# Notes on Using Run Charts for Quality Improvement

Kevin Little, Ph.D. 26 January 2015

## Summary

The probability of a set of consecutive values on one side of the median and the probability of a consecutive sequence of values increasing or decreasing are both increasing functions of the series length for random data. Restricting the simple shift and trend rules proposed by Perla et al (2011) to relatively short series (e.g. n <=20) will keep the rate of signals to less than 5%, one-sided, for one case of the shift rule in random data and less than 10%, one sided, for the trend rule and the other case of the shift rule. We provide R simulation code to allow others to replicate our calculations and as a basis for further investigation.

## Context and Models for Run Chart Shift and Trend Rules

### Context

We don’t want to fool ourselves in claiming evidence of improvement where there is none. On the other hand, we don’t want to be so conservative that we ignore initial evidence of improvement and prematurely abandon useful change ideas.

To counter the complexity of the too few or too many runs table shown as Table 1 in Perla et al (2011), in which the critical number of runs depends on the length of the series to be analyzed, the authors proposed simple rules do not depend on the the length of the series plotted on the run chart. See Perla et al (2011) and Provost and Murray (2011).

To be useful, the threshold length of the shift or trend should not occur too often in random sets of observations but not be so rare that the rules will filter out real signals of improvement too often.

The rules are easy to remember and apply as improvers seek to be neither foolish nor too conservative.

**Shift Rule**: Six consecutive values either all above or all below the median. Values that fall on the median do not add or break a shift. Skip all values that fall on the median and continue counting.

**Trend Rule**: Five or more consecutive points all going up or all going down. If the value of two or more consecutive points is the same, only count the first point and ignore the repeating values; like values do not make or break a trend.

Figure 1 Simple Rules proposed for shift and trend in Perla et al (2011).

Perla et al. (2011) use six consecutive values for shift and five consecutive points for trend as the thresholds; these rules are approximately guided by a 5% chance of these patterns occurring in random data.

Lloyd Provost clarified that he views the run chart as most useful in improvement settings with a limited number of observations, e.g. 8 or 10 to 20 values in series (email 19 January 2015). If you have 20 or more values, a Shewhart control chart typically provides more insight and that’s what Llloyd recommends.

These notes clarify the performance of the simple rules and support Lloyd’s view that the rules are most appropriate with relatively short series of observations.

One-sided or Two-sided reference probabilities?

In improvement work, there is almost always a natural direction of improvement: up or down, depending on the measure. The major exception occurs when we are trying to maintain predictable performance at a specific level, with suitably low variabiltiy. In this case, we will need to use control charts to guide our work.

Before we start improvement, we should state which direction is “better”, either up or down.

In terms of the simple trend and shift rules, looking for better peformance implies that reference probabilities should be one-sided, e.g. shift above reference median; shift below reference median; trend up; or trend down.

### Models for the two simple rules

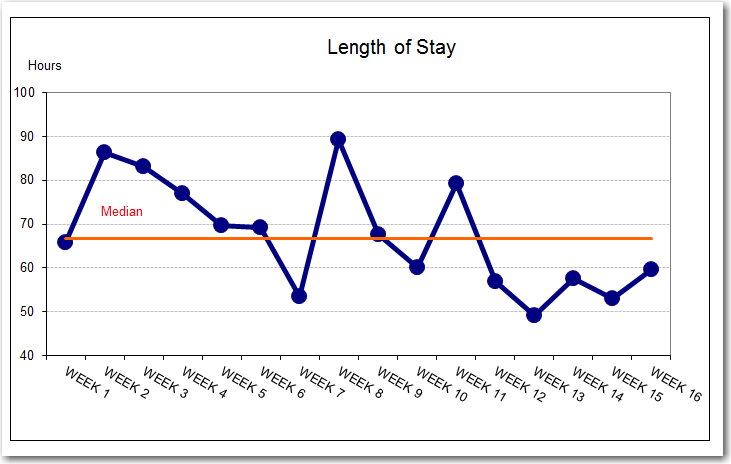
#### Trend

The trend rule is independent of a median reference. We can simulate the chance of seeing a run up or run down of k consecutive values in a series of N by permuting the vector (1:N) and counting runs up and down of k consecutive values, working with the first differences of the vector.

We use the definition of run up or run down given by Olmstead (1945) cited by Provost and Murray (2011) and Perla et al. (2011):

*"...let a1,a2,...,an be n numbers, no two alike, and let the sequence S = (h1,h2,...hn) by any permutation of a1,a2,...,an,where S is to be considered a chance variable, and each of the n! permuations of a1,a2,...,an is assigned the same probability. Consider the derived sequence R whose ith element is the sign (+ or -) of hi+1 - hi, (i=1,2,...,n-1). A sequence of p consecutive + signs immediately preceded by a - sign is called a run up of length p or more; a sequence of p consecutive - signs immediately preceded by a + sign is called a run down of length p or more."* p. 24-25

Olmstead’s definition requires k consecutive increasing differences or decreasing differences to be called a run up or run down of length k; in Figure 2, a run down of length five is circled, the Week 5 value is greater than the Week 6 value.

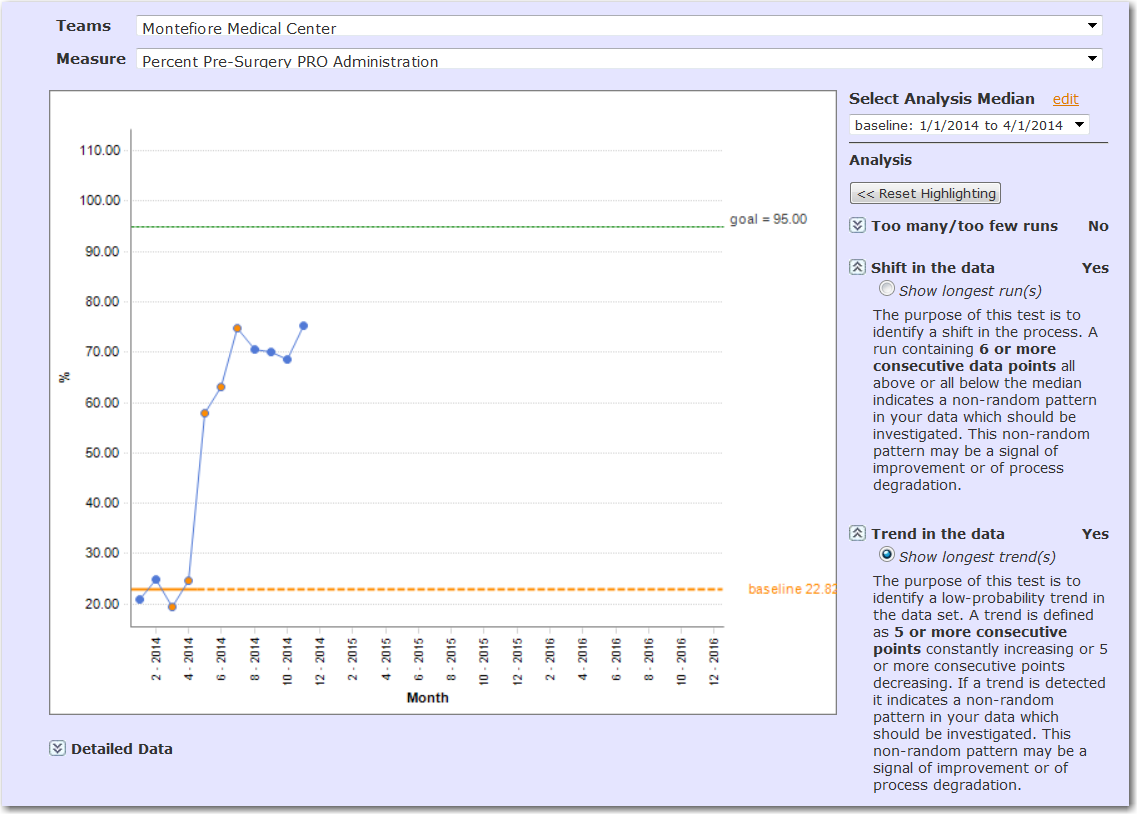


Better

Figure 2 example of a series that shows a run down of length 5 (five consecutive negative differences starting at point plotted at Week 2); confusion in describing the run down arises because there are six points circled. Careful application of the trend rule requires the number of discharges that drives the length of stay number to be roughly constant week to week, varying no more than about 20% throughout the series.

The IHI Extranet uses a rule for a run up or run down of length 4 as defined by Olmstead:

Figure 3 Screen shot of IHI Extranet display shows that k=4 is the value used for the run up or run down rule. View obtained from *Advanced Reports Measure Run Charts* selection in the Reports section of the Extranet. Accessed 23 January 2015.



Olmstead provides exact and estimated frequencies for runs up and runs down in random series for a range of series lengths n; we also programmed a simple simulation in R to estimate the frequencies as an exercise to understand Olmstead’s calculations.

Here’s the relevant **two-sided** probability table for runs up or down of length k=4 and k= 5 for random data:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Exact** | | **Poisson Approximation** | | **R simulation (25000 replications)** | |
| n | k=4 | k=5 | k=4 | k=5 | k=4 | k=5 |
| 4 | 0.0000000 |  | 0.0028 |  |  |  |
| 5 | 0.01666667 | 0.0000000 | 0.0165 | 0.0004 | 0.0174 |  |
| 6 | 0.03055556 | 0.00277778 | 0.0301 | 0.0028 | 0.0301 | 0.0032 |
| 7 | 0.04444444 | 0.00515873 | 0.0435 | 0.0052 | 0.0462 | 0.0058 |
| 8 | 0.05833333 | 0.00753968 | 0.0567 | 0.0075 | 0.0606 | 0.0089 |
| 9 | 0.07183642 | 0.00992063 | 0.0697 | 0.0099 | 0.0722 | 0.0095 |
| 10 | 0.08520117 | 0.01230159 | 0.0825 | 0.0122 | 0.0823 | 0.0134 |
| 11 | 0.09836806 | 0.01466991 | 0.0952 | 0.0146 | 0.0976 | 0.0145 |
| 12 | 0.11134204 | 0.01703439 | 0.1076 | 0.0169 | 0.1128 | 0.0163 |
| 13 | 0.12412981 | 0.01939305 | 0.1200 | 0.0193 | 0.1234 | 0.0182 |
| 14 | 0.13673594 | 0.02174688 | 0.1321 | 0.0216 | 0.1375 | 0.0211 |
| 15 | NA | NA | 0.1441 | 0.0239 | 0.1502 | 0.0217 |
| 16 | NA | NA | 0.1559 | 0.0262 | 0.1607 | 0.0276 |
| 17 | NA | NA | 0.1675 | 0.0286 | 0.1729 | 0.0278 |
| 18 | NA | NA | 0.1790 | 0.0309 | 0.1864 | 0.0329 |
| 19 | NA | NA | 0.1903 | 0.0332 | 0.1984 | 0.0327 |
| 20 | NA | NA | 0.2015 | 0.0355 | 0.2072 | 0.0352 |

Table 1 Reference probabilities for run up or run down of length k in a series of length n. Source for Exact and Poisson Approximation: Olmstead (1945), tables 2 and 3. Shaded cells in table calculated by Little using Olmstead’s Poisson approximation

As the probabilities in Table 1 are two-sided, divide by two to obtain the reference probablity for a run up or a run down of length k, given n.

The table suggests that the chance of k=4 runs in the good direction arising from chance—the Extranet implementation--is approximately bounded by 0.05 for n <=11 or 12. Chances of seeing a run up or down at random with k=5 are much smaller, by a factor of about six in the range on n values shown in the table..

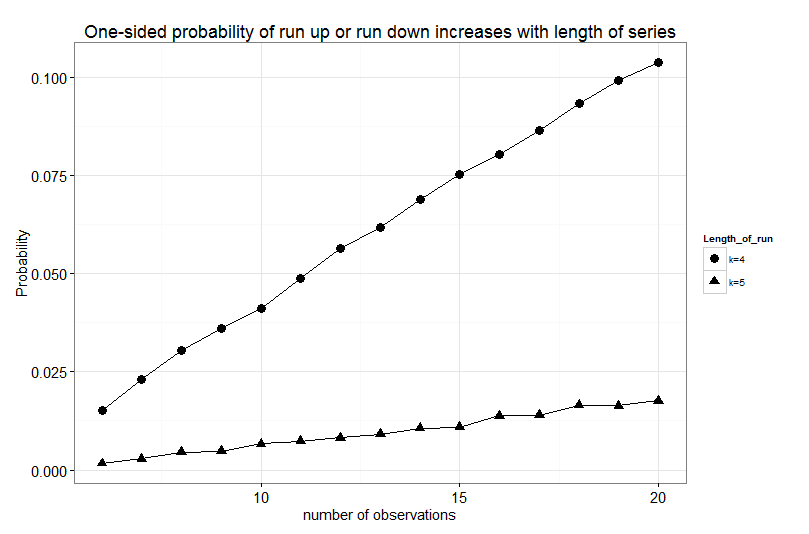


Figure 4 Estimated probability of seeing a run up or run down of length 4 or 5 (the “better” d irection) by chance in a random data series. Note the value k=4 is used by the Extranet. Probabilities estimated from R simulation.

Shift

The shift rule requires a median as reference. There are two cases in practice.

Case 1: No data are available as baseline prior to project; you gather data as you organize and test change ideas. Compute median for entire data series and look for a shift signal, k consecutive points on one side of the median.

In case 1, we use this “no improvement” reference model for probabilities: determine the frequency of sequences of length k or greater in a vector of length n, n even, where half the values are +1 and half the values are -1. (Any vector of length n, n even, will be divided by the median into n/2 values above the median, labeled +1 and n/2 values below the median, labeled -1.)

We need n>= 2k points to have k points on one side of the median in this model.

Now calculate or estimate the the probability of seeing e.g. k=6 points in sequence on the side of the median that is the good direction, in n values. We use the range 12 <= n <=20, ,matching Provost’s advice. Table 2 shows that beyond n=20, the one-sided probability of seeing a signal in random data continues to grow beyond 0.05

|  |  |  |
| --- | --- | --- |
| **n** | **Two-sided (10000 simulations)** | **One-sided (“good direction”)** |
| 12 | 0.02 | 0.01 |
| 14 | 0.03 | 0.015 |
| 16 | 0.06 | 0.03 |
| 18 | 0.08 | 0.04 |
| 20 | 0.10 | 0.05 |
| 22 | 0.13 | 0.06 |
| 24 | 0.16 | 0.08 |
| 26 | 0.19 | 0.09 |
| 28 | 0.21 | 0.11 |
| 30 | 0.23 | 0.11 |

Table 2 estimated probabilities for k=6, reference model “no improvement” using the R simulation model zero.

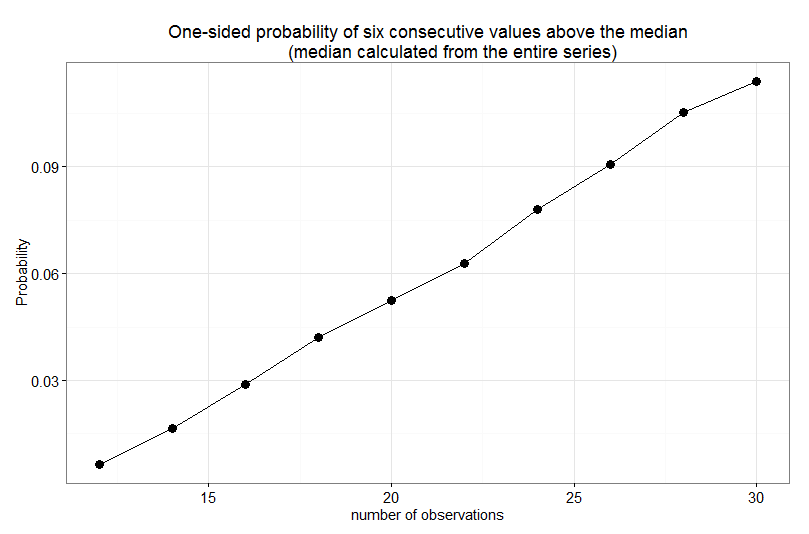


Figure 5 Estimated probability of seeing six consecutive values on one side of the median (the “better” d irection) by chance in a random data series. The median is computed from the entire data series. Probabilities estimated from R simulation.

Case 2: Baseline available before interventions based on nb observations. Compute median for baseline data; plot subsequent points and check for shift signals in the nb+1th, nb+2th, …, 20th values relative to the baseline median.

Our context: If you have more than 20 observations, you should consider an appropriate Shewhart chart.

Reference model for probablities: probability of seeing 6 consecutive heads in a series of

20 – nb coin tosses of a fair coin, where heads represents the good direction. Under the assumption of “no improvement” and the system continues to operate substantially after baseline period as during the baseline period, each subsequent point will be above or below the median with probability 0.5.

In order to see 6 values above the baseline median and restrict the series to 20 values, nb must be no more than 14.

|  |  |  |  |
| --- | --- | --- | --- |
| **n** | **nb** | **Two-sided (10000 simulations)** | **One-sided (“good direction”)** |
| 6 | 14 | 0.03 (exact: 0.03125) | 0.015 (exact: 0.15625) |
| 7 | 13 | 0.05 | 0.025 |
| 8 | 12 | 0.06 | 0.03 |
| 9 | 11 | 0.08 | 0.04 |
| 10 | 10 | 0.10 | 0.05 |
| 11 | 9 | 0.11 | 0.06 |
| 12 | 8 | 0.12 | 0.06 |
| 13 | 7 | 0.14 | 0.07 |
| 14 | 6 | 0.15 | 0.08 |
| 15 | 5 | 0.17 | 0.09 |
| 16 | 4 | 0.18 | 0.09 |
| 17 | 3 | 0.19 | 0.10 |
| 18 | 2 | 0.21 | 0.10 |

Table 3 Model with baseline median; probability of seeing six or more consecutive values on one side of the median in random data grows with series length n and crosses the 0.05 level at about n =10 or 11. Table values derived from R simulation model B.

The simulated one-sided probablities are shown in Figure 6.

Comparing Table 3 to Table 2, we see that the shift of 6 rule applied to a baseline median is at least twice as likely to occur at random than when the median is calculated from all the values.

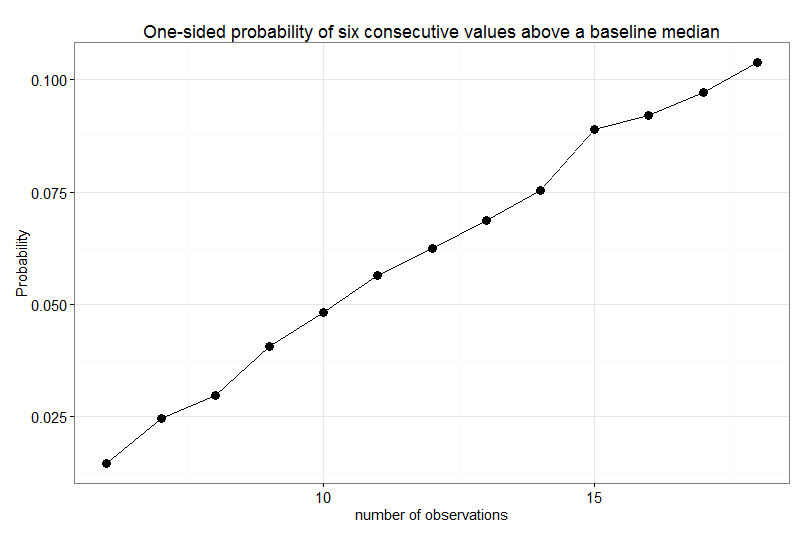


Figure 6 Estimated probability of seeing six consecutive values on one side of a baseline median (the “better” d irection) by chance in a random data series. Probabilities estimated from R simulation.

## References

Kahneman, D (2011), *Thinking, Fast and Slow*, Farrar, Straus and Giroux, New York.

Olmstead P, “Distribution of Sample Arrangements for Runs Up and Down,” *Annals of Mathematical Statistics*, 1945, March, **17**, pp 24–33

Perla, R J, Provost, L P, Murray, S K, “The run chart: a simple analytical tool for learning from variation in healthcare processes”, *BMJ Quality & Safety*, 2011, **20**, pp. 46-51 doi:10.1136/bmjqs.2009.037895

Provost, L P and Murray, S K (2011), *The Health Care Data Guide: Learning from Data for Improvement,* Jossey-Bass, San Francisco, chapter 3.

## Appendix 1: Background Comments Provost and Little

### Lloyd Provost’s comments in January 2015 emails:

* + The context for run chart use: looking for confirmation that interventions have made a difference in performance (answer to second question of M for I: how will we know that a change is an improvement?)
  + Run charts should be used as a simple guide in early stages of an improvement project with limited data, i.e. 8 or 10 to 20 observations in sequence.
  + Beyond 20 observations, Lloyd recommends an appropriate Shewhart control chart in most applications to display data and provoke insights.
  + The “too few runs/too many runs” table is hard for many people to apply correctly and motivated the simpler rules to detect shifts or trends as aids for improvers.

### KL notes:

* plotting data in time order is always appropriate and potentially useful, as a starting point for analysis.
* The simpler the causal system (smaller scope in time, space and relationships), the lower the technical complexity of tools needed to gain insight and gauge improvement.
  + The lowest-tech tool is just human observation and reflection. However, we typically need help to guide our thinking; logic supported by data tools provide that help (see Kahneman’s System 1 and System 2 dichotomy in *Thinking, Fast and Slow*).

## Appendix 2: Degree of belief from a Bayesian Viewpoint

(embedded spreadsheet, click to open and vary the first three cells in the Value column to see effect of changing the ingredients to the Bayes arithmetic.)



Increasing posterior degree of posterior belief: trade off between reference signal (P(s|not C) and P(s|C) and P(C ). P(C ) should be more than 0.5 (1:1 odds that we are causing improvement).

If you make the signal criterion more stringent –lower P(s|not C)—then when you see a signal, you should have higher belief that improvement caused the signal for any fixed level of P(s|C).

For a given signal threshold P(s|not C) , an increased probability P(C) that your change will cause improvement (e.g. based on evidence from other improvement effort) also increases P(C|s) the posterior probability that changes caused improvement, given the signal.

Finally, if you can increase the sensitivity of the chart to signals when there is improvement, P(s|C), you also increase P(C|s) for any fixed level of P(s|not C).

The table suggests you can be willing to allow for a higher level of false positives e.g 0.1 rather than 0.05 or 0.01—P(s|C), chart signals when there is no improvement—if either P(C) or P(C|s) or both are high.