Paper AR04

A Simulation-Based Approach to Establish Quality Tolerance Limits

Debbie Adkins, Phastar, Manchester, UK
Susan Lovick, Phastar, Manchester, UK

ABSTRACT

Over recent years advances in technology have provided the promise of improving the conduct & oversight of clinical trials. Across industry, multi-disciplinary teams are working closely together to reap the rewards of risk based monitoring approaches. However, whilst the potential advantages of Risk Based Monitoring (RBM) are clear, the best way to successfully embed RBM in clinical trials is still emerging. This paper describes an example of how we created a quality threshold limit for protocol deviations using simulated data based on previous studies, using an approach which is used in statistical process control. We created simple, easy to understand limits which are effective in detecting unusual levels of protocol deviations at one site or at one visit. Although unusually high levels of deviations may indicate an issue at that site which may need to be quickly addressed, it is also important to be able to detect unusually low levels of deviations as this may indicate underreporting. Our limits were able to do both. Using such limits, any issues identified can be investigated in real time, giving the best chance of determining the root cause.

INTRODUCTION

Quality Tolerance Limits (QTLs) form a key part of the RBM approach. A QTL is defined as a level, point, or value associated with a trial variable that should trigger an investigation if a deviation is detected, in order to determine if there is a possible systematic issue (i.e. a trend has occurred).

QTLs are monitored at the trial level and are pre-defined before the trial commences from a review of historical data from similar trials and where possible, using statistical methods and modelling.

If there is limited historical information available about the trial variable of interest, difficulties may arise in developing a QTL. In order to overcome the issue of limited real historical data, we used a simulation approach to generate 'historical' data and develop a QTL for Protocol Deviations. For the purposes of our work, we used only major protocol deviations, but a similar approach could be used to create a QTL for minor protocol deviations or for the total. We then simulated further data to test our limits.

CREATING A QTL

BACKGROUND DATA

Throughout this paper, the term 'PD rate' is used to describe the rate of major protocol deviations (PDs) per subject-visit. We use a rate, rather than the actual number, to allow for differences in the number of subjects at each site. Major PDs are as classified by the monitor from an operational perspective, rather than at database lock by the study team to reflect the decision making if the data were being monitored in real time.

The PD rate per subject-visit for each site is defined as

$$\frac{\text{total number of PDs at that site}}{\text{number of subject visits at that site}}$$

Actual historical data on PD rates, falling into each of 5 categories, were available for 10 sites, as follows:

Table 1: Historical data on major PDs

	Number of Major PDs									
Site	Category 1	Category 2	Category 3	Category 4	Category 5	Total count of major PDs	Number of subjects	Subject visits	Overall PD rate per subject by site	Overall PD rate per subject-visit by site
201	11			1		12	22	94	0.55	0.13
202	3		4	9		16	15	63	1.07	0.25
203	6	4	5	2		17	15	71	1.13	0.24
204		5		2		7	9	47	0.78	0.15
205	1	3	10	3		17	12	53	1.42	0.32
206				1		1	4	13	0.25	0.08
207	3	9	3	5	1	21	8	41	2.63	0.51
208		1	1	1		3	3	12	1.00	0.25
209			3			3	7	25	0.43	0.12
210		3	1			4	1	3	4.00	1.33
Total	24	25	27	24	1	101	96	422		

The rate can be seen to vary widely between sites, and in particular the rate at site 210, with only 1 subject but 4 PDs over 3 visits, may not be a reliable estimate of the actual underlying PD rate at that site. For this reason, care must be taken in deciding whether to incorporate data from very small sites or relatively low subject-visits.

Although only the overall rate of PDs is shown in the table above, the PD rates for each of the individual categories may also be calculated by dividing the total number of PDs in that category by the total number of subject visits. These rates for the individual categories will be used later in the simulation step.

SIMULATING HISTORICAL DATA

To generate the historical data which will be used to create a QTL, a trial is initially simulated involving 500 subjects over 20 sites at each of 5 visits. The subjects are assumed to be unevenly distributed across the sites, and so the first step is to simulate the number of subjects at each site. We put a lower limit of 10 on the number of subjects per site to try to avoid potential issues such as those seen above with site 210.

The 'RANDBETWEEN' function in Excel is used to generate the number of subjects for each of 20 sites. We found that a lower limit of 10 and an upper limit of 40 gave appropriate subject numbers. It is unlikely that the total over all the sites would exactly add up to 500, and so following the RANDBETWEEN step, the numbers of subjects at each site may be slightly inflated / deflated by a common factor to ensure the subject numbers across the whole simulated study add up to 500.

Having generated the numbers of subjects at each site, we can then simulate the protocol deviations across the study for each of the 500 subjects within the study.

We assigned each subject a unique number and then created a SAS dataset with 2500 rows, each subject repeated 5 times, once for each visit. Five columns correspond to the number of PDs in each of the five categories. For each subject at each visit, a flag of 0 or 1 for each category of PD was simulated. To do this, we use the Bernoulli distribution, which takes the value 1 with probability p, and the value 0 with probability 1-p. A flag of 1 in the simulated dataset meant that that subject had that type of PD at that visit, a flag of 0 meant they did not. Note that our method here makes the assumption that all subjects at all sites and at all visits are equally likely to have a PD from a given category.

The SAS code to populate the columns of the dataset is as follows.

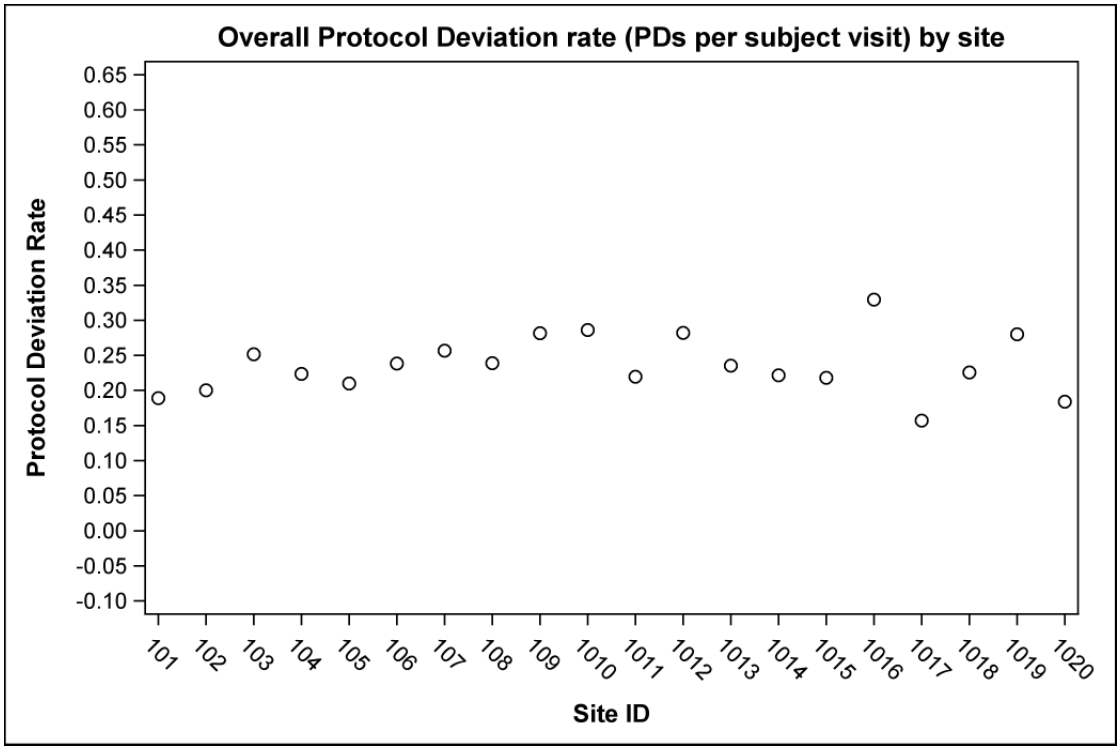
```
data allsites;
      set allsites;
pd_cat1=rand("Bernoulli",0.057);
pd_cat2=rand("Bernoulli",0.059);
pd_cat3=rand("Bernoulli",0.064);
pd_cat4=rand("Bernoulli",0.057);
pd_cat5=rand("Bernoulli",0.002);
pd_total= pd_cat1 + pd_cat2 + pd_cat3 + pd_cat4 + pd_cat5;
      run;
```

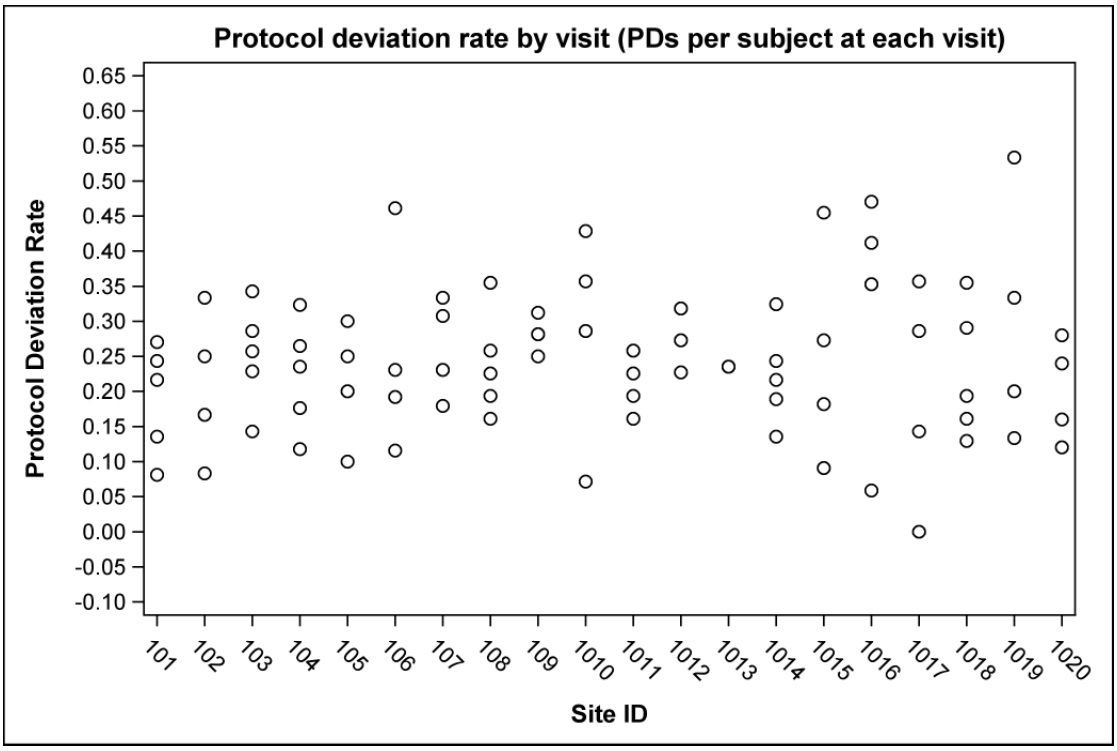
The rates in the Bernoulli simulation step are the PD rate for the relevant category as estimated from the historical data above.

An excerpt of the simulated dataset is shown below.

	usiteid	usubjid	visitno	pd_cat1	pd_cat2	pd_cat3	pd_cat4	pd_cat5	pd_total
204	102	102_12	4	0	1	0	0	0	1
205	102	102_12	5	0	0	0	0	0	0
206	102	102_13	1	0	0	1	0	0	1
207	102	102_13	2	1	0	0	0	0	1
208	102	102_13	3	0	0	1	0	0	1
209	102	102_13	4	0	0	0	0	0	0
210	102	102_13	5	0	0	0	1	0	1
211	102	102_14	1	0	0	0	0	0	0
212	102	102_14	2	0	0	0	0	0	0
213	102	102_14	3	0	0	0	0	0	0
214	102	102_14	4	0	0	1	1	0	2
215	102	102_14	5	0	0	0	0	0	0
216	102	102_15	1	0	0	0	0	0	0
217	102	102_15	2	1	0	0	0	0	1
218	102	102_15	3	0	0	0	0	0	0
219	102	102_15	4	0	0	0	0	0	0
220	102	102_15	5	0	0	0	0	0	0
221	102	102_16	1	0	1	0	0	0	1
222	102	102_16	2	0	0	0	0	0	0
223	102	102_16	3	0	0	0	0	0	0
224	102	102_16	4	0	0	0	0	0	0
225	102	102_16	5	0	0	1	0	0	1
226	102	102_17	1	0	0	1	0	0	1
227	102	102_17	2	0	0	0	0	0	0
228	102	102_17	3	0	0	0	0	0	0

The overall rate of PDs at each site across all visits may be calculated using the formula PD Rate = (total number of PDs at that site/number of subject-visits at that site). The average rates by site across all visits, and the rate per site per visit, are plotted below.


Overall Protocol Deviation rate (PDs per subject visit) by site
Protocol Deviation Rate
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
101
102
103
104
105
106
107
108
109
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
Site ID

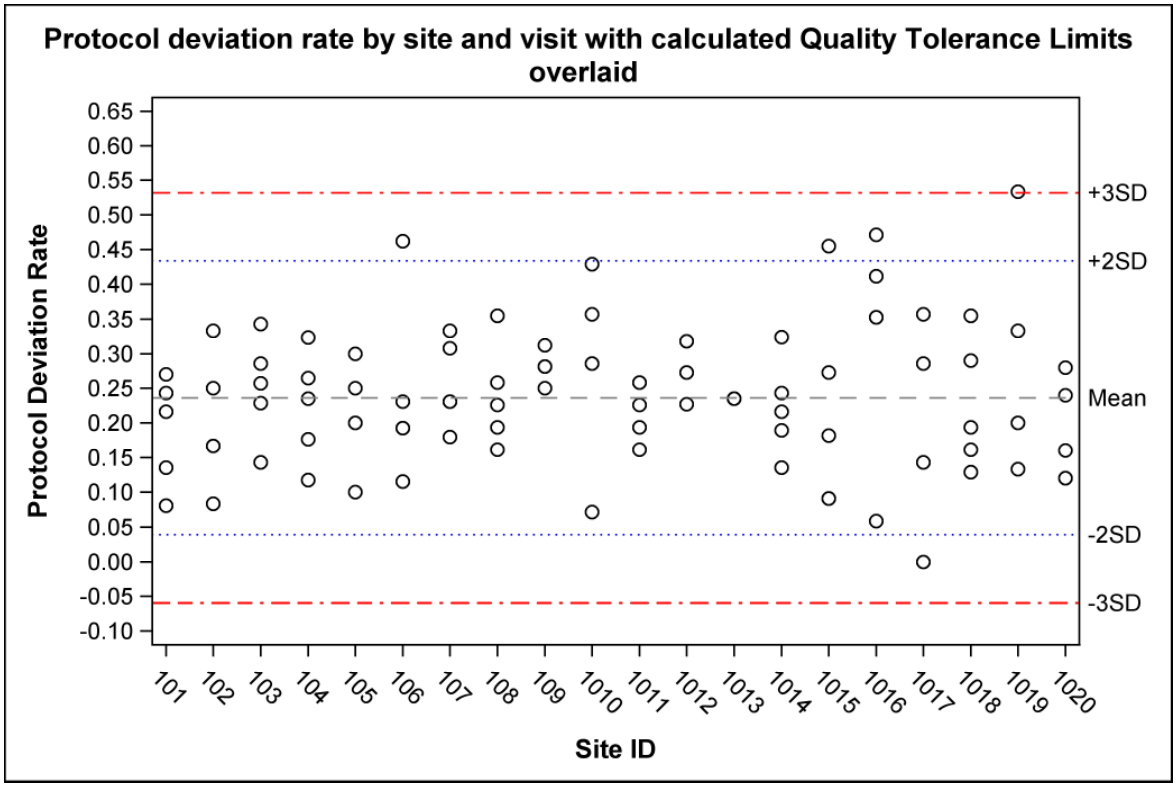

Protocol deviation rate by visit (PDs per subject at each visit)
Protocol Deviation Rate
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
101
102
103
104
105
106
107
108
109
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
Site ID

CALCULATING THE QTL

Basic principles of statistical process control may be used to create QTLs for the rate of PDs at each site. By calculating the overall mean and the standard deviation across sites and visits (i.e. across the 100 data points displayed in the figure above), limits on the overall number of PDs for each site can be computed. These limits are fixed at Mean +/- 3 Standard deviations (SD). 'Warning' limits may also be computed at Mean +/- 2 SD.

In our simulated data, the overall mean was calculated as 0.236 i.e. there were 0.236 major protocol deviations per subject per visit across all sites.

These lead to limits of (-0.059, 0.532), with the 2SD warning limits of (0.039, 0.433). For all practical purposes, since there cannot be a negative number of PDs, the lower limit will be 0.



TESTING THE QTL

INFLATING THE PD RATE

We can test the QTLs which we have just created using a similar simulation approach. First, we test whether the limits can pick up an abnormally high rate of PDs at a site.

To begin with another set of 20 sites is simulated, with between 10 and 40 subjects at each, and totalling 500 subjects across the trial, as before.

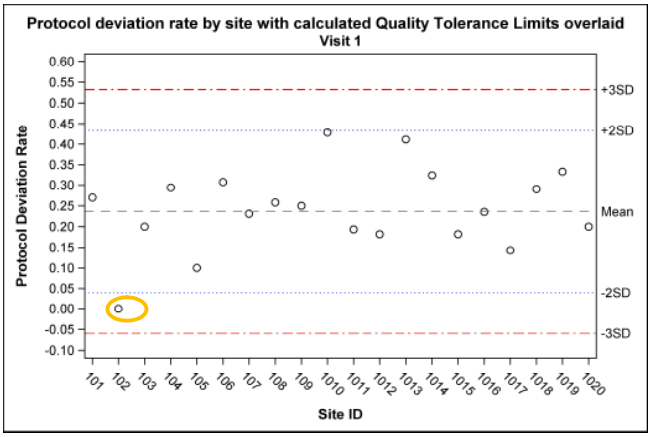
Following the same method used to create the QTL, a PD flag is generated using the Bernoulli function. This time we simulate a study with 6 visits. A 6-visit study design is chosen simply to demonstrate that the design of the historical study does not have to match the design of the new study. The number of PDs for each subject at each of those visits for each PD category is simulated, and the totals summed.

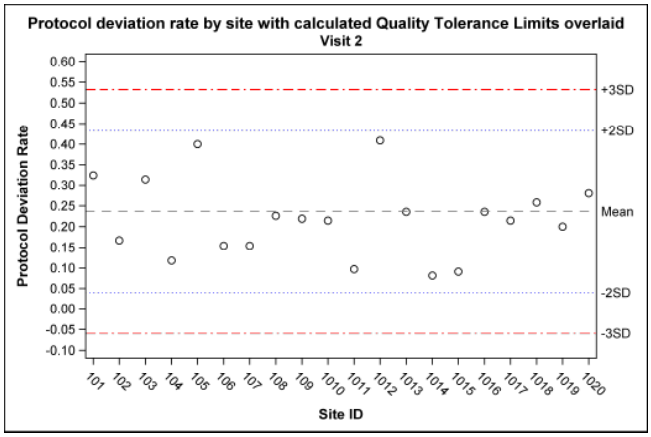
Now using this new simulated dataset, one of the types of PD may be artificially inflated to mimic an issue arising at a site. Suppose category 2 relates to study procedure, and site 107 develops an issue with some aspect of the study procedure at visit 3 which leads to 40% of subjects having a major PD in that category at that visit. The data for this site at this visit is generated as follows:

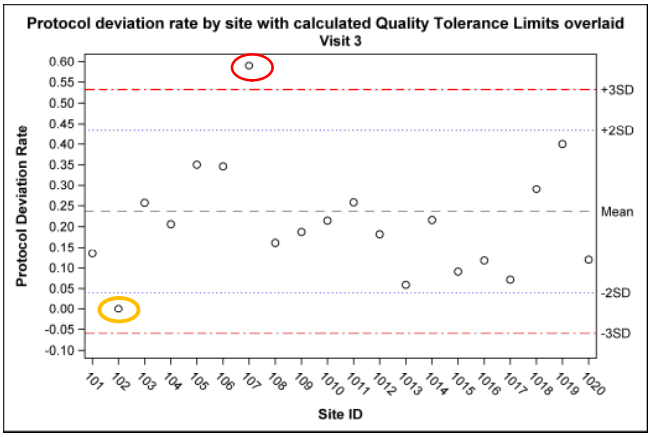
```
data newsites1;
       set newsites;
       if usiteid=107 and visitno=3 then do;
              pd_cat2=rand("bernoulli",0.4);
              pd_total= pd_cat1 + pd_cat2 + pd_cat3 + pd_cat4 + pd_cat5;
       end;
run;
```

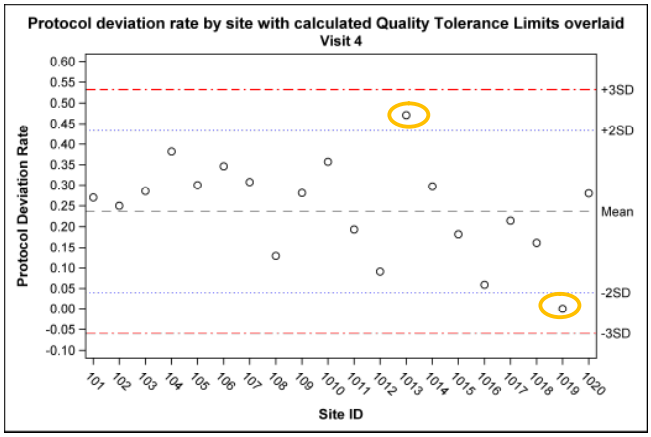
Having generated the data for this new simulated trial, the overall rates of PDs are calculated in the same way as before using the formula PD Rate = (total number of PDs at that site/number of subjects at that site) for each visit. These rates may be plotted against the QTL calculated above. Note that the QTLs are based on the historic data, in this case the original simulated data. They are not recalculated once new data from the current study become available.

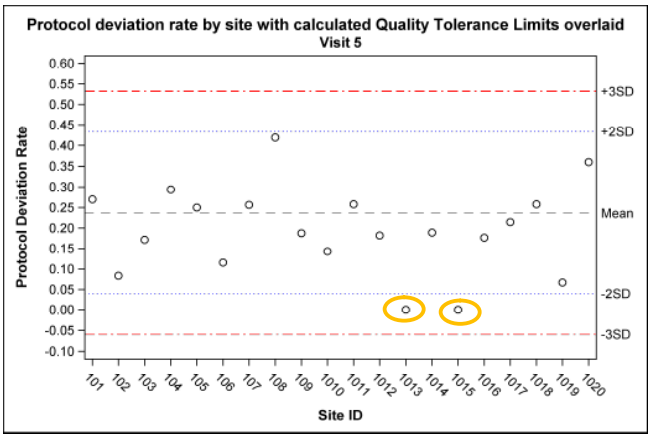
In traditional approaches to monitoring data from clinical studies, data have been reviewed at the end of the study, but here the data are monitored in real time, and for each site and each visit, the rate of PDs is compared to the QTL.

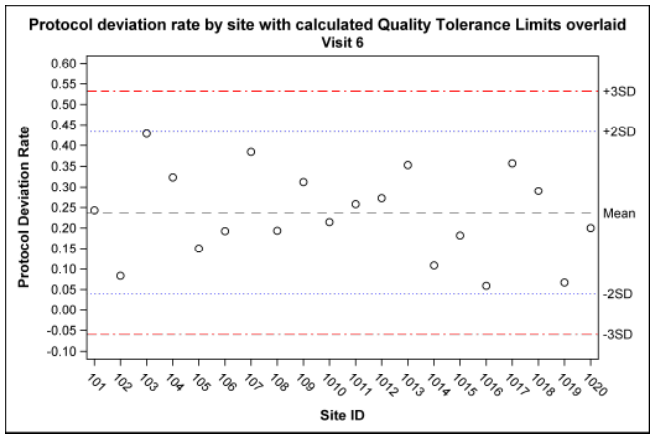












Looking at the protocol deviations on a visit-by-visit basis, the rate at site 107, visit 3 (marked in red), can be clearly seen as falling above the upper QTL. This should then lead to an investigation into the rate of major PDs at this site.

We also note a number of points outside of the 2SD warning limits (marked amber); however, there is no pattern to these points and so they would generally not trigger any form of investigation.

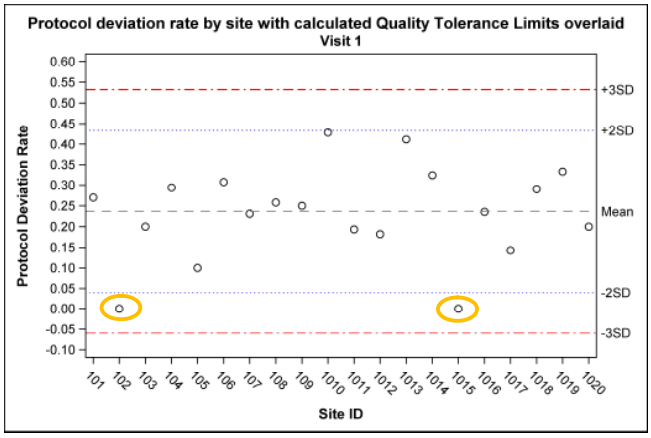
UNDERREPORTING

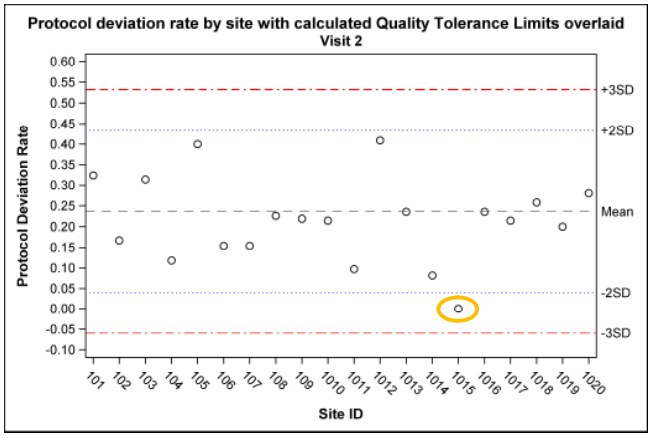
Now we go back to our newly simulated data, but instead of artificially inflating the result at one site, we artificially reduce the rate for all types of PD at one site, to model underreporting. The code to produce this simulated data is as follows.

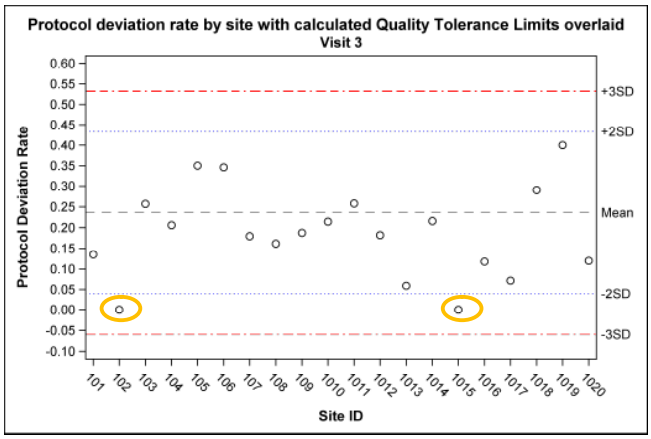
```
data newsites2;
      set newsites;
      if usiteid=1015  then do;
            pd_cat1=rand("bernoulli",0.01);
            pd_cat2=rand("bernoulli",0.01);
```

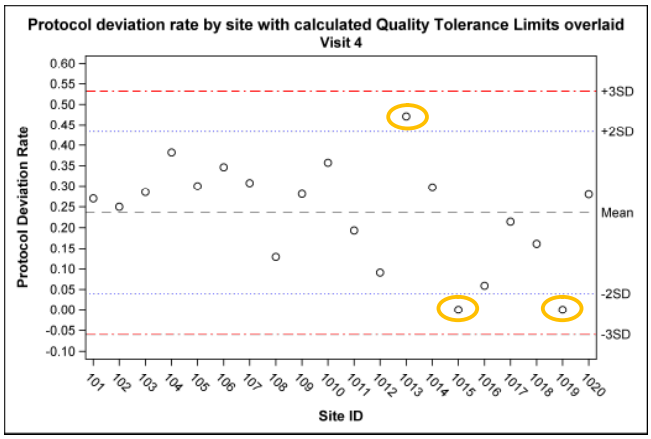
```
                pd_cat3=rand("bernoulli",0.01);
                pd_cat4=rand("bernoulli",0.01);
                pd_cat5=rand("bernoulli",0.01);
                pd_total= pd_cat1 + pd_cat2 + pd_cat3 + pd_cat4 + pd_cat5;
        end;
run;
```

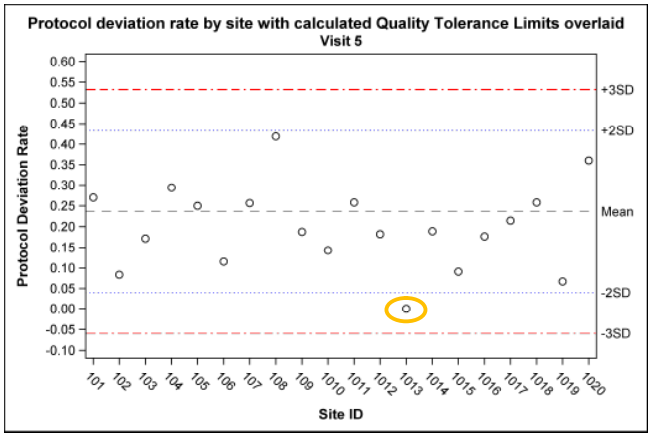
Again, we plot the data visit by visit against the QTLs calculated initially.

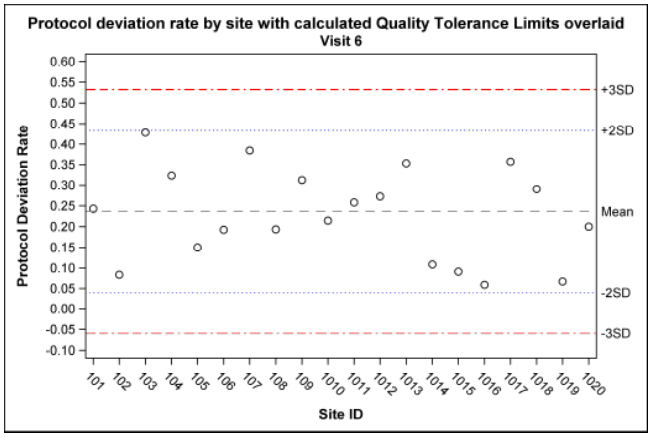












Although there are no breaches of the 3SD action limits in this simulation, there are a number of breaches of the warning limits, marked in amber. Some of these breaches would not be a cause for concern, but we note that site 1015 falls outside the lower warning limit for Visits 1-4. A pattern such as this should lead to an investigation of site 1015. It is likely that the investigation would be triggered after the second or third consecutive visit outside the warning limits.

Site 102 has two breaches of the lower warning limit, at visits 1 and 3, even though we know in this instance that the rate at site 102 has not been altered and these low results are just due to chance. It is worth noting however that site 102 is one of the smaller sites, with only 12 subjects. In fact, it can be noted that all those points outside the warning limits correspond to smaller sites, with the maximum size being 17 subjects. This emphasises the importance of setting a lower threshold on the number of subjects at each site, both when looking at actual historical study data to determine the rates of PDs, and when simulating data to calculate limits.

CONCLUSION

Further work was carried out, but is not presented here, to investigate how reliably issues such as those simulated above can be detected. Perhaps unsurprisingly, we discovered when artificially inflating the rate that the higher the inflated rate, the more likely an issue will be picked up by the limits.

As noted above, smaller sites may potentially show some indications of an issue, even where there is none. Part of any investigation into the cause of a breach of the limits needs to carefully take account of the size of the site.

QTLs calculated using this methodology will not be updated with emerging data from a new study and so it is important to ensure that they are reliable and are not derived using clear outliers. Care must be taken when including data from sites with few subjects and from sites where issues have been seen previously. However, a balance must be struck between excluding data which is potentially unreliable and ensuring there is sufficient data to calculate the QTLs.

ACKNOWLEDGMENTS

Thanks to Zoe Macintyre of Phastar for help with the SAS coding.

CONTACT INFORMATION

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

PHASTAR
Unit 2A 2 Bollo Lane
London W4 5LE

debbie.adkins@phastar.com
susan.lovick@phastar.com

www.phastar.com