =_index_ ));
RUN;

```

Then through a data step, the value of the mean proportion `_p_` is overwritten with the `&CenterLine` macro variable value before re-running PROC SHEWHART to produce Figure 6:

```

DATA limits;
    SET limits;
    _p_ = &CenterLine;
RUN;

```

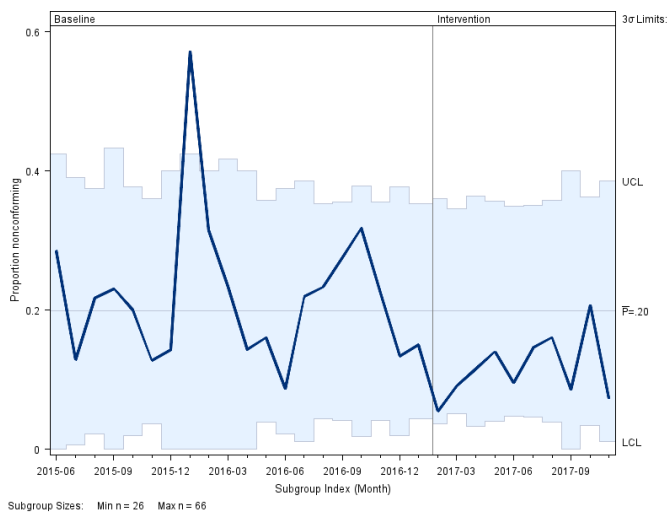


Figure 6. *p* Chart of Nonconforming Units over a 20-Month Baseline and 10-Month Intervention Period with the Baseline Mean Extended

With the baseline mean extended into the intervention period, special cause tests can commence. In healthcare, the most useful of these is Test 2, defined as nine points in a row in Zone C or beyond on one side of the central line, because it is used to quantify a process shift in the desired or undesired direction. The number of consecutive points is adjustable, and an 8-point run is also commonly used. The PCHART statement options can be supplemented with this setting through the use of the TEST2RUN option, as well test-positive labels using TESTLABEL:

```

PROC SHEWHART data=NumDenom
      LIMITS=      Limits ;
PCHART Nonconforming*&TemporalUnit / SUBGROUPN = population
      READPHASE = ALL
      READINDEXES = ALL
      PHASEREF
      PHASELEGEND
      DISCRETE
      SKIPLABELS=1
      TOTPANELS=1
      TESTS=1 to 4
      TESTNMETHOD=STANDARDIZE
      TEST2RUN=8
      TESTLABEL1="*"
      TESTLABEL2="8 pts on 1 side"
      TESTLABEL3="6 point trend"
      TESTLABEL4="alternating pattern"
      OUTTABLE=OUTTABLE
      OUTHISTORY=HISTORY;

LABEL Nonconforming ="Proportion nonconforming"; RUN;

```

An examination of Figure 7 and the `_TESTS_` variable in the `OUTTABLE` data set reveals that the eighth point of the run is 2017-09.

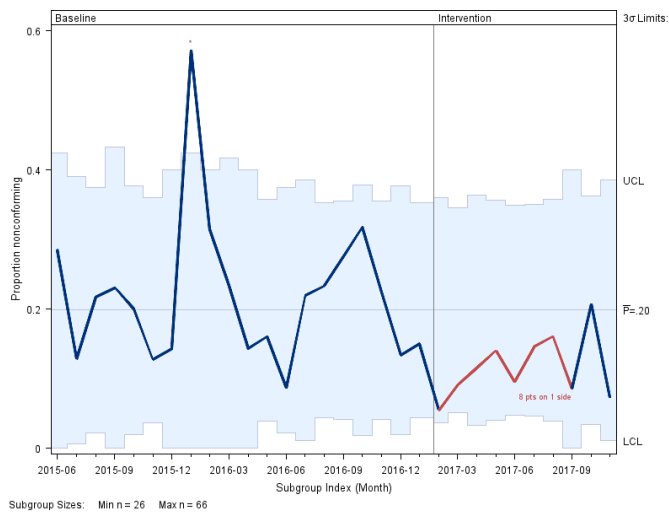


Figure 7. *p* Chart of Nonconforming Units over a 20-Month Baseline and 10-Month Intervention Period with the Baseline Mean Extended and Test 2 positive

However, there are 10 consecutive points below the central line – from 2016-12 through 2017-09. The eight-point run identified in the results reflect only the intervention phase. This finding underscores the potential usefulness of the `TESTACROSS` option, which applies tests across phase boundaries. In healthcare, `TESTACROSS` is often appropriate because the impact of a process change might be desired or observed even before a policy becomes official or prior to the date of a technology-related “go-live”. This is particularly true with respect to behavior and process changes that are publicized, promoted, or developed within the same group of people who are carrying out an intervention. Adding `TESTACROSS` to the `SHEWHART` procedure has the effect of signaling a positive Test 2 for 2017-07.

Whether TESTACROSS is used or not, the simulation data provides evidence of a shift – sustained, lower rates of nonconformity. This shift requires that a new mean proportion be calculated and projected forward into the future, so special cause tests cease to be calculated against the baseline mean. To produce the appropriate SPC chart now requires repetition of several of the steps above: calculating the value of \bar{p} in the months of the 8-point run, extending it forward from the beginning of those 8 points, redefining and renaming the intervention phase to identify it as a shift, and obtaining test results prospectively as new months of data become available. With more data added, the simulation data set produces Figure 8.

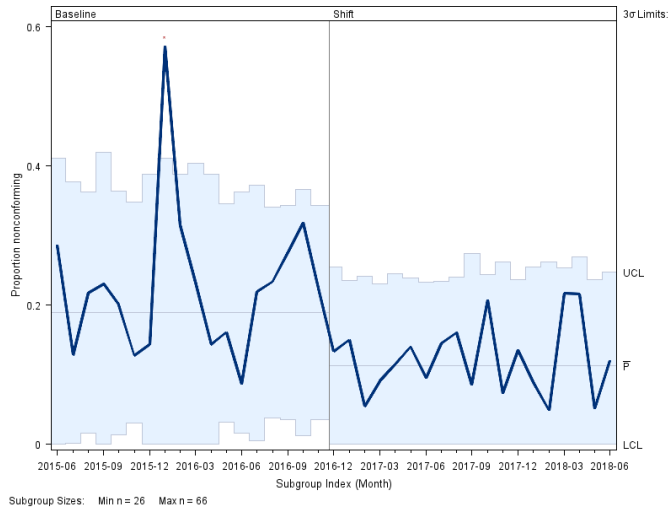


Figure 8. p Chart of Nonconforming Units over a 20-Month Baseline and 17-Month Intervention Period with One Shift

CONCLUSION

Leveraging the options provided by PROC SHEWHART facilitates the creation of SPC charts that extend central lines from baseline periods and display shifts to represent process changes. In developing a p chart for a proportion metric, one of the most common scenarios when using SPC in the healthcare setting, these options are essential:

- DISCRETE and SKIPLABELS allows us to manage the labeling and sparseness of the X-axis.
- TOTPANELS=1 ensures the output is contained within a single graphic image.
- OUTLIMITS, OUTTABLE and OUTHISTORY generate output datasets containing key statistics that are used in subsequent development steps.
- NOCHART suppresses graphic output when the primary purpose of a PROC SHEWHART run is to generate output datasets.
- READPHASE=, READINDEXES=, PHASEREF and PHASELEGEND introduce phases, such as baseline and intervention, and display them in the graphic output.
- TESTS= establishes the special cause tests to apply; TEST2RUN= is an example of how the number of points required for a consecutive-point run can be set by the programmer; TESTLABEL options can be used to label the positive tests on the chart.
- TESTNMETHOD=STANDARDIZE permits the tests to be applied to data with unequal subgroup sizes.
- TESTACROSS applies special cause tests across phase boundaries.

To further enhance PROC SHEWHART, we recommend these additional methods. First, use ODS GRAPHICS to create image files and control additional aspects of the output, such as titles and point markers. Second, leverage macro variables, parameter files containing the values for those macro

variables, and iterative %DO loop processing to run the processes on multiple metrics.

REFERENCES

Nelson, L. S. 1989. "Standardization of Shewhart Control Charts". *Journal of Quality Technology*, 21:287-289.

Provost, L. P. and Murray, S. K. 2011. *The Health Care Data Guide: Learning from Data for Improvement*. San Francisco, CA: John Wiley & Sons.

Institute for Healthcare Improvement. "Science of Improvement." Accessed August 3, 2018. Available at <http://www.ihl.org/about/Pages/ScienceofImprovement.aspx>.

RECOMMENDED READING

- *SAS/QC® 14.1 User's Guide*

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Roderick Jones
Ann & Robert H. Lurie Children's Hospital of Chicago
RCJones@luriechildrens.org

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.