

A Simulation Based Approach to Establish Quality Tolerance Limits

Debbie Adkins

Principal Statistician, PHASTAR

Background

- Historically paper CRFs have been monitored by source data verification, but advances in technology offer new opportunities to increase efficiency and focus on relevant activities.
- Interest in Risk Based Monitoring (RBM) is increasing and several guidance and consultation documents have emerged that encourage the use of risk based quality management systems.
- RBM aims to “identify, prioritize and control for risks based on probability, detectability, and impact within the conduct of a clinical trial”¹.

1) Risk-based quality management: Quality Tolerance Limits and Risk Reporting.
Transcelerate Biopharma Inc., 2017

Background

- Quality Tolerance Limits (QTLs) are an integral part of RBM and are used to monitor key parameters.
- They apply to the whole trial and are usually created following a review of historical data.
- If historical information is limited, it may be difficult to create a QTL.
- In this presentation, we use the historical data that is available, and simulate additional data in order to create a QTL.

Background

- Our example focuses on developing a QTL for the number of protocol deviations (PDs).
- We chose only major PDs, as classified by the monitor from an operational perspective.
- This reflects the decision making that would happen on an ongoing basis throughout the trial – so is representative of the data that would be available as the trial is monitored in real time.

Background data

- Throughout this presentation, the term 'PD rate' is used to describe the number of protocol deviations per subject-visit at a given site, and is defined as:

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

- This allows us to account for different numbers of subjects at each site in the trial.
- We use data from 10 sites from one historical study as our basis. At each site, the number of major PDs in each of 5 categories is given.

Background data

Site	Number of Major PDs					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
	Category 1	Category 2	Category 3	Category 4	Category 5					
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		

Across all sites and all visits, the PD rate is $101/422 = 0.239$

Background data

	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		

Note that site 210 only has data for 1 subject, with a total of 4 major PDs. This leads to a high PD rate for that site.

Background data

- Although only the overall PD rate is shown in the table, we can use the same principle to calculate a PD rate for each of the 5 individual categories of deviation.
- It is these individual rates that we use in the simulation steps that follow.

Category	1	2	3	4	5	Overall
Total PDs	24	25	27	24	1	101
Subject Visits	422	422	422	422	422	422
PD rate	0.057	0.059	0.064	0.057	0.002	0.239

Simulating historical data

- We first simulate some additional historical data, based on the actual data collected from these first 10 sites.
- A trial is constructed with 20 sites and 5 visits, and a total of 500 subjects across the 20 sites. The subjects are not evenly spread across sites to better reflect a real trial.
- We put a lower limit of 10 on the number of subjects at each site to try to avoid issues with small sites being unduly influenced by one subject.

Simulating historical data

- There are a number of ways that the numbers of subjects at each site could be simulated; we used the 'RANDBETWEEN' function in Excel, randomly generating a number between 10 and 40 and forcing the overall total to be 500.
- Having determined the number of subjects at each of the 20 sites, we then assign each subject a unique number and create an empty dataset in SAS which has one row for each subject at each visit, 2500 rows in total. Five columns of this dataset will correspond to the PDs in each of the five categories.

Simulating historical data

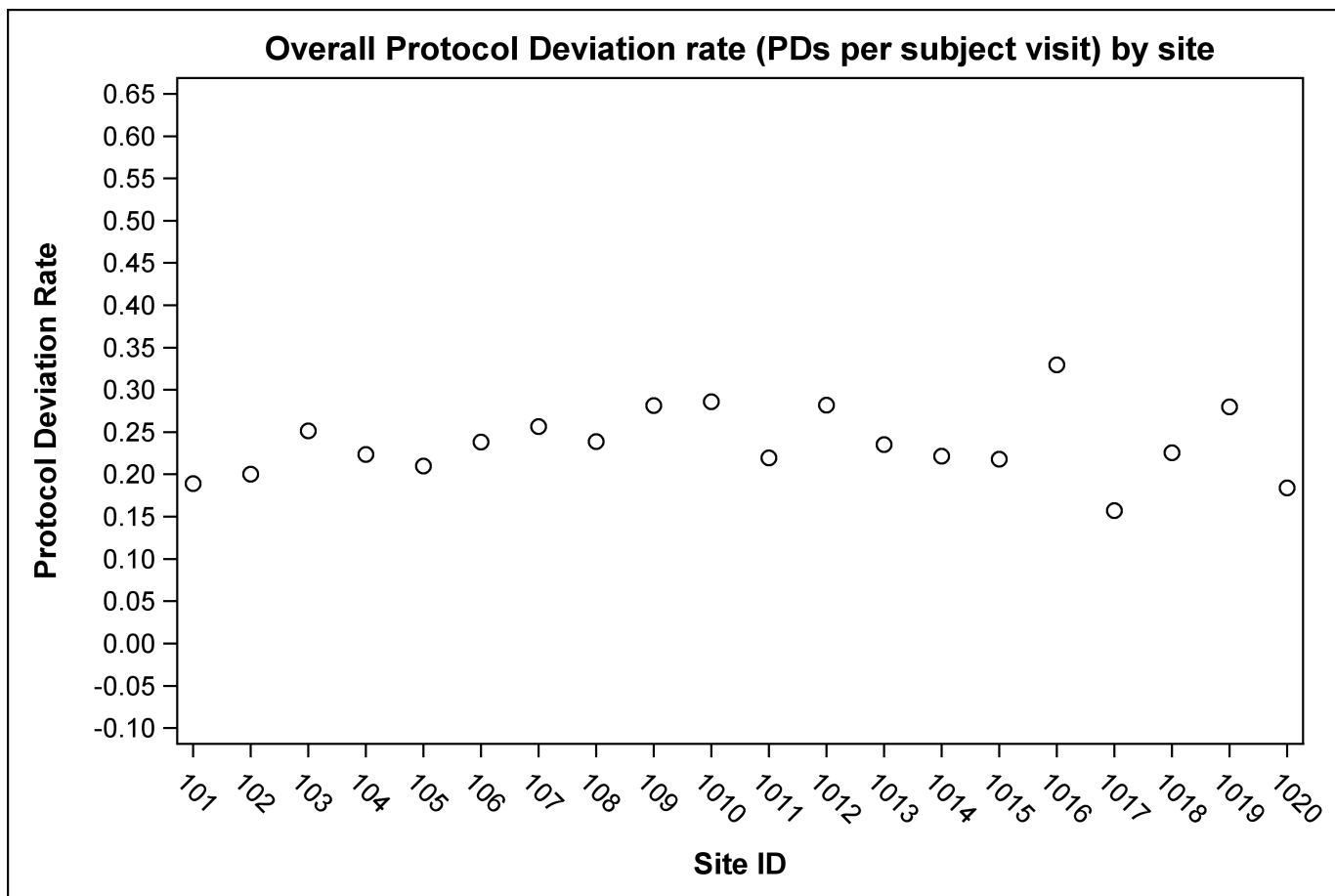
- For each subject at each visit, a flag of 0 or 1 for each category of PD was simulated.
- To do this, we use the Bernoulli distribution, which takes the value 1 with probability p , and the value 0 with probability $1-p$. We use the individual rate for each category of PD as the probability p in the simulation step. e.g.:

```
pd_cat1=rand("Bernoulli",0.057);
```
- 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.

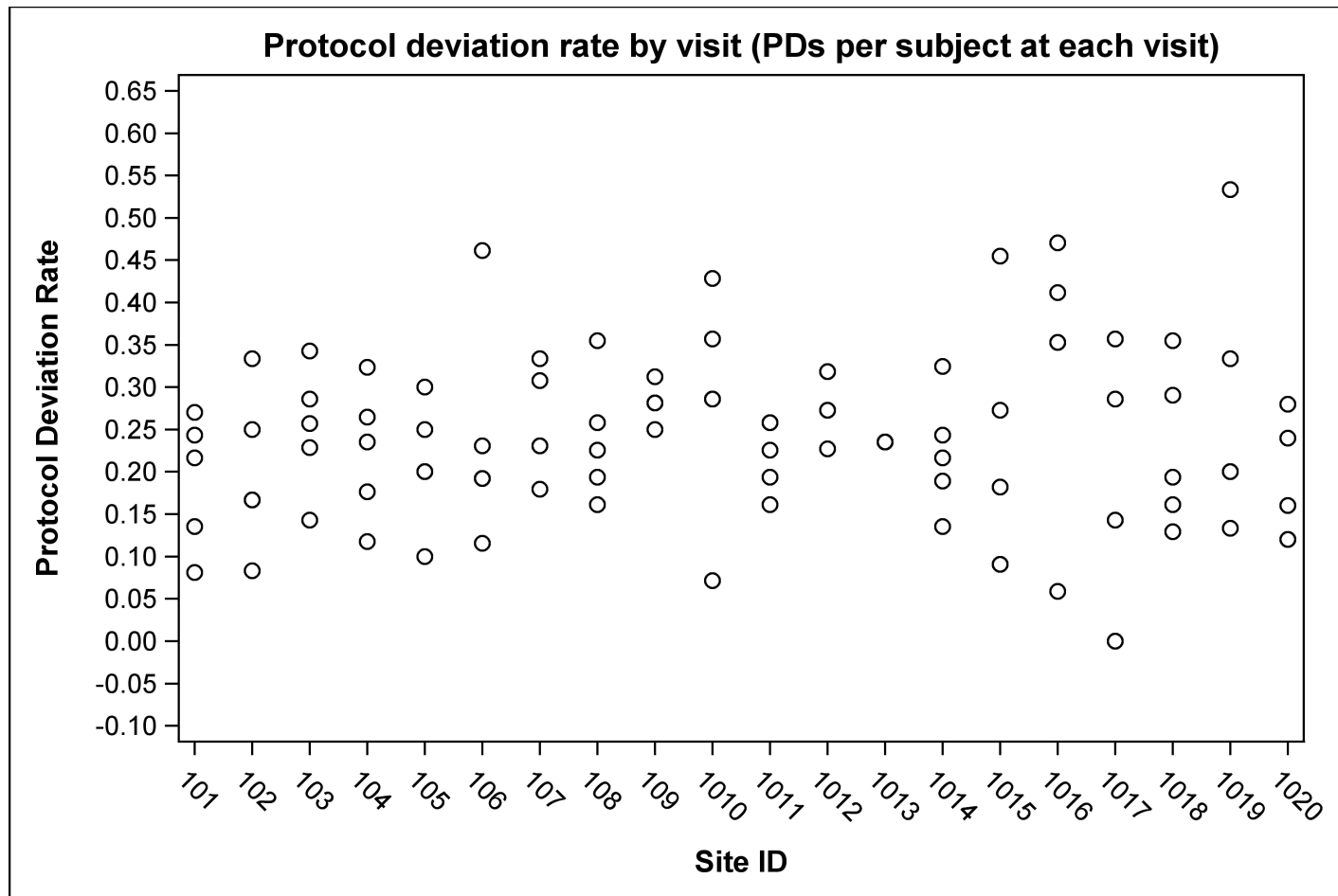
Simulating historical data

- Having generated a flag for PDs in each of the 5 categories, the 5 categories are summed to give the total PDs for each subject at each visit.
- We then average across the subjects at each site and for each visit.
- We can also calculate an overall rate per subject-visit for each site, as described earlier.

Simulated data – overall PD rate per site



Simulated data – PD rate per visit per site

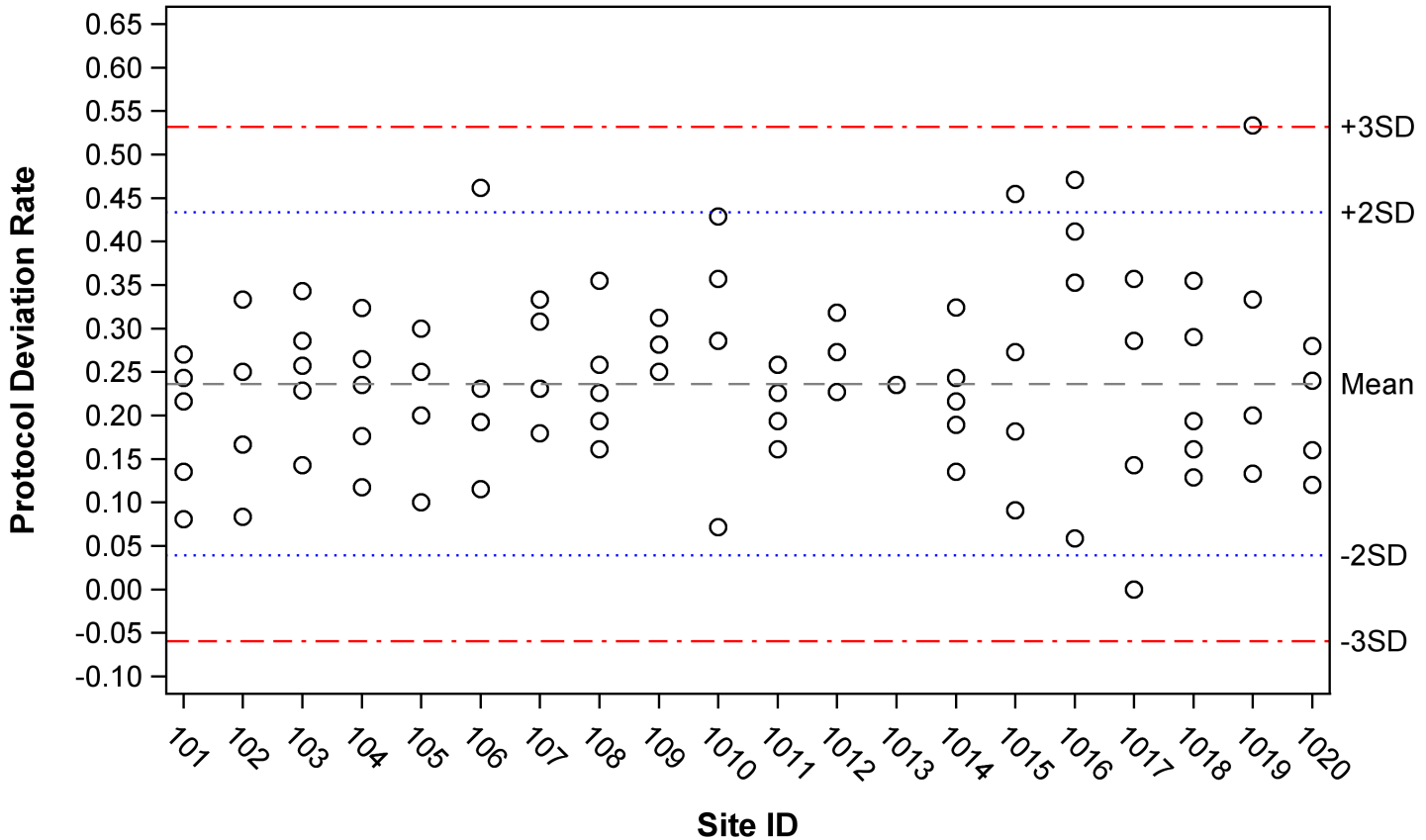


Generating the QTL

- To generate the limits that will be used going forwards, we calculate the mean and standard deviation (SD) of these 100 data points (20 sites x 5 visits).
- We set limits at the mean ± 3 SDs, with 'warning' limits at ± 2 SDs.
- In our example, the mean is 0.236 i.e. there were 0.236 major protocol deviations per subject per visit across all sites. We note that in the actual historical data used to set up the simulation, the overall mean was 0.239.
- 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.

Visualising the QTL

Protocol deviation rate by site and visit with calculated Quality Tolerance Limits overlaid



Testing the QTL

- Having created a QTL using simulated data, we can then test that it is able to detect unusual rates of PDs.
- As well as high rates of PDs, it is equally important to be able to detect unexpectedly low rates of PDs, which may indicate under-reporting.
- First, we test whether the QTL can pick up an unusually high rate of PDs.

Testing the QTL

- To test the QTL, we simulate data from a new trial.
- This time, we still use 20 sites, but now create a trial with 6 visits.
- We distribute a total of 500 subjects across the 20 sites, as before, and then simulate a PD flag (0 or 1) for each category using the Bernoulli distribution, setting the rates as before.

Testing the QTL

- We choose one site and one visit (in this instance, site 107 and visit 3) and we artificially inflate the rate of PDs at that site for that visit.
- We choose one category of PDs, and overwrite the PD flags for that category using a different Bernoulli function:

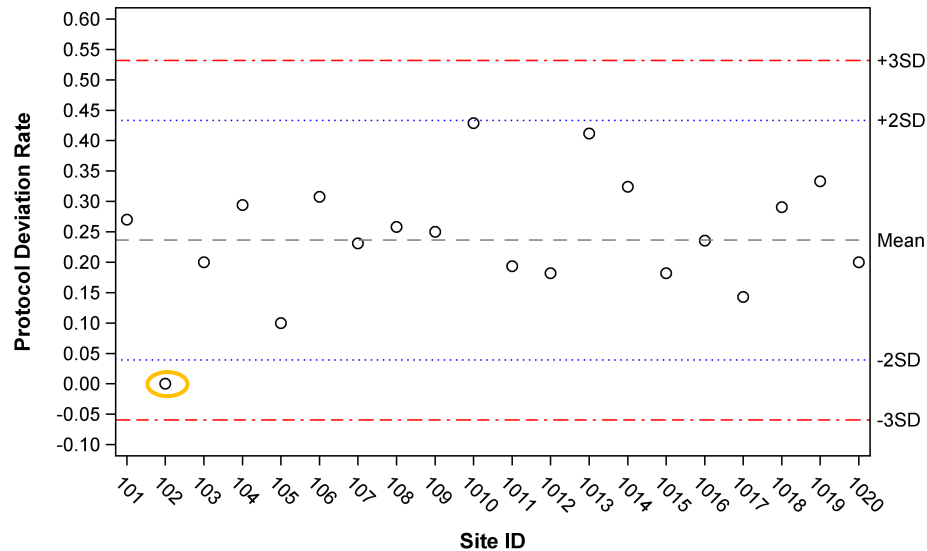
```
if usiteid=107 and visitno=3 then do;  
    pd_cat2=rand("bernoulli",0.4);  
    pd_total= pd_cat1 + pd_cat2 + pd_cat3 +  
              pd_cat4 + pd_cat5;  
end;
```

Testing the QTL

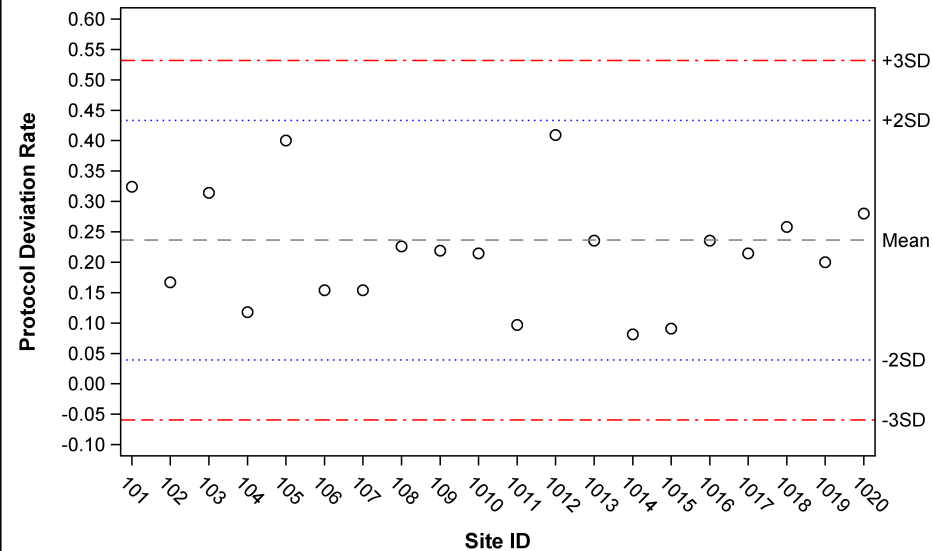
- Historically, the approach has been to look at data at the end of the trial.
- In RBM, we look at data on an ongoing basis.
- Consider the data from our new trial on a visit-by-visit basis.

Testing the QTL

Protocol deviation rate by site with calculated Quality Tolerance Limits overlaid
Visit 1

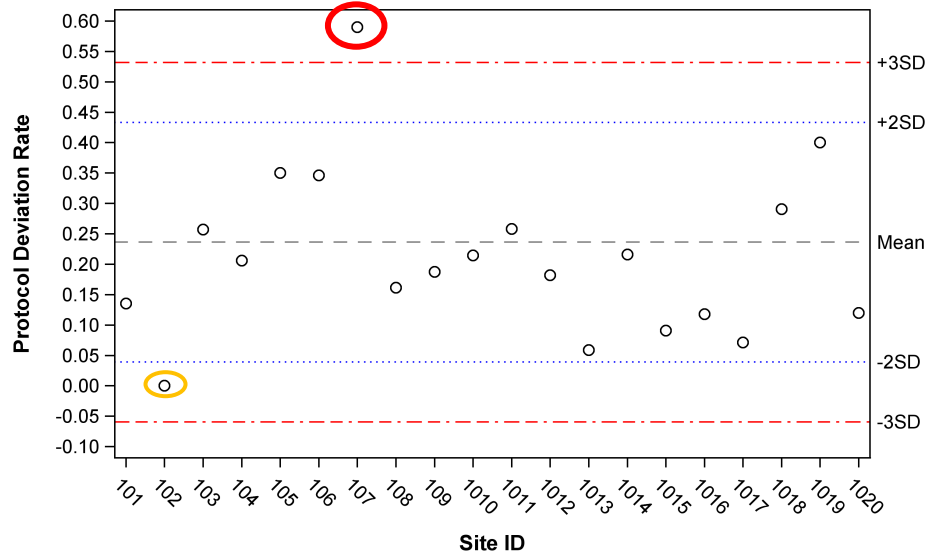


Protocol deviation rate by site with calculated Quality Tolerance Limits overlaid
Visit 2

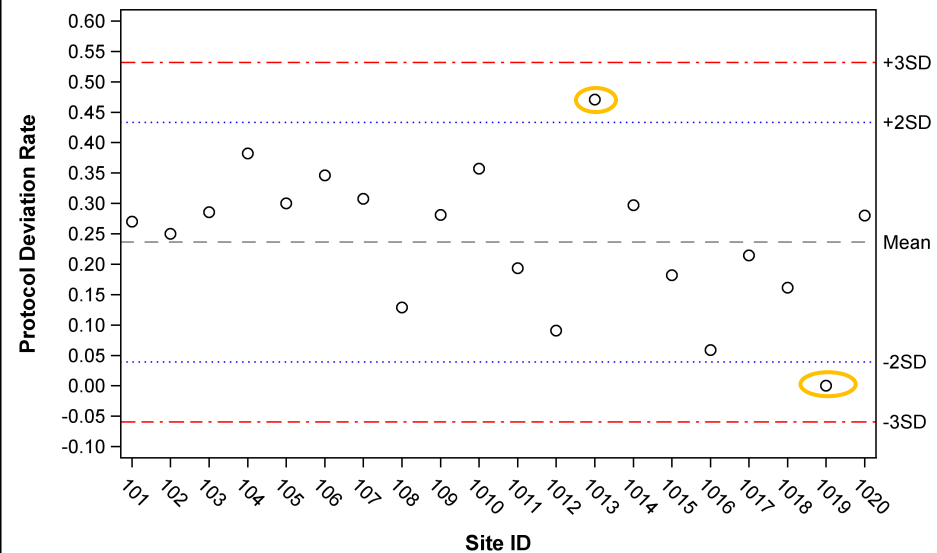


Testing the QTL

Protocol deviation rate by site with calculated Quality Tolerance Limits overlaid
Visit 3

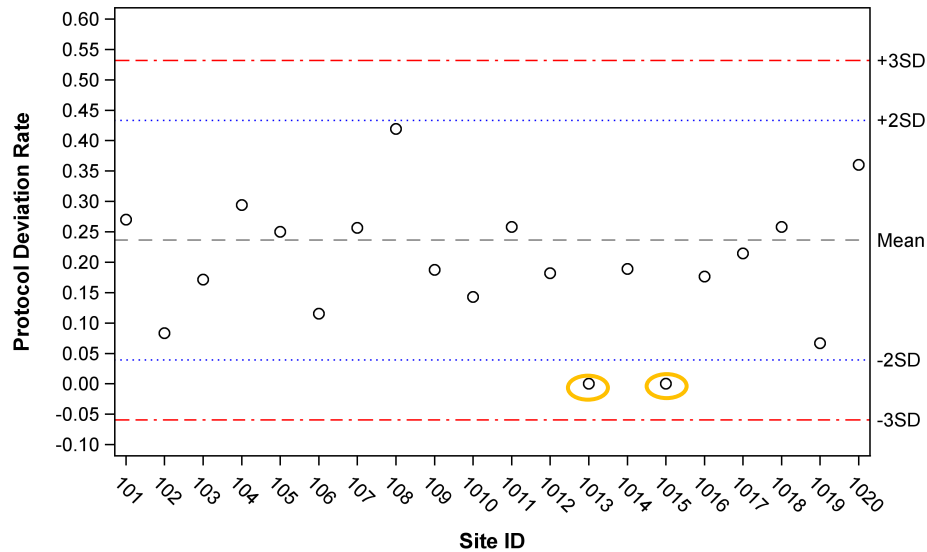


Protocol deviation rate by site with calculated Quality Tolerance Limits overlaid
Visit 4

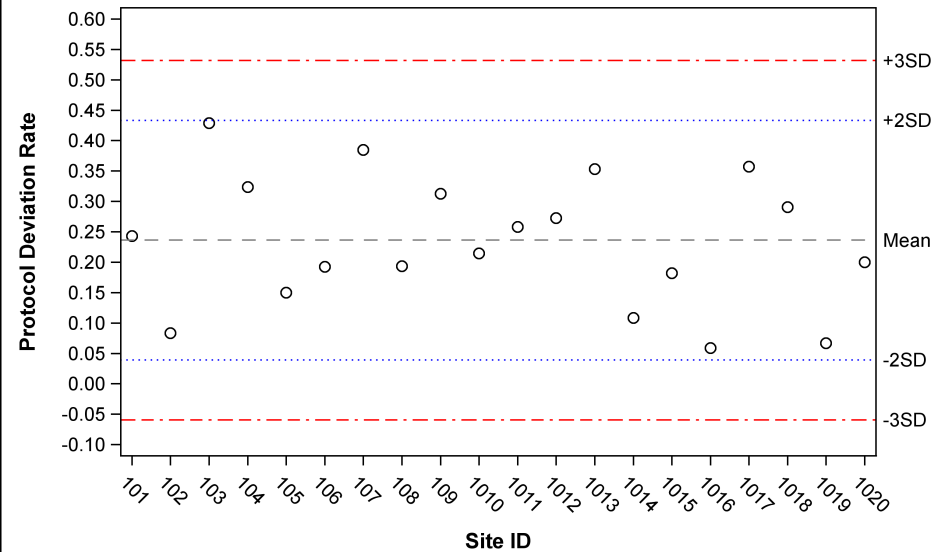


Testing the QTL

Protocol deviation rate by site with calculated Quality Tolerance Limits overlaid
Visit 5



Protocol deviation rate by site with calculated Quality Tolerance Limits overlaid
Visit 6



Testing the QTL

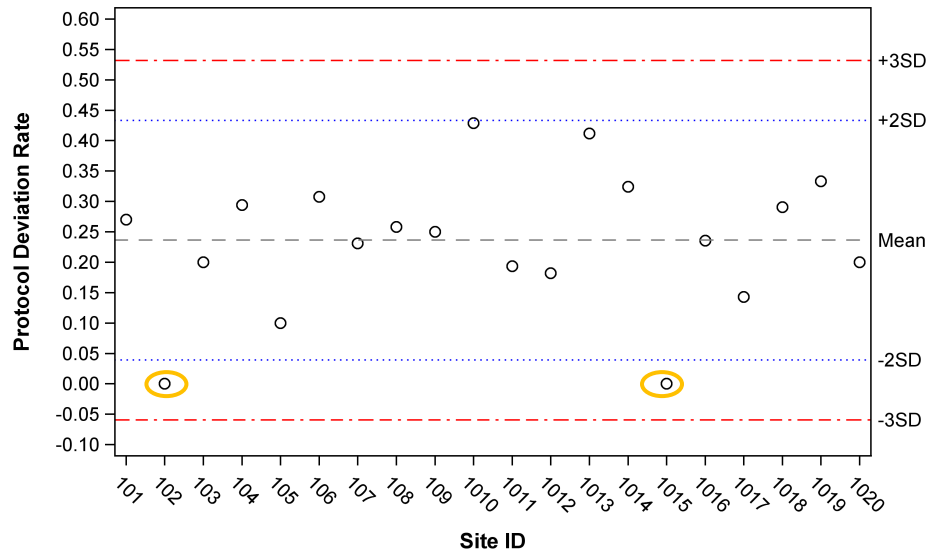
- Looking at the protocol deviations on a visit-by-visit basis, the rate at site 107, visit 3 (marked in red), can be clearly seen as falling above the upper QTL. This should then lead to an investigation into the rate of major PDs at this site.
- This investigation should be carried out as soon as the data are available, so that corrective actions can be put in place before the end of the trial, if appropriate.
- 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.

Testing the QTL - Underreporting

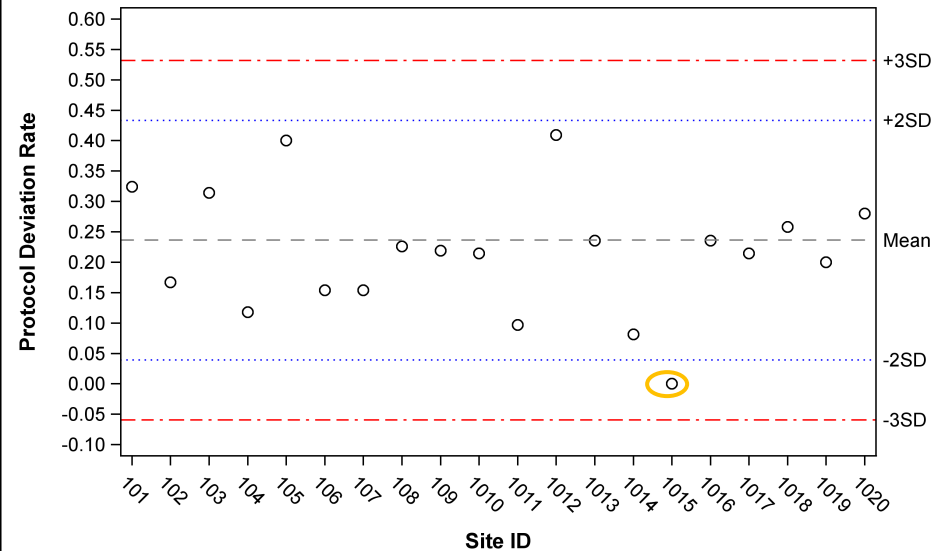
- It is equally important to be able to detect cases of potential underreporting, as well as excessively high rates of PDs.
- We use the data for our new trial, but instead of artificially inflating the results, we set the PD rate at one site (1015) to be extremely low (0.01) for all categories and at all visits, to model underreporting.
- Again we plot the data against our original QTLs. Clearly, as the lower limit is 0, rates cannot fall below the limit, so how do we pick this issue up?

Testing the QTL - Underreporting

Protocol deviation rate by site with calculated Quality Tolerance Limits overlaid
Visit 1

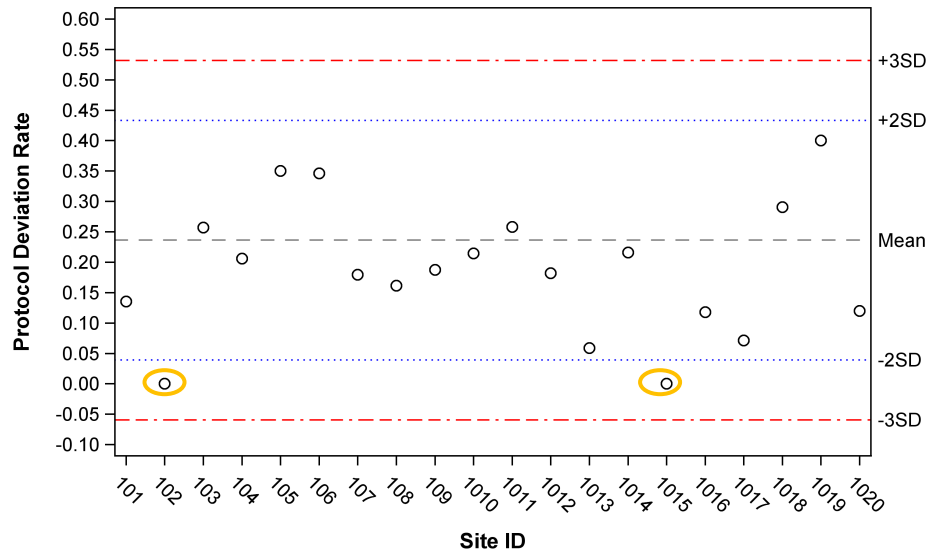


Protocol deviation rate by site with calculated Quality Tolerance Limits overlaid
Visit 2

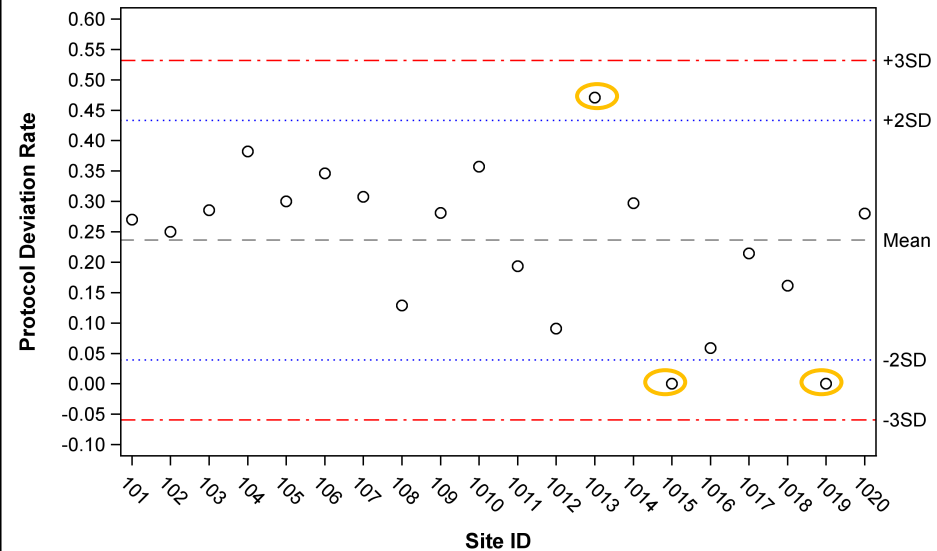


Testing the QTL - Underreporting

Protocol deviation rate by site with calculated Quality Tolerance Limits overlaid
Visit 3

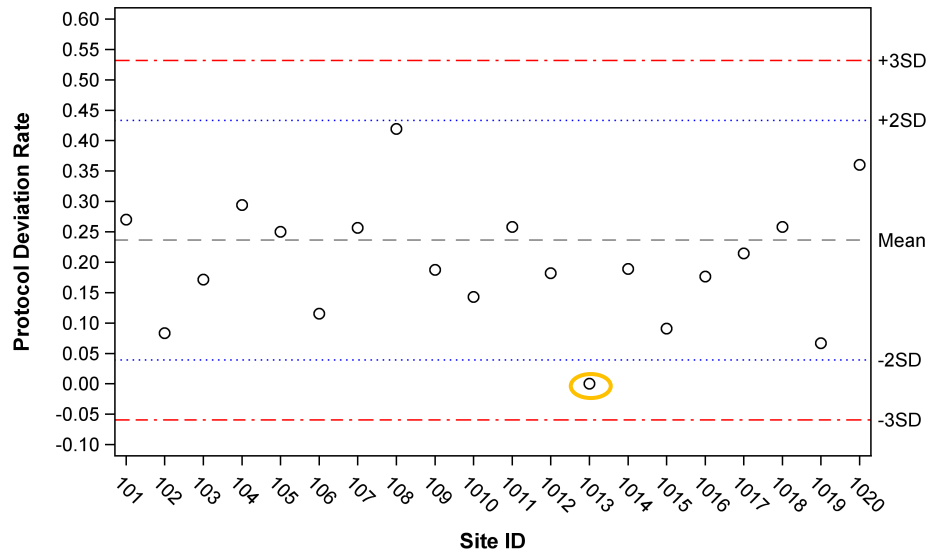


Protocol deviation rate by site with calculated Quality Tolerance Limits overlaid
Visit 4

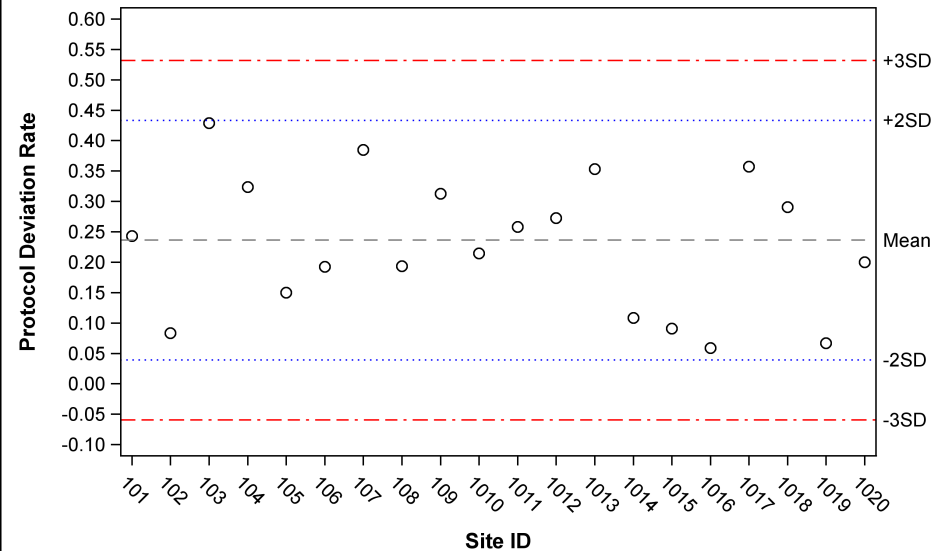


Testing the QTL - Underreporting

Protocol deviation rate by site with calculated Quality Tolerance Limits overlaid
Visit 5



Protocol deviation rate by site with calculated Quality Tolerance Limits overlaid
Visit 6



Testing the QTL - Underreporting

- There are no breaches of the red limits in this case, but there are several 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.

Testing the QTL - Underreporting

- 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.
- 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.

Conclusions

- QTLs are created using historical study data. If there is insufficient historical data to create a QTL, further data may be simulated in order to create the limit.
- 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.
- A balance must be struck between excluding data which is potentially unreliable and ensuring there is sufficient data to calculate the QTLs.

Conclusions

- This method of calculating a QTL for PDs is able to detect both unusually high rates and unusually low rates of PDs.
- 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.

