Supplementary File 2

Categorizing Patterns of Change in Control Charts

Michael A. Murray, Peter G. Norton, Gary Teare
(2012)[1]

We used standard empirical rules commonly used in the interpretation of control charts[2] to develop a set of coding rules. Our rules were then used to make a decision about whether a team's quality indicator measurements were improving, worsening, staying the same, or were indeterminate across time. These coding rules were refined over a series of iterations. For each iteration, a new set of 20 to 25 control charts based on clinical data were prepared and scored independently by two expert reviewers. After the scoring, a consensus conference of the reviewers was held to resolve differences and the rules modified as necessary.

Desired Classification Coding

Codes for 'predictable' patterns
1= better at the end
2= worse at the end
3= about the same at the end, and relatively stable across time
4= about the same at the end, but with instability in the middle (codes 3 & 4 can be combined)

Code for unpredictable or unclassifiable pattern
5= unclassifiable as to performance; erratic

NOTE: the rules were designed for the RAI results, which had relatively short series of data. For longer runs of data, consider investigating whether change happened before the end of the data series.

Specific Code Rules

Codes 1 & 2: The first two codes are concerned with the question, "Do the current (or most recent) measurements show that results[3] are better or worse than where they started?"

- By 'current' we mean the most recent few quarters of data; we would like evidence that the process is both changed (for better or worse) and somewhat predictable at a new level for the future.
    - For longer series of data, one on the order of many years for example, we might consider whether improvement happened at any point in the series regardless of whether it was sustained. This makes for harder coding.
- We are also concerned with where the process started. Given the myriad of possible patterns in processes over time, it is possible a process could become much better or worse over time and then change again. The 'current' process might look better or worse than an intermediate point BUT still not be better or worse than the beginning. We are

[1] Cite as: Murray MA, ~~Poss JW,~~ Norton PG, Teare G. *Categorizing patterns of change in control charts*. Edmonton, AB: Knowledge Utilization Studies Program, University of Alberta, 2012.
[2] Commonly called the "Western Electric rules" although there are many different versions of these.
[3] Results could be 'process measures' or 'outcome measures'; our methodology would work the same way for both.

not concerned with whether the process is now better or worse than *it ever was*, just whether it is better or worse than *at the beginning*.

- A single 3 SD rule violation either at the very beginning of the process *or* right at the end of the process is not sufficient evidence of a process that is better or worse at the end, or that the process is unstable and therefore unclassifiable. In combination with other nearby signals, however, it might help provide evidence of change. Especially consider discounting a widely divergent point at the beginning or end of the series in making judgements.
- In judging whether a process is better or worse at the end, it is useful to consider how many points at the beginning and end of the data series are involved in providing evidence of change.
    - At the beginning of the process we have no prior knowledge about it and would like to see evidence based on more than 1 or 2 points before deciding on the initial level of performance. Use at least 3 points (9 months when looking at RAI quarterly data) to decide. Often it appears that the first data point is wildly different than subsequent points. By itself, as noted earlier, it provides insufficient evidence to base subsequent decisions on but, if there are other rule violations at the beginning, that might be sufficient. A first data point need not be out-of-control to be discounted as establishing initial performance, but it should be wildly divergent from the next few values to be discounted.
    - At the end of the process, we know much more about its performance. A single divergent point at the end of the process shouldn't be used to judge improvement or deterioration, but a 2-of-3 rule violation might be sufficient. The two other multiple point rules (4-of-5 or 8-on-a-side) are better evidence of sustained change.
- When viewing a process with relatively stable beginning and ending periods but instability in the middle, be wary of considering a process as having changed (up or down) when mean differences are small. One of the longer multi-point rule violations (4-of-5 or 8-on-a-side) at the beginning or end might be sufficient to say the process is better or worse, but just seeing slightly lower or higher scores is insufficient evidence of change.
- Trend rules are hard to use when evaluating processes. A trend violation at the beginning of the data series says that the process isn't stable to begin with and deciding the actual starting level of the process then becomes hard. At the end, a trend might indicate improvement or decline BUT what matters is where the trend ends. A process could be improving but still be worse than where it started, or vice versa. Trend violations are unlikely, therefore, to be good measures of improvement or decline, but may, in combination with other rules, give some evidence of change.

Codes 3 & 4: The third and fourth codes are concerned with the question, "Given that the current scores look to be at about the same level as where they started (i.e. not coded 1 or 2), were the results relatively stable across time or did they show substantial variation in the middle?"

- A 'pure' code 3 pattern would show a process with **no** rule violations over its entire length.

- A relatively stable process might have a rule violation that is inconsequential in understanding long term performance. It might be a 3 SD violation in the middle, or a trend, or an isolated 2-of-3 violation. Consider this a code 3.
- The current process could have roughly the same level as the beginning but with substantial variation in between. Multiple rule violations, especially when they are on both sides of the centre line, are probably evidence of a code 4.

Code 5: A process might be sufficiently erratic that deciding whether it has improved, worsened or stayed the same is just not possible. A process might be just too erratic over the whole series to decide what is happening, or with too few stable points at the end to be confident about its most recent performance. In that case, consider using code 5.

- The last few subgroup values might not be stable enough to judge how the process is actually performing and to get any idea about whether it is better or not.
- There might be multiple rule violations on both sides of the centre line from beginning to end, and the process may appear to be just not stable.
- Deciding on whether a process is 'about the same but with instability' or 'unclassifiable' may be difficult. The distinction is based on how many points are available to judge the ending of the series. If a period of higher or lower performance in the middle or near the end of the process gives insufficient points to be confident in a judgement or statement that a process has changed, then consider the process 'unclassifiable'.

General coding considerations and issues

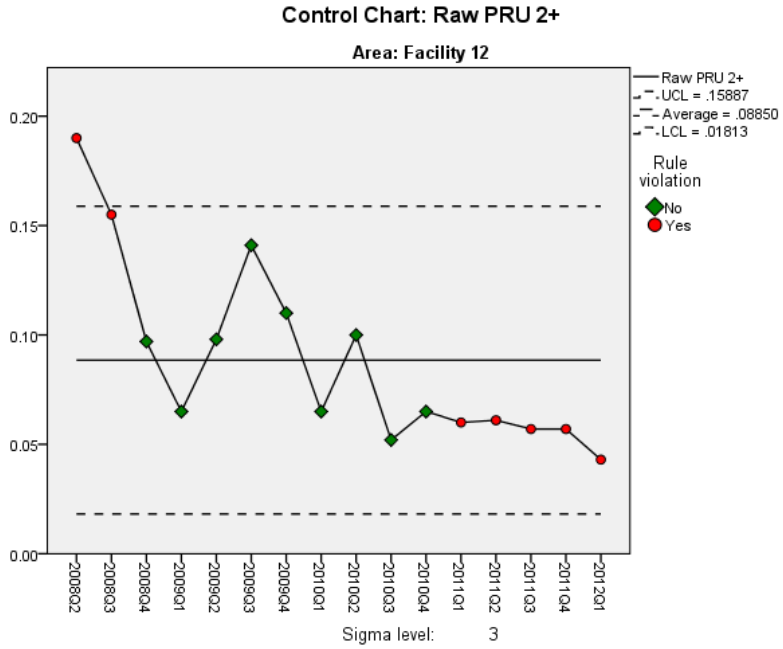
- If you find yourself waffling between codes, especially a code 4 or something else, consider using code 5.
- For research purposes, you might want to be conservative in saying a process is better or worse, erring on the side of codes 3 through 5.
- When considering giving feedback to organizations, these codes and rules might be *too* conservative and not give sufficient credit for improvements over the medium term. Yes, performance might not have changed over the long term, but over the most recent year or so it might have.
- Given the relatively short data series available and the way control limits are calculated using Moving Ranges, a single HUGE change in values might broaden limits and change the averages. The Nelson method compensates for this and you might want to consider running a few charts through it[4]. As a shortcut, when you see a subgroup value that is wildly different from all others, consider whether some values might become a violation as the limits shrink slightly AND change your classification. We can investigate those cases in subsequent analyses.

[4] In an XmR (individuals) chart, the Nelson rule removes points with a 3 SD violation ON THE MOVING RANGE (mR) CHART from calculation of limits on the X chart. Huge jumps in points, which affects average moving range, might cause points to be removed from calculation of limits; it isn't just about high or low points.

Examples to show coding rules

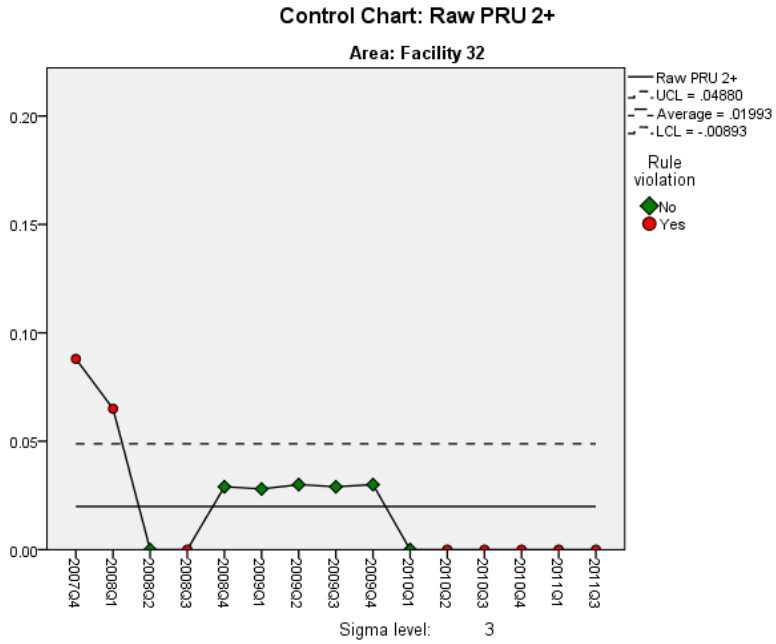
Charts show 3 SD limits & 3 SD violations as well as other rules.

1=Better

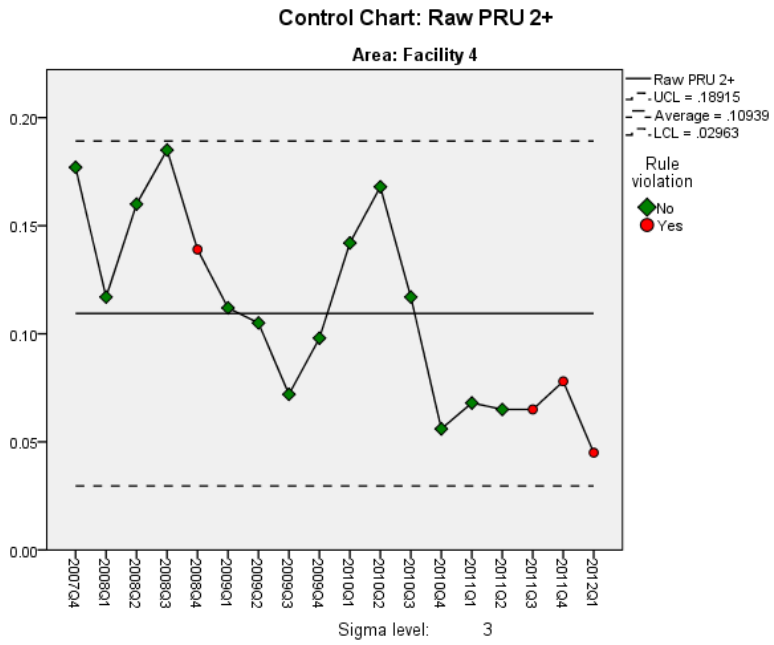


1=Better

Starts with 2 out-of-control high points (1 wouldn't have been enough and even 2 is suspect), but ends with 6 zeros.

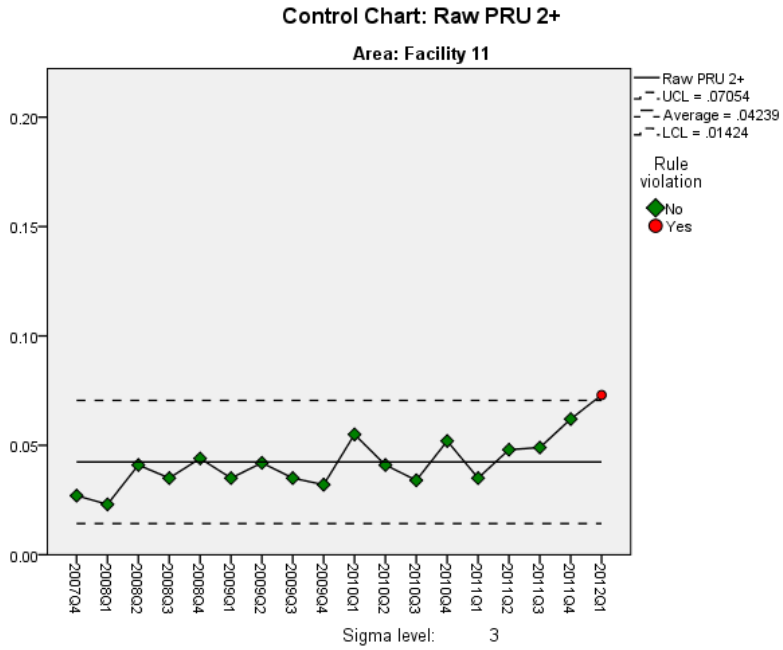


1=Better



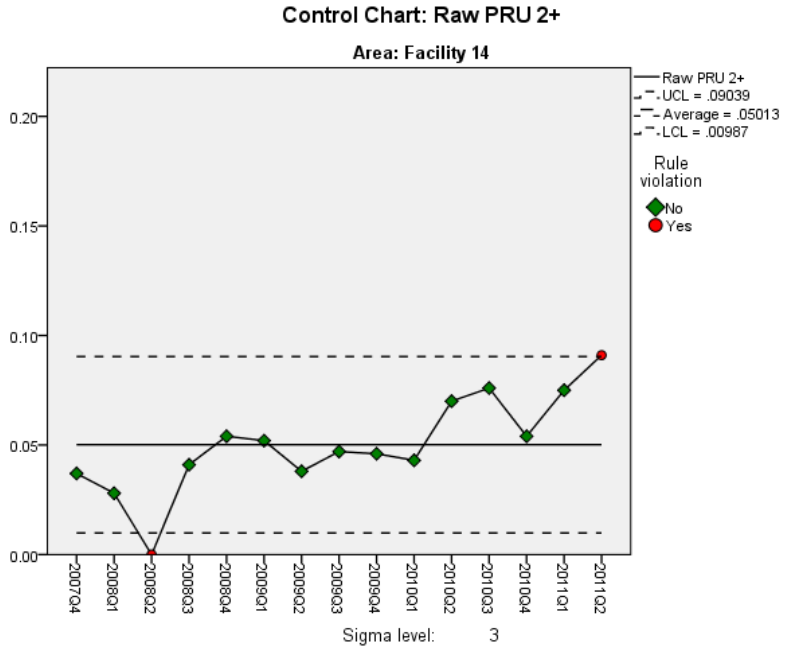
2= Worse

Just barely worse because the final 2 points are out-of-control. One point alone at the end would have been insufficient evidence of worsening. Using a Nelson rule would probably NOT have been enough to cause the first 2 points to violate a rule.

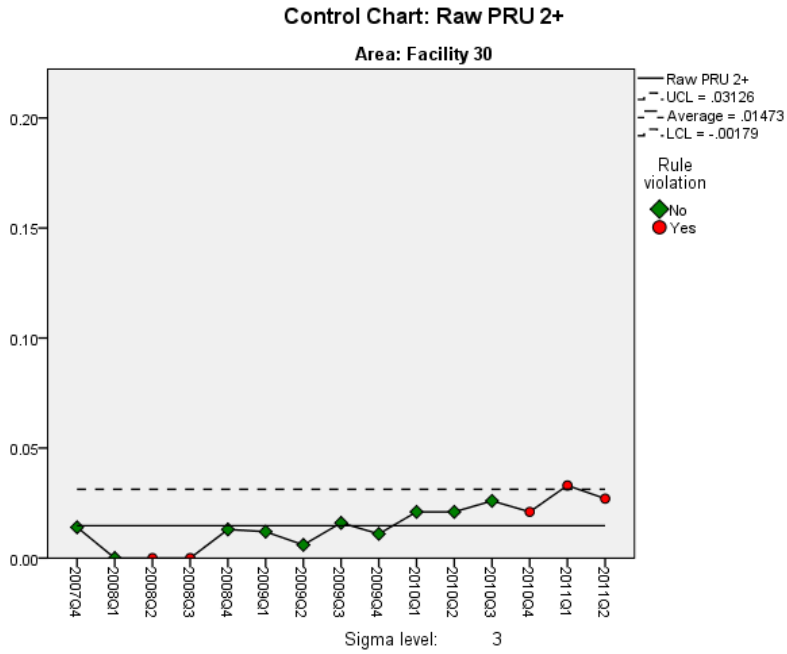


2=Worse

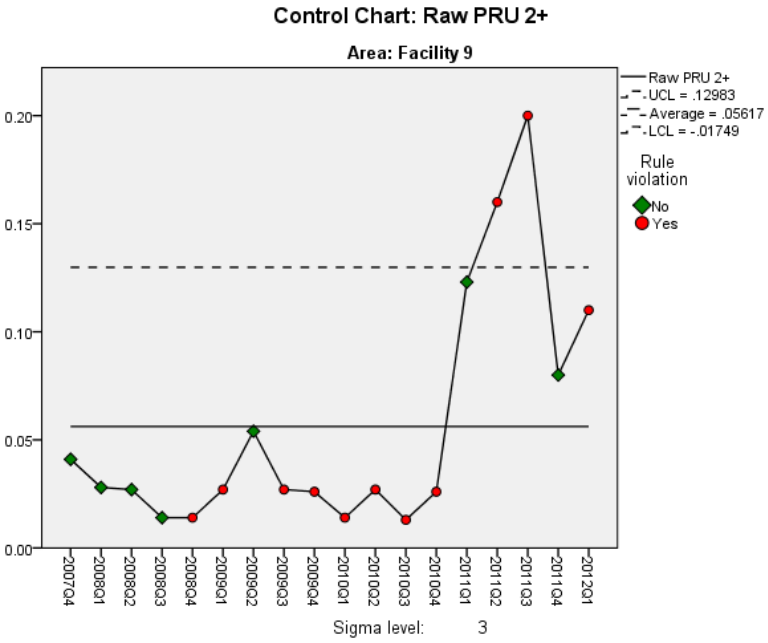
Starts low, ends high (4 of last 5 above 1 sigma)



2=Worse

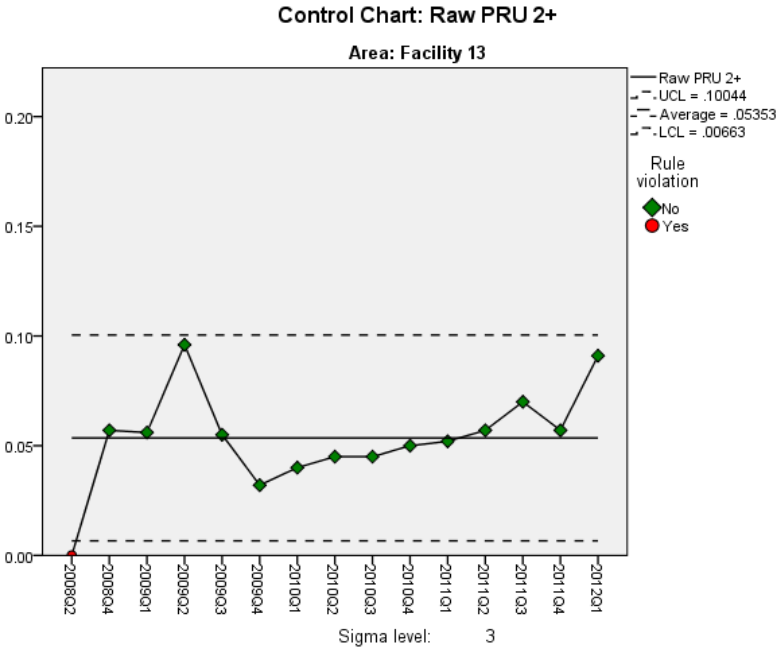


2=Worse



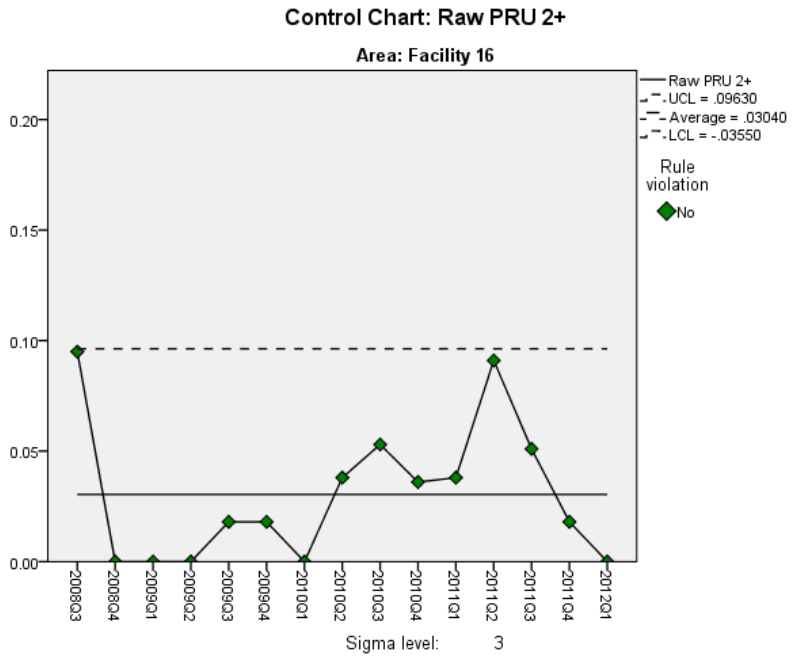
3=Same

Ignore first point and spike in last point. The worsening trend in the middle is NOT sustained long enough.

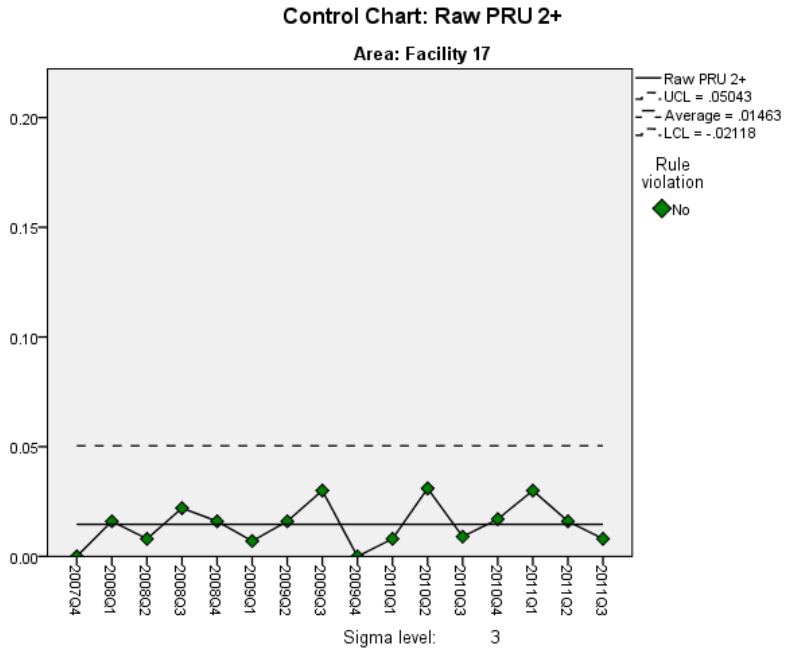


3=Same

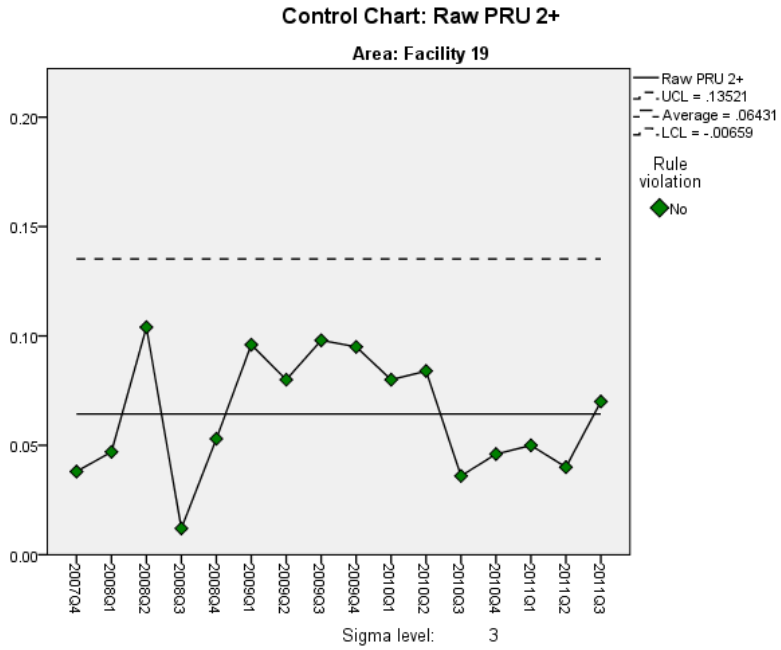
No rule violations. Re-running with the Nelson rule would probably push the first point and fourth-last point out of control, but the process would probably still be stable.



3=Same

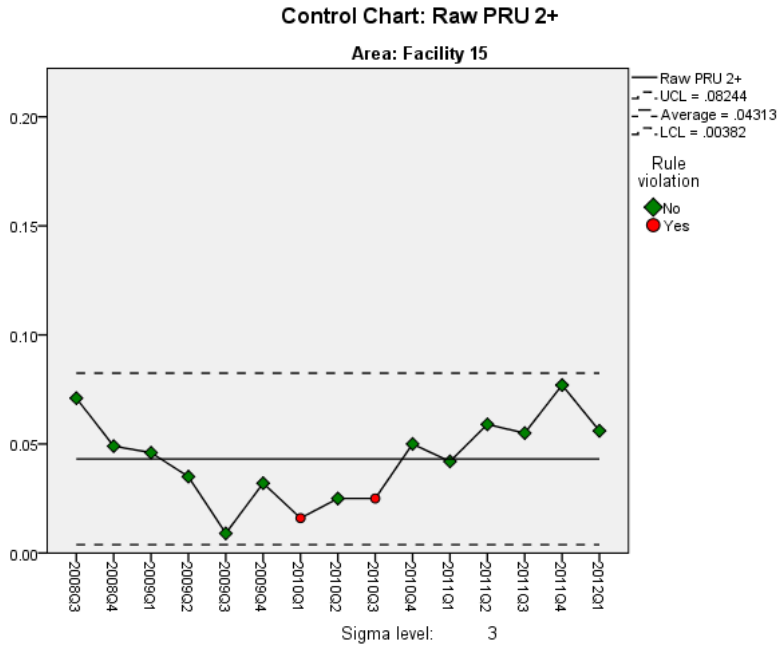


3=Same

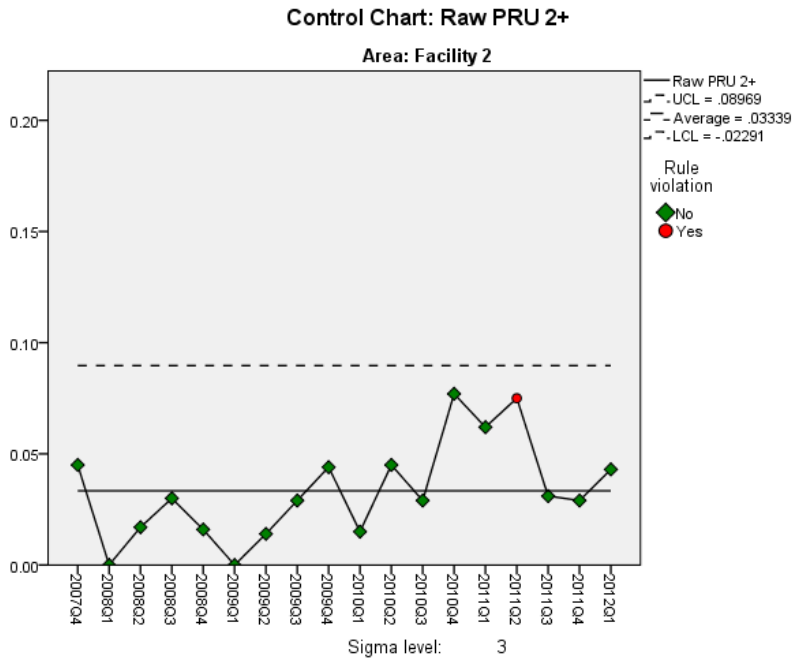


4=Same with instability

The last 4 points are not sufficiently higher than the first few to say with confidence that the process is worse. A further 1 or 2 points at the same level might say so.

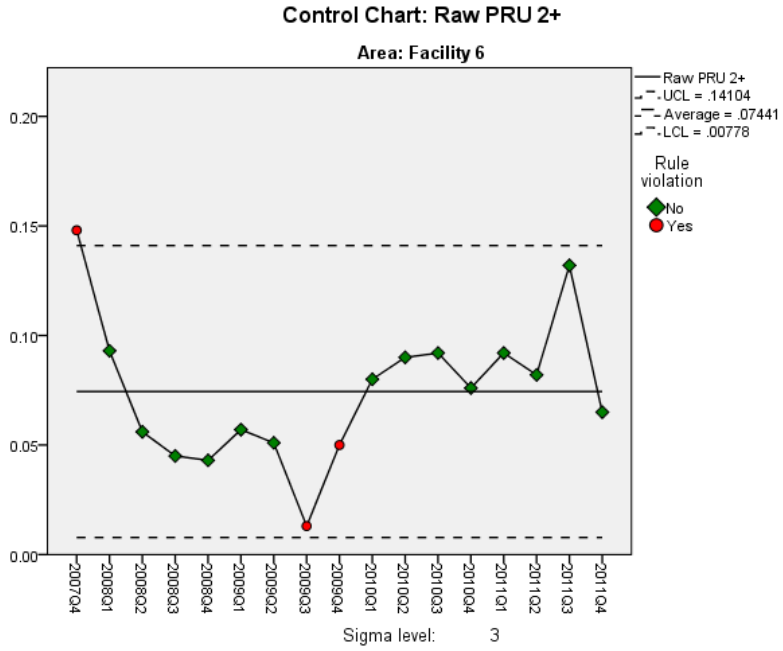


4=Same with instability



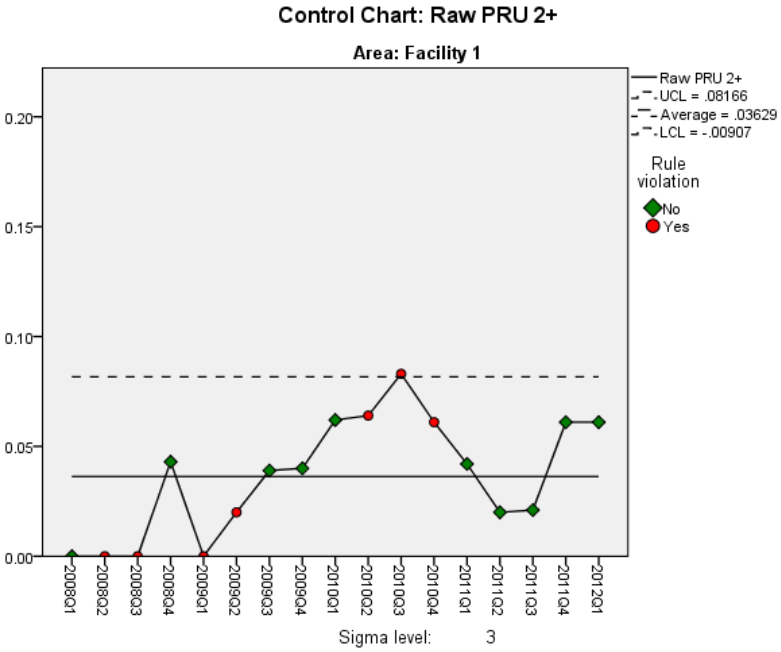
5= Unclassifiable

The first two points are higher, then the next 5 are about the same level, but 2 more points show that the process is better. The process then goes back to the middle, then high (which might be out-of-control using the Nelson rule), then back down. It is hard to predict the next point and impossible to say if the process is better or worse since its start. The last high is worse compared to most of the middle half.



5=Unclassifiable

Although the process starts off at a good level, it is worse in the middle and the final 6 points are confusing. Whether the process will continue higher (and therefore worse) or back down to the third- and fourth-last points is unpredictable.



5=Unclassifiable

Technically this is a code of 3 because there are no rule violations, but application of the Nelson rule would push the high point out. Absolute differences in performance are very large. The start and end points are zero but clearly the final few points aren't zero just as the first 4 points aren't zero.