ISPE

ISPE Quality Metrics Initiative

Quality Metrics Pilot Program
Wave 2

June 2016

ISPE
7200 Wisconsin Ave., Suite 305
Bethesda, Maryland 20817 USA

Tel: +1-301-364-9201 – Fax: +1-240-204-6024 – ask@ispe.org

www.ISPE.org

Table of Contents

1 Executive Summary

Following issue of the "ISPE Quality Metrics Initiative: Wave 1 Report" [1] it was broadly agreed that there is a continuing appetite in the pharmaceutical manufacturing industry for information about quality metrics to support continual improvement. ISPE therefore initiated a Wave 2 Pilot, which commenced in July 2015. Initial goals were:

- Expand the data set across segments, geographies, and time to expand the knowledge gained from Wave 1 and evaluate trends.
- Continue to develop measures, tools, and dialogue related to quality culture and process capability to facilitate industry self-development and assessment.
- Enable continued objective and data-driven dialogue with FDA and other health authorities.

During the Wave 2 planning and setup phase, FDA issued a Federal Register Notice (FRN) [2] and "Request for Quality Metrics" Draft Guidance. [3] As a result, the design of Wave 2 was adjusted to include the following additional objectives:

- Test the proposed FDA metrics
  - Help develop appropriate definitions
  - Understand data collection challenges
- Evaluate the logistics and effort of gathering data at a product application level

WAVES 1 AND 2 HAD A COMBINED TOTAL OF **83 SITES** FROM **28 COMPANIES**

Wave 2, like Wave 1, was conducted in partnership with McKinsey and Company, who performed a confidential data collection and analysis. ISPE received only aggregated data; individual sites could not be identified.

The total number of sites increased from 44 sites and 18 companies in Wave 1 to 83 sites from 28 companies in Wave 1 and 2 combined. The total number of companies in Wave 2 was 21.

Three of the four metrics proposed by the FDA in Section V, Part B of the FDA Draft Guidance were evaluated:

- Lot Acceptance Rate
- Product Quality Complaint Rate
- Invalidated Out-of-Specification (OOS) rate

No Annual Product Review (APR) or Product Quality Review (PQR) on Time Rate was included in Wave 2, since findings in Wave 1 indicated that it was not differentiating.

Findings

ISPE Pilot Program Wave 2 met its objectives and confirmed findings from Wave 1. Main findings were:

- Effort to collect FDA Draft Guidance metrics was approximately three times that given in the FRN. This is probably an underestimate, especially for over-the-counter companies and companies with complex supply chains.
- ISPE estimates that using its recommended calculations for the three FDA Draft Guidance metrics evaluated requires one-third the effort of collecting data according to FDA calculations.
- When calculated using the FDA definitions, the three FDA Draft Guidance metrics evaluated did not exhibit relationships with either external quality outcomes or culture indicators.
  - Lot Acceptance Rate, using the FDA-recommended lots attempted as the denominator, did show relationships with some other internal quality outcomes, Invalidated OOS per test or lot, and Deviations Recurrence.
- When alternative calculations as defined in the Wave 2 Pilot were used, the same three FDA metrics did show relationships with culture indicators.
- Consistent with the ISPE comments to the FDA Draft Guidance, the Wave 2 Pilot study confirmed that alternative definitions of the three FDA Draft Guidance diagnostic metrics evaluated should be considered.
- Wave 2 confirmed the importance of quality culture, with some further relationships identified.
- Determining quality culture using simple-to-collect metrics—such as the three proposed FDA metrics—is confirmed as extremely difficult and may be not possible.
- Process capability/performance measures are extensively used by companies to help control processes and identify continual improvement opportunities. Use of these indices varies between companies.
- An internal metric, Deviations Recurrence rate was identified as one that could be used by companies to help predict external quality outcomes.

1

Preliminary findings from Wave 2 were used to develop ISPE's response to the FDA Draft Guidance and FRN. Final analysis confirmed those points, which are that ISPE:

- Supports FDA's effort to implement a quality metrics program
- Supports a small, targeted start to minimize the burden
  - This would also allow FDA and industry to learn, for example, about topics like implementation of standardized definitions, and the collection, submission, and analysis (e.g., statistical analysis) of data.
- Recommends a phased introduction as an option of "starting small"
- Supports starting with three of the proposed metrics while simultaneously considering varied definitions of these metrics
- Recommends deferring some metrics and data points
- Is concerned that the burden is underestimated
- Requests greater transparency in the manner in which data will be assessed, and outcome and conclusions determined and communicated

Participating companies reported that they derived great value from the metric data they received, and from participating in the confidential benchmarking exercise.

ISPE extends sincere gratitude to the 21 participating companies and their staff for the excellent input, support, and enthusiasm they provided throughout this Wave 2 Pilot.

2 Background

ISPE conducted a Quality Metrics Pilot Program Wave 1, which was reported in June 2015 [1]. The Wave 1 Pilot met its overall objectives and a summary of the insights gained include:

- It is feasible to collect and submit a standardized set of metrics.
- The majority of companies that participated reported the following benefits:
  - Gaining a deeper understanding of the standardized metrics definitions and design
  - Establishing a centralized submissions process trial
  - Developing access to a benchmarking report that allowed them to examine their progress against aggregated data from their peers
- Central collection and submission of metrics will create a burden for industry, primarily because standardized metrics will inevitably differ from current company metrics.
- Many companies will perform metrics collection for FDA reporting in addition to their established programs.
- Understanding context is crucial to interpreting results.
- The Wave 1 Pilot also provided some key insights in relation to the prevailing quality culture within an organization that merit further exploration.

Following presentation of the Wave 1 Pilot results at the ISPE Quality Metrics Summit in Baltimore on 21-22 April 2015, it was broadly agreed that there was a continuing appetite within industry for additional learning with respect to metrics to measure quality performance. ISPE therefore initiated a Wave 2 Pilot, which commenced in July/August 2015. The initial goals of this Wave 2 Pilot were:

- Expand the data set across segments, geographies, and time to expand the knowledge gained from Wave 1 and evaluate trends
- Continue to develop measures, tools, and dialogue related to quality culture and process capability to facilitate ongoing industry self-development and assessment
- Enable continued objective and data-driven dialogue with FDA and other health authorities.

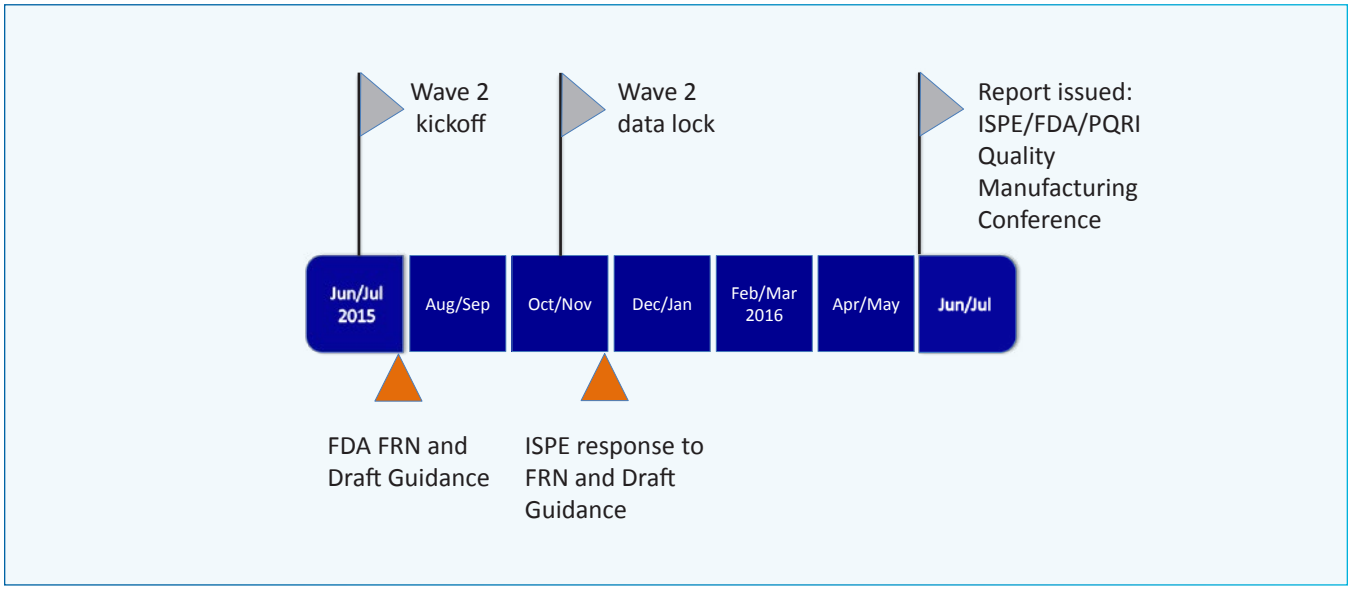
2

During Wave 2 planning and setup, FDA issued a Federal Register Notice (FRN) [2] and "Request for Quality Metrics" Draft Guidance. [3] As a result, the design of Wave 2 was adjusted to include the following additional objectives:

- Test the proposed FDA metrics
  - Help develop appropriate definitions
  - Understand data collection challenges
- Evaluate logistics and effort of gathering data at a product application level

Milestones for Wave 2 are given in Figure 1.

Figure 1: Wave 2 Pilot milestones



Given the convenient timing of the Wave 2 Pilot, preliminary data from Wave 2 analysis were included in ISPE's data-driven response to the FDA FRN and Draft Guidance. [4]

3 Wave 2 Pilot Design

3.1 Project governance and operating model

As described in the Wave 1 report, [1] Wave 2 was conducted in partnership with McKinsey and Company. Key points were:

- Only McKinsey personnel saw data from individual companies
- McKinsey conducted the data analysis
- ISPE Quality Metrics Project Team had access only to aggregated data across all companies or to subsets of companies where numbers were sufficient to maintain anonymity
- ISPE Project Team subgroups held regular teleconferences with participant company and site leads to:
  - Brief them on progress
  - Provide an overview of the data analysis for their review
  - Seek their input

The continuing importance of clear definitions of data to be collected and metrics to be calculated was again fully recognized with the Definitions sub team producing the definitions given in Appendix 1.

The Definitions sub team also worked with the Quality Culture sub team to develop Culture Indicators, which are also given in Appendix 1.

Quality Survey questions are the same as used in Wave 1 and are given in Appendix 2.

The Definitions subteam also helped shape survey questions relating to process capability/ performance values, which had initially been developed by the Process Capability subteam. These are given in Appendix 2.

Questions relating to potential drug shortages are given in Appendix 2.

McKinsey created separate templates for site- and product-based metrics and questions, and gave them to participants for completion.

3.2 Wave 2 Pilot metrics and survey questions

Metrics, culture indicators, and survey questions studied in Wave 2 were derived from the Wave 1 findings and those requested in the FRN/FDA Draft Guidance.

A summary of metrics, culture indicators, and surveys collected in Wave 2 is given in Figure 2.

Figure 2: Summary of Wave 2 Pilot metrics and survey questions

Quantitative Metrics	Culture Indicators	Surveys
• ***Product Quality (Total) Complaint Rate*** [1,4]	• CAPAs with Preventive Actions	• Quality Culture[2]
• Critical Complaints Rate [1]	• Planned Maintenance	• Process Capability[3]
• Recall Events	• Employee Turnover	• Drug Shortages[5]
• ***Lot Acceptance Rate*** [1,4]	• Human Error Deviations	
• ***Invalidated OOS Rate*** [1,4]	• Deviations with no Assigned Root Cause	
• Right-First-Time Rate	------	
• Deviations Rate[1]	• ***CAPAs Requiring Retraining[4]***	
• Recurring Deviations Rate		

[1] To be evaluated in multiple variants (e.g., lots attempted vs. dispositioned, packs released vs. lots released, lots tested vs. total tests performed).
[2] Only for sites that have not participated in Wave 1.
[3] Includes additional questions to the ones evaluated in Wave 1
[4] Aligned with the FRN guidance (also given in bold italics). Additionally, OOS Rate and Lots Pending Disposition for 30+ Days can also be calculated from the data.
[5] Questions to assess if standardized metrics can assist prediction of drug shortages.

3.2.1 Wave 2 metrics derived from Wave 1

Based on findings from the Wave 1 Pilot, ISPE recommended that the following set of starting metrics continue to be monitored in Wave 2:

- Lot Acceptance Rate (normalized by lots dispositioned), collected at site level
- Lot Acceptance Rate (normalized by lots dispositioned), collected at product level within a site
- Critical Complaints Rate (normalized by packs released), collected at product level by each product application, not broken down by site
- Critical Complaints Rate (normalized by packs released), collected at site level, undifferentiated by product
- Deviations Rate at site level

Definitions of metrics are those given in the Wave 1 Report.

The rationale for continued monitoring was:

- In most cases the metric had already been captured in the Wave 1 Pilot, and continued monitoring is desired over a longer time frame and broader set of companies, technologies, and regions.
- It demonstrated a statistically significant relationship in Wave 1 to one of the following:
  - Deviations Recurrence Rate
  - Quality culture values
  - Critical Complaints Rate
  - Lot Acceptance Rate
- It has proven relatively easy to collect and submit.
- It was deemed an important metric for determining site quality performance.
- It will help the company identify continual improvement opportunities.
- While the Critical Complaints Rate (normalized by number of packs released at product level by application, not broken down by site) was not included in the Wave 1 Pilot, it is thought to have merit and should be explored by product application.

3.2.2 Metrics from FDA FRN/Draft Guidance

After the FDA FRN/Draft Guidance was issued, three of the four FDA-proposed metrics with definitions given in the guidance were included in Wave 2. These are highlighted in bold in Figure 2 and are:

- Product Quality (Total) Complaint Rate
- Lot Acceptance Rate
- Invalidated OOS Rate

Annual Product Review (APR) or Product Quality Review (PQR) on Time Rate metric was not included in Wave 2 due to findings in Wave 1 that it was not differentiating.

FDA Draft Guidance states that data for FDA metrics are collected on a product application level and broken down by site (establishment). Templates were prepared to collect data on this basis. For Wave 2, sites were asked to complete data for two to five products supplied to the US market, ideally covering different technologies (e.g., solids/steriles) and supply chains (internal vs external, single vs. multi-site). Data were requested to cover a 12-month reporting period. For over-the-counter (OTC) products, which may not be subject to a product application, an alternative definition for a product application was provided.

Other Wave 2 metrics were collected on a site basis.

A more detailed list of Wave 2 metrics and quality culture indicator survey questions sub divided into External Quality Outcomes, Internal Quality Outcomes and Culture Indicators is given in Figure 3. Figure 3 shows the different denominators, which were used.

Figure 3: Detailed Wave 2 metrics and survey questions

Blue = FRN metrics[2] (tested) at site and at product level.

External Quality Outcomes

- Total Complaints Rate
  - Per million packs, incl. lack of effect
  - Per million packs, excl. lack of effect
  - Per ’000 attempted lots released, incl. lack of effect[2]
  - Per ’000 attempted lots released, excl. lack of effect[2]
- Critical Complaints Rate
  - Per million packs
  - Per ’000 attempted lots released
- Total Recall Events per year[1]

Internal Quality Outcomes

- Lot Acceptance Rate (%)
  - Per finally dispositioned lots
  - Per attempted lots[2]
- Invalidated OOS Rate
  - Per ’000 lots tested
  - Per ’000 tests performed
  - Per total OOS per tests performed[2]
- Right First Time Rate (%) per released lots attempted
- Deviations Rate
  - Per ’000 finally dispositioned lots
  - Per ’000 attempted lots
- Recurring Deviations Rate (%)
- Lots pending disposition more than 30 days (%) per lots attempted[2]

Culture Indicators

- Culture survey scores (% top boxes)
  - Total score
  - Leadership score
  - Integrity score
  - Mindset score
  - Governance score
  - Capabilities score
- CAPAs with Preventive Actions (%)
- Planned Maintenance Rate (%)
- Employee Turnover Rate (%)
- Human Error Deviations (%)
- Deviations with No Assigned Root Cause (%)
- CAPA Requiring Retraining (%)[2]

[1] Recalls are normalized on annual basis for sites that have submitted periods different from 12 months
[2] FRN metrics, tested at site and product level

3.2.3 Quality culture indicators

Wave 1 Pilot provided some key insights in relation to the prevailing quality culture within an organization that merited further exploration. The Quality Culture sub team, therefore, included in Wave 2 a series of additional Cultural Indicators to probe the relative importance of these indicators of quality culture. These Indicators were:

- Corrective Actions and Preventive Actions (CAPAs) with Preventive Actions (%)
- Planned Maintenance Rate (%)
- Employee Turnover Rate (%)
- Human Error Deviations (%)
- Deviations with No Assigned Root Cause (%)

These were collected on a site basis.

Sites that had not completed a quality culture survey questionnaire in Wave 1 were asked to do so in Wave 2. Wave 1 participants were not required to complete a second quality culture survey.

In addition, optional metrics related to quality culture given in the FDA Draft Guidance were included:

- Proposed Optional Metric 1 related to APR or PQR review and approval
- Proposed Optional Metric 2 related to CAPA effectiveness and the proportion of CAPAs involved retraining

APR/PQR approval metric was collected on a product basis, and CAPAs that required retraining were collected on a site basis.

3.2.4 Process capability

Wave 1 assessed the tools and processes used to monitor process capability. Findings indicated that the capability approach varies by company in terms of use and applicability. The tool employed (e.g., process performance index [Ppk], process capability index [Cpk], and control charts) is contextual and there is no one tool that can be applied to all situations.

Wave 2 questions were designed to explore more deeply than Wave 1 questions such as how process capability values were developed and used in industry. In addition, Wave 2 included questions relating to process capability/performance given in the FDA Draft Guidance as Proposed Optional Metric 3. [3] These questions were asked on a site basis.

3.2.5 Drug shortages

An objective of the FDA Draft Guidance [3] is to "better detect manufacturing conditions that may lead to a drug shortage," and stated "FDA intends to use these quality metrics, in part, as a tool to identify risk-based factors that could increase or decrease inspection frequency and that could potentially be predictive of drug supply disruption." [3]

Wave 2 included questions that asked if a company/site had experienced a drug shortage, and if there were some metrics in the Pilot that were predictive or could mitigate a potential drug shortage. Wave 2 also asked a company's view on whether quality metrics help predict potential drug shortages.

These questions were asked on a product basis in relation to all products in a site or US portfolio, not just the two to five products for which data were requested on a product-application level. The questions did not differentiate by shortage type, duration, or significance.

3.3 Data collection and submission effort estimates

In addition to data collection on a site basis, participating companies were asked to record time spent gathering and reviewing product data at application level (on two to five products) for the FDA-proposed metrics of Product Quality Complaint Rate, Lot Acceptance Rate, and Invalidated OOS Rate.

3.4 Data collection period

New participants (who had not participated in Wave 1) were requested to provide 12 or 24 months of retrospective data. Participants who had submitted data in Wave 1 were requested to provide 9 or 12 months of retrospective data. Participants were allowed to provide data that was most convenient for them.

3.5 Systems

Sites were asked to provide information on which information technology (IT) systems they used to collect and submit data.

4 Data Analysis

Based on the objectives of the Wave 2 Pilot given in Section 2 above, there were objectives for analysis of the data, these being:

- Understand the metrics' variability and consistency
- Assess the effort required (in time and cost) to collect and report the data requested in the FDA Draft Guidance data
- Test variations of metrics, to determine the degree of difference between the variations and what may be optimal definitions or calculation methods for each metric
- Identify significant connections/relationships between metrics (one to one or in combination), e.g. site practices that influence outcomes, internal quality metrics that influence external quality outcomes.

Effort/burden analysis was performed on three of the four metrics requested in the FDA Draft Guidance – Lot Acceptance Rate, Product Quality Complaint Rate and Invalidated OOS Rate. Annual Product Review (APR) or Product Quality Review (PQR) on Time Rate metric was not included in Wave 2 due to findings in Wave 1 indicating that it was not differentiating.

Relationships between metrics were identified based on:

- Using relatively simple linear regression analysis as a practical approach, consistent with that used in with Wave 1.
- The sample size in Wave 2 is such that relationships identified using linear regression analysis could be compared with relationships identified from Wave 1 data, which were derived using the same analysis.
- Only statistically significant relationships are highlighted i.e. less than 5% likelihood of a coincidence. *p* value is probability that correlation between X and Y is zero. A *p* value below 0.05 indicates a significant result.
- Potential relationships were investigated for the total Wave 2 sample for a particular comparison, and also for sub sets of similar technologies (e.g. solids or sterile products) where the sample size was sufficiently large to support blinding of individual sites.
- Multivariate relationships (what combination of metrics significantly relates with the leading indicators or outcomes to take into account the inter-relationships between these metrics) using multivariate analysis
- Time lags (whether one metric influences another after a time lag of several months and consistently for multiple sites)

It must be stressed that a statistically significant relationship does **NOT** imply causation. To derive greater levels of understanding of the many factors involved in and direction of a potential relationship requires much further work. The Findings section below includes a discussion of possible reasons for a relationship and gives some background as potential hypotheses to be evaluated in any further work.

The data generally are 'noisy' with relatively high variation and hence a relatively simple statistical analysis approach was utilized for analysis. The strength of the relationships reflected by a correlation coefficient (R^2) varies and may be relatively low. For example, some relationships may have values of R^2 of 30% or 40%, which is expected since these metrics are influenced by multiple factors.

Incomplete data sets and outliers were treated as follows:

- Incomplete data sets were excluded from the analysis
- Outliers were excluded using a consistent rule for all samples, excluding any data point more than 2 standard deviations from average, or two interquartile ranges from median for non-normal distributions. Due to limited knowledge of the specific context and variables in participating sites, this consistent rule was applied to avoid judgment on what comprised an outlier result. For example if a data point was located far from the main cluster, there was insufficient information available (even after consulting with the site) to conclude its location was due to unique and unlikely-to-repeat circumstances, a different method of tracking used at the site despite common definition, or a legitimate result even if different from the other sites. Otherwise put, the information available was insufficient to determine whether a data point was a true outlier or represented a subset of the larger industry population. Hence the consistent exclusion rule was selected as the most objective and independent approach.

Linear regression was used appropriately to find the best directionally correct relationship. All data were continuous or at least ordinal (the latter in the case of annualized recalls). More sophisticated statistical analysis such as data transformation through logistics function or development of confidence intervals was not deemed necessary.

While some data points had hard boundaries, R^2 has been used to describe the fitness of the linear trend for 2 variables from the sample within the range and R^2 is independent to the scale of each variable. There is no attempt to use the observed linear trends for prediction or extrapolation of values, but rather to illustrate the observed relationship between the variables.

Presence of a relationship is indicative of a "direction," which may be worthy of further exploration, either as a group of sites, or by individual sites.

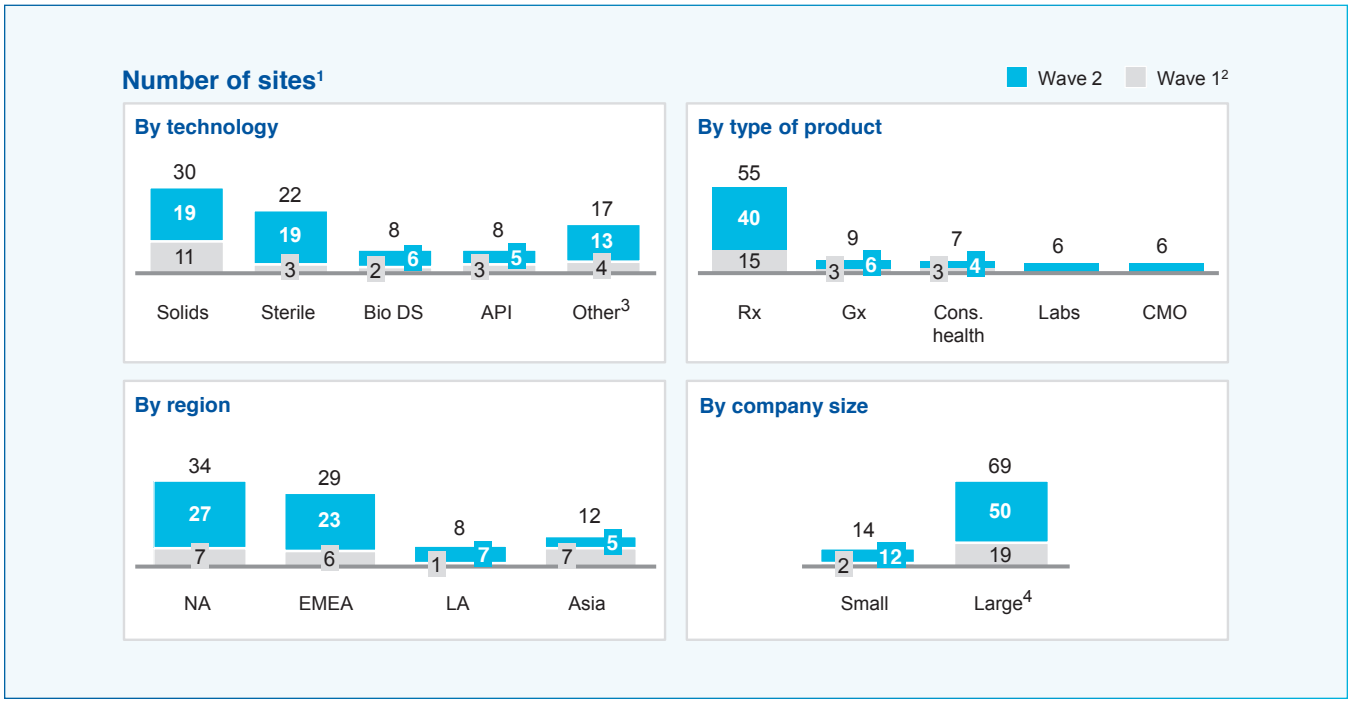
As a first "directional question," a linkage between metrics and ‘compliance status’ of a site was investigated. McKinsey obtained publically available information relating to compliance status of sites (Consent decrees, class 1 recalls, warning letters, number of 483s) and attempts were made post Wave 2 data lock to relate compliance status with metric data.

5 Findings from ISPE Pilot Program Wave 2

5.1 Sample size

Final enrollment from Wave 1 and Wave 2 is given in Figure 4, showing the total number of sites broken down by technology, type of product, region, and company size.

Figure 4: Combined enrollment for Wave 1 and Wave 2



[1] If a site has more than one technology we count the number of separate templates they will fill, usually one per technology
[2] Sites that participated in both Wave 1 and Wave 2 are reported under Wave 2 only
[3] e.g., soft gels, transdermal
[4] Over $1 billion in annual revenue

Abbreviations
Bio DS: Biopharmaceutical or biological drug substance site
API: Active pharmaceutical ingredient/small molecule drug substance
Rx: Originator company
Gx: Generic company
Cons. Health: Consumer health or OTC company
CMO: Contract manufacturing organization laboratory
Labs: Contract research and testing laboratories
NA: North America
EMEA: Europe, Middle East, and Africa
LA: Latin America
Small company: < $1 billion in revenues
Large company: > $1 billion in revenues

Sites that were in both Wave 1 and Wave 2 are reported under Wave 2 only; those reported as Wave 1 participated in Wave 1 only.

Total number of sites increased from 44 (18 companies) in Wave 1 to 83 (28 companies) in Waves 1 and 2 combined. The total number of companies in Wave 2 was 21.

Observations relating to the Wave 2 sample:

- Sample sizes increased across all technologies, giving good representation in all technologies.
- The sample is dominated by originator (Rx) companies and sites.
- The number of contract manufacturing organizations (CMOs) and laboratories participating in both Wave 1 and 2 remained the same.
- The sample is dominated by companies with revenues greater than $1 billion, but the proportion of smaller companies increased from about 10% to about 17%.

5.2 Effort analysis

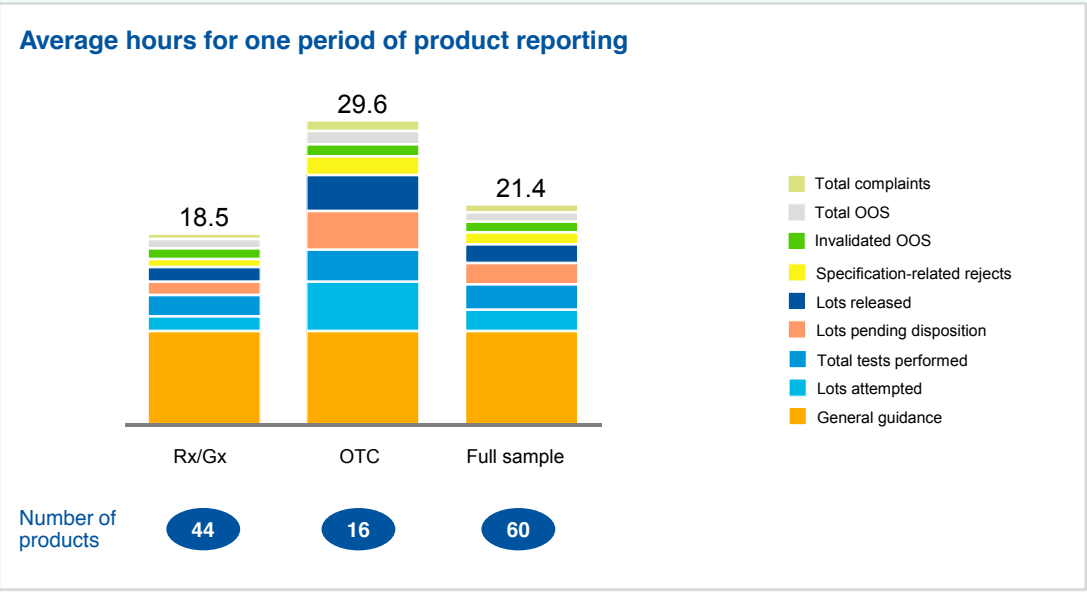
Wave 2 used several data analysis methods to estimate the effort required to collect metric data as described in the FDA Draft Guidance.

5.2.1 Overall industry effort

In Figure 5, collection effort for Rx and generic (Gx) sites is compared to OTC/consumer health sites; an average is also given for the full sample.

5

Figure 5: Data collection effort



The 21.4 hours of data reporting for the full sample consists of 12.3 hours for pure data collection and 9.1 hours for guidance and coordination. In this context, guidance and coordination are defined as: clarifying definitions of data to be collected, answering data-related questions from sites, and reviewing data.

Extrapolating these results to quarterly data collection (four data points per year) was based on three assumptions drawn from McKinsey's POBOS (pharma operations benchmarking) experience with repeat data collection at the same site over many years (see Wave 1 Report, Section 2 for further explanation of McKinsey POBOS programs):

- For Lot Acceptance Rate, quarterly data increases effort ~3 times
- For other metrics, quarterly collection will add only ~20% effort
- Guidance time and coordination is assumed to be unchanged

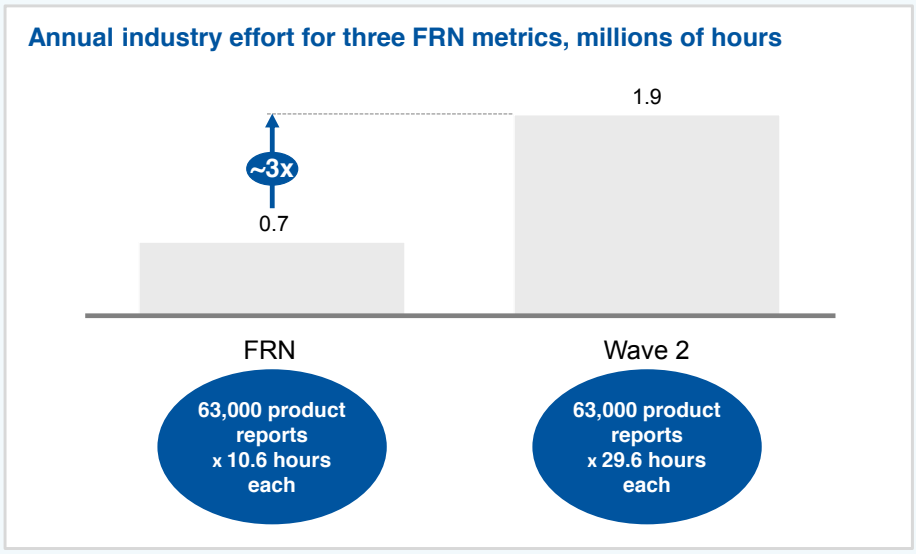
Based on these assumptions **the extrapolated annual effort would be a total of 29.6 hours** (20.5 hours average for pure data collection and 9.1 hours for guidance and coordination).

Figure 5 shows that data collection effort for consumer health/OTC sites is 60% higher than Rx/Gx sites, which were found to be similar.

5.2.2 Industry effort to collect FDA Draft Guidance metrics

An estimate of the amount of effort required to collect the three FDA Draft Guidance metrics is given in Figure 6.

Figure 6: Estimate of effort to collect FDA Draft Guidance metrics

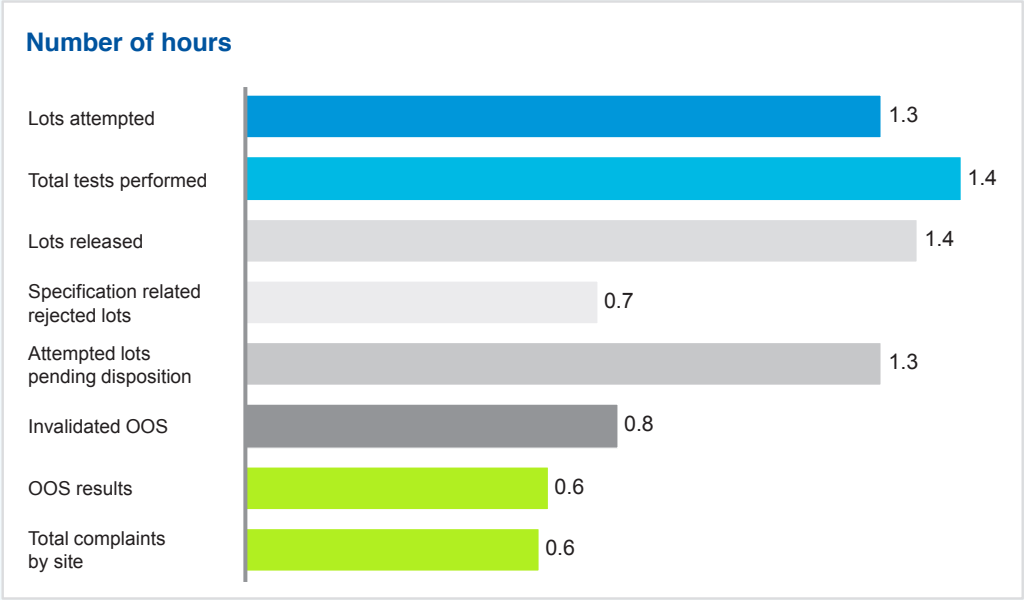


To compare findings from Wave 2 with estimates given in the FRN, the estimated annual effort of 29.6 hours was multiplied by 63,000 total annual responses given in the FRN Table 1. [2] This equals 1.9 million hours, compared with 667,800 hours given in the FRN, approximately **3 times greater**.

As observations, if lots pending disposition and total OOS data points are omitted, industry collection effort would decrease, as these represent about 13% of the total effort. Lots pending disposition is a measurement tool and would be used to verify data supporting Lot Acceptance Rate values. Total OOS result data points are used in the Invalidated OOS Rate calculation; in ISPE's view, this is not a helpful denominator.

The analysis considered if and how this value might change with experience, for example, with systems designed to collect and report FDA-requested metrics. Collection-effort values for each FDA-requested data point at a site is given in Figure 7.

Figure 7: Data collection effort for each data point by site, per product



These data indicate it would be extremely difficult to predict if and how effort might change when FDA Draft Guidance is implemented. Most likely a range of approaches would be adopted across the industry, and it is unlikely that a single solution would apply for all companies. Despite automation (see Section 5.7) manual effort may be required to record and check values before entering them into a metrics IT system, or to check automatically collected values before they are entered into a formal regulatory system.

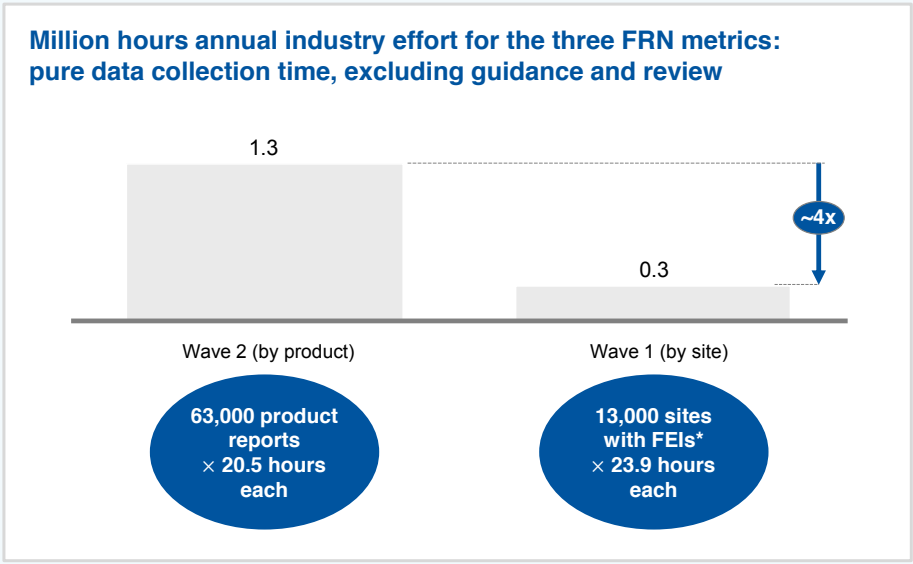
FDA Draft Guidance states that metrics be aggregated to the product level. This means sites will have to be conscientious when inputting data into cross-site and potentially cross-company IT systems. Consequently, effort reduction after implementation may not be significant. Regardless, it is not anticipated that the guidance and coordination effort estimate will change.

IT system development experience suggests that considerable effort and time will be required to develop and validate a system (or systems) applicable across a supply chain and suitable for that site or company. Such a system would have to be flexible to allow changes to supply chains. Payback may take many years.

5.2.3 Collecting FDA-proposed metrics on a site basis

An analysis was performed to determine what the impact would be of collecting FDA metric data in the more site-based manner and using definitions recommended by ISPE, i.e. largely on a site-basis with different definitions. Lot Acceptance Rate and Invalidated OOS Rate would be collected on a site-basis with Product Quality Complaint Rate being collected centrally. This was the design of Wave 1 and hence the values of effort can be abstracted and an estimated total industry value calculated as in Figure 8.

Figure 8: Comparison of data collection effort by product and by site



Note: the comparison excludes the effort for guidance, coordination, and review, since these were not evaluated during Wave 1.
The Wave 1 estimate is for site-level only, excluding any additional effort per product within the site.
* Federal Establishment Identifier numbers

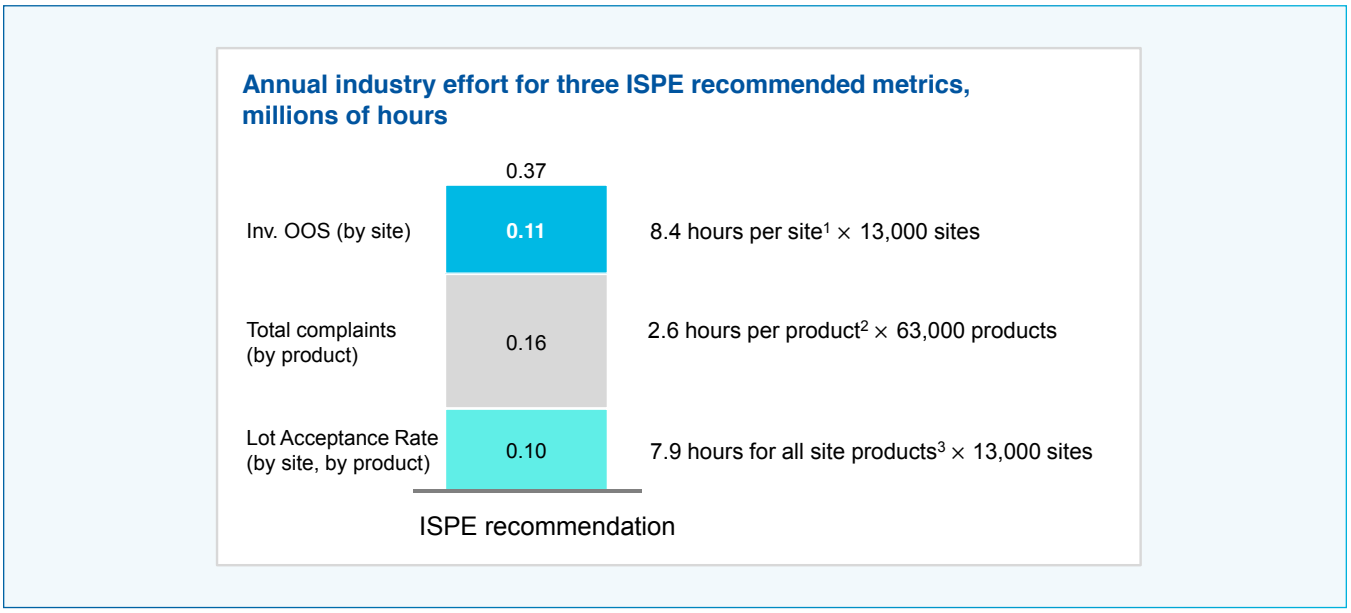
On the left, Wave 2 annual data collection effort is estimated for 63,000 product reports [2] using an average value of 20.5 hours for data collection (without guidance and coordination). On the right, Wave 1 annual data collection effort on a site-only basis is estimated for the 12,949 registered drug establishments identified by Federal Establishment Indicator (FEI) numbers in 2014. [5] Using the same three FDA Draft Guidance metrics, an average of 23.9 hours are required for each site (omitting guidance and coordination effort, since these were not collected in Wave 1). Effort to collect these three metrics on a product basis is four times greater than collecting them on a site basis.

Since the data in the right column in Figure 8 are based on site-only collection, an attempt was made to estimate the difference in effort between collecting data on a site-only basis and ISPE's recommendation of collecting data for:

- Lot Acceptance Rate on a site-by-product basis initially, potentially moving to a product-differentiated-by-site basis
- Invalidated OOS Rate on a site-only basis
- Product Quality Complaint Rate on a product basis

The calculations are presented in Figure 9, with assumptions listed in the footnotes.

Figure 9: Data collection effort using the ISPE-recommended approach



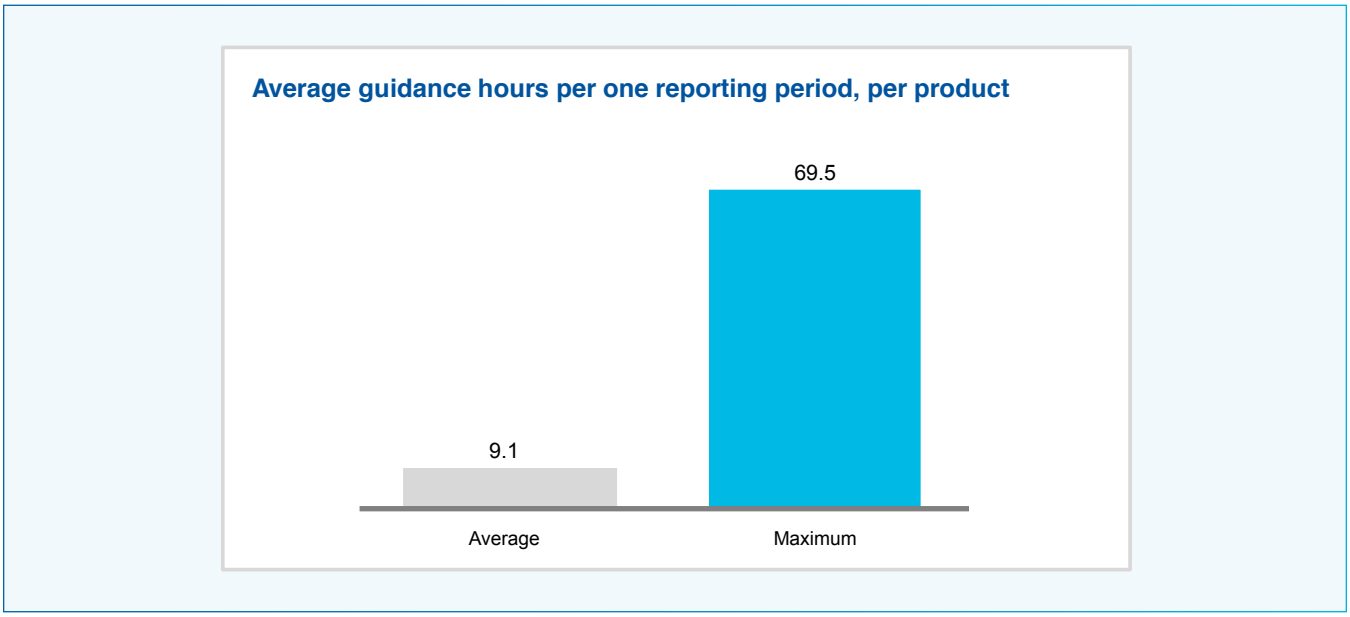
[1] Based on Wave 1 data for "unconfirmed OOS" as reported by sites for site-only effort
[2] Based on Wave 2 data as reported by companies
[3] Based on Wave 1 data for "Lot Acceptance Rate normalized by lots dispositioned" as reported by sites for site- and product-level effort, for all products on-site. Site-only effort only for Lot Acceptance Rate was 6.5 hours.
Note: Comparison excludes the effort for guidance, coordination, and review, since these were not evaluated in Wave 1.

In Figure 9, annual industry collection effort is estimated on the basis recommended by ISPE using values for each metric and the appropriate site or product multiplier. The vertical column shows the summation of 0.37 million hours annually for the ISPE-recommended approach compared with 0.3 million hours using a site-only basis (Figure 8). This increase is small and does not affect the overall conclusion that the effort to collect data on IPSE's recommended basis is significantly—by about a third—lower than ISPE's estimate of the approach requested in the FDA Draft Guidance.

5.2.4 Guidance and coordination effort

The amount of effort given by companies to provide guidance and coordination to sites and individuals who collect and submit data showed wide variation. Figure 10 shows that although the average value was 9.1 hours per period of product reporting, one company reported a value of 69.5 hours.

Figure 10: Guidance effort



Other observations related to general guidance were:

- Guidance and coordination effort was highest for companies:
  - Outside of US/Europe
  - With complex supply chains (10 or more sites)
- Some companies decided not to submit product data for the Wave 2 pilot, citing the degree of effort required: “It will take much more effort and coordination to provide data at the product application level, and if we need to set up the internal organization/reporting structure to provide data in this manner we will do so, but not at this time for the means of the pilot.”
- Ease of obtaining data from partners (e.g. CMOs) outside the company was also variable.

5.2.5 Other observations

Data and feedback from participants also indicates that actual industry effort will likely be higher than that reported in the Wave 2 sample because:

- Products selected by companies for inclusion in Wave 2 had relatively simple supply chains—over 60% of Wave 2 products were manufactured at a single site. FDA estimates [6] that there will be 5 to 10 sites per product.
- The Wave 2 Pilot involved collection for 8 data points; FDA Draft Guidance states that 10 should be collected.
- The Wave 2 sample is relatively small (60 products from 14 companies) compared to the industry as a whole.
- Some companies will experience a high effort burden:
  - Companies with complex supply chains
  - OTC sites, which needed 60% more time to collect data than Rx/Gx sites
  - Companies using CMOs, given the need for coordination effort and confidentiality arrangements to support reporting data from CMOs

5.2.6 Conclusions

1. The Wave 2 overall effort estimate for the three FDA Draft Guidance metrics requested and evaluated (Lot Acceptance Rate, Product Quality Complaint Rate, and Invalidated OOS Rate) exceeds the FRN estimate by approximately three times.
2. Collecting data for the three FDA metrics evaluated by product requires approximately four times more effort than collecting data on a site-only basis.
3. Collecting data in the manner recommended by ISPE requires only slightly more effort (0.3 vs. 0.4 million hours annually) than collecting data on a site-only basis, and is approximately 1/3 the ISPE-estimated effort to collect data as required in the FDA Draft Guidance.
4. Actual industry effort for reporting quality metrics to FDA is likely to be even higher than estimated.
5. Some companies and sites will experience higher burdens than others.

5.3 Analysis of metrics data and relationships

5.3.1 General observations of quantitative metric data

Median values of all metrics collected by technology are given in Figure 11.

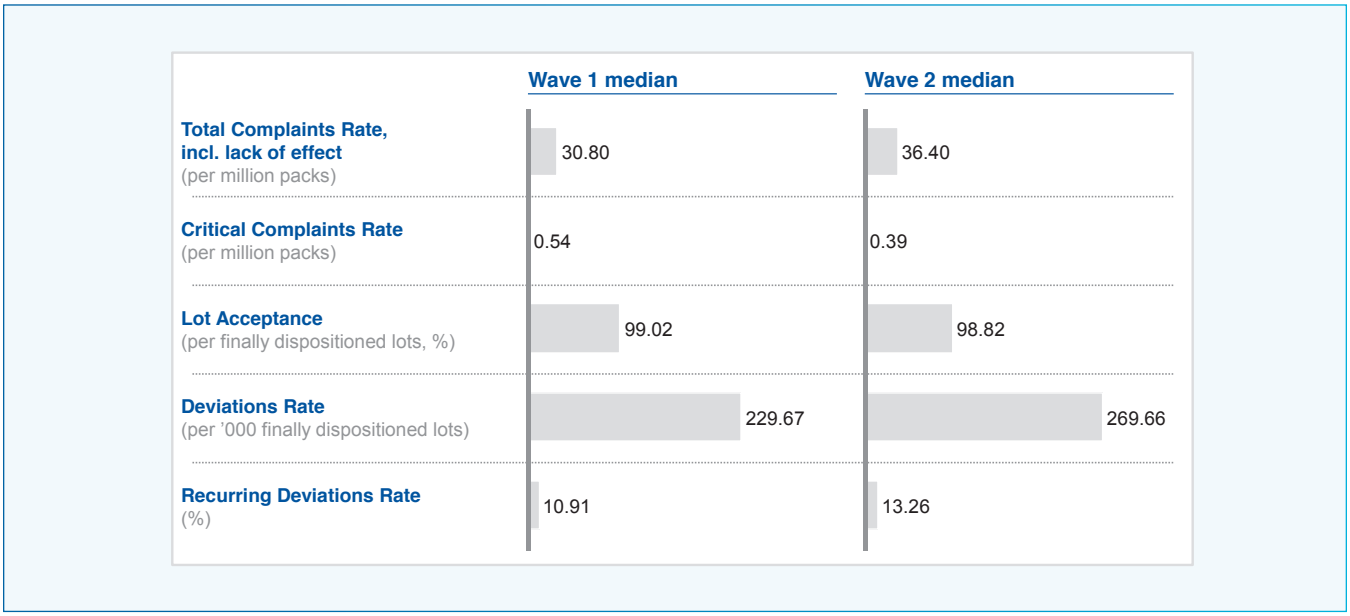
Figure 11: Median values of metrics by technology

		Solids	Steriles	Liquids/Creams	API	Bio DS
External Quality Outcomes	**Total Complaints Rate, incl. lack of effect** (per million packs)	18	41	49		
	Total Complaints Rate, incl. lack of effect (per '000 attempted lots released)	131	369	1,861	0	0
	Total Complaints Rate, ex. lack of effect (per million packs)	14	45	46		
	Total Complaints Rate, ex. lack of effect (per '000 attempted lots released)	116	311	1,752	0	0
	Critical Complaints Rate (per million packs)	0	0	1		
	Critical Complaints Rate (per '000 attempted lots released)	3	11	27	0	0
	Total Recall Events (Recalls per year)	0	0	0	0	0
Internal Quality Outcomes	**Lot Acceptance** (per finally dispositioned lots, percent)	99	99	99	100	99
	Lot Acceptance (per attempted lots, percent)	100	100	100	100	99
	Lots pending disposition more than 30 days (per attempted lots, percent)	6	7	1	67	181
	Right First Time Rate (%)	89	76	98	79	0
	Invalidated OOS Rate (per '000 lots tested)	4	6	2	11	21
	Invalidated OOS Rate (per '000 test performed)	1	0	1	1	2
	Deviations Rate (per '000 attempted lots)	123	186	15	386	8,326
	Deviations Rate (per '000 finally dispositioned lots)	143	283	77	616	7,043
	Recurring Deviations Rate (%)	11	9	19	6	11
Culture indicators	**Deviations with No Assigned Root Cause** (%)	5	4	0	7	6
	Human Error Deviations Rate (%)	33	20	36	27	32
	CAPAs with Preventive Actions Rate (%)	26	20	31	53	35
	Planned Maintenance Rate (%)	59	74	82	35	70
	Employee Turnover Rate (%)	7	6	13	4	7
	CAPAs Requiring Retraining Rate (%)	9	8	8	9	9

Based on full sample of Wave 2 data and Wave 1 data for sites that did not participate in Wave 2, no outliers excluded.

A comparison of Wave 1 and Wave 2 data for site-only metrics common to both is given in Figure 12.

Figure 12: Median values of Wave 1 and Wave 2 metrics



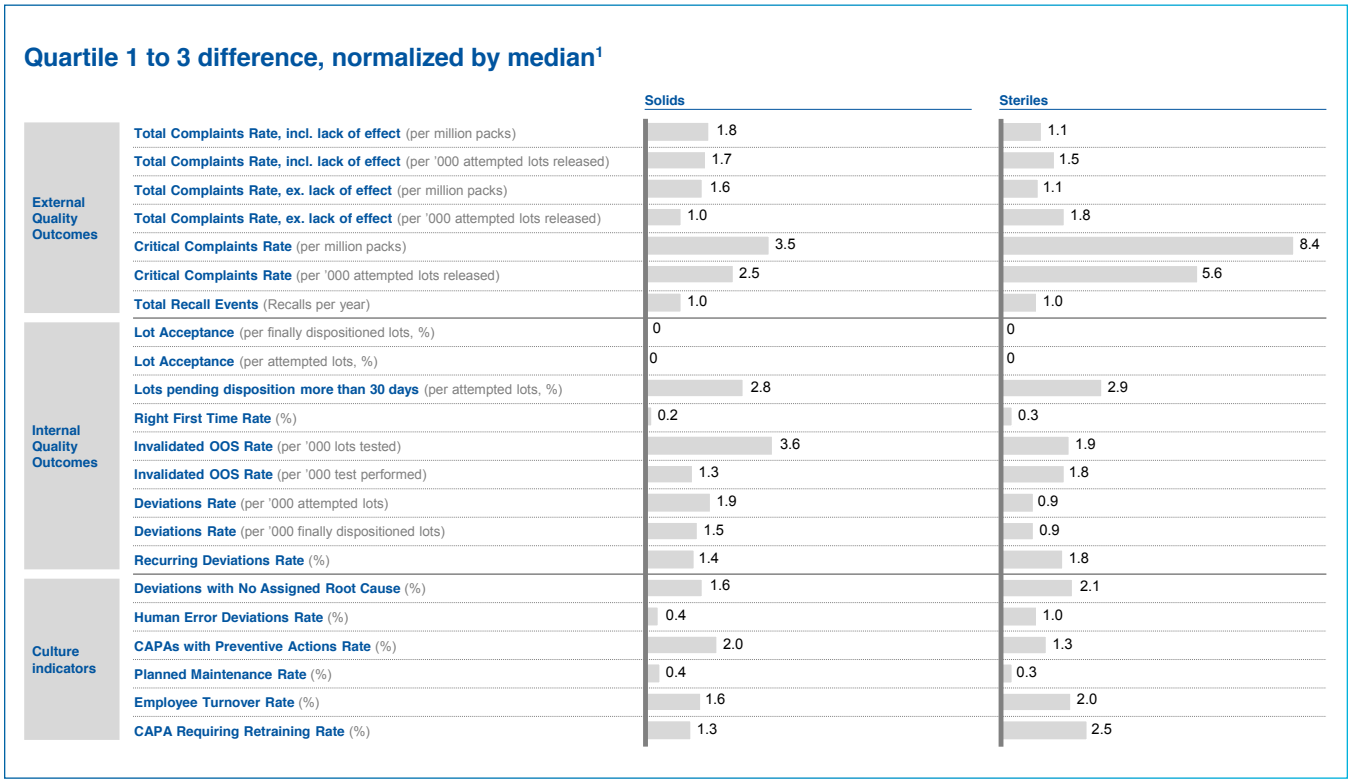
Based on full sample of Wave 2 data and Wave 1 data for sites that did not participate in Wave 2, no outliers excluded.

This comparison shows that Wave 1 and Wave 2 median values are similar, with consistent data in both.

5

Figure 13 shows the variability of metrics data.

Figure 13: Metrics data variability, measured by difference in quartile values

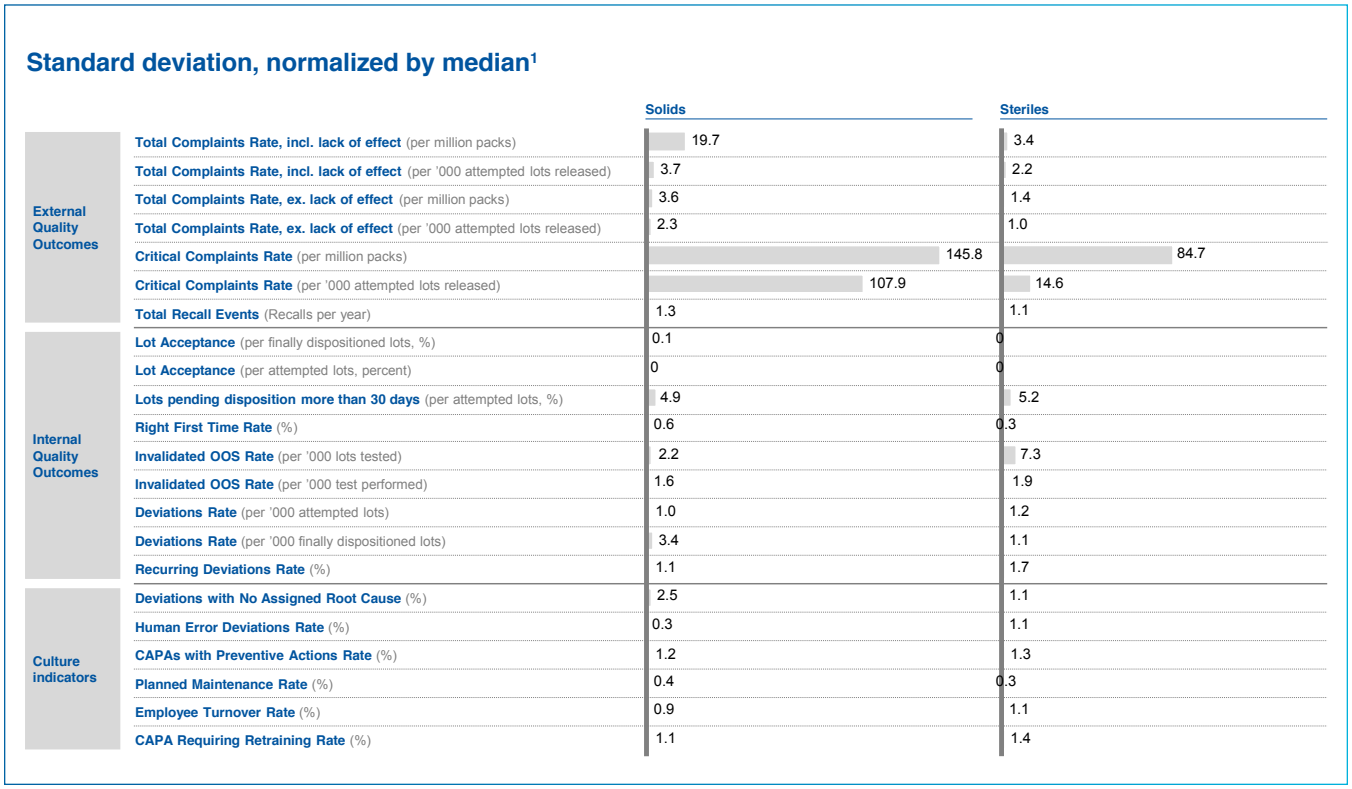


Based on full sample of Wave 2 data and Wave 1 data for sites that did not participate in Wave 2, no outliers excluded.
[1] For recalls, interquartile ranges only

The Critical Complaints Rate shows especially high variability in sterile products, with low variability in Lot Acceptance Rate and Right First Time (RFT) rate. This is demonstrated by interquartile ranges (difference between quartile 1 and 3, normalized by the median).

Comparing standard deviations as shown in Figure 14 also shows high variability in the Critical Complaints Rate.

Figure 14: Metrics data variability, measured by standard deviation



Based on full sample of Wave 2 data and Wave 1 data for sites that did not participate in Wave 2, no outliers excluded.

[1] For recalls, standard deviation only

5.3.2 General observations from quality culture surveys

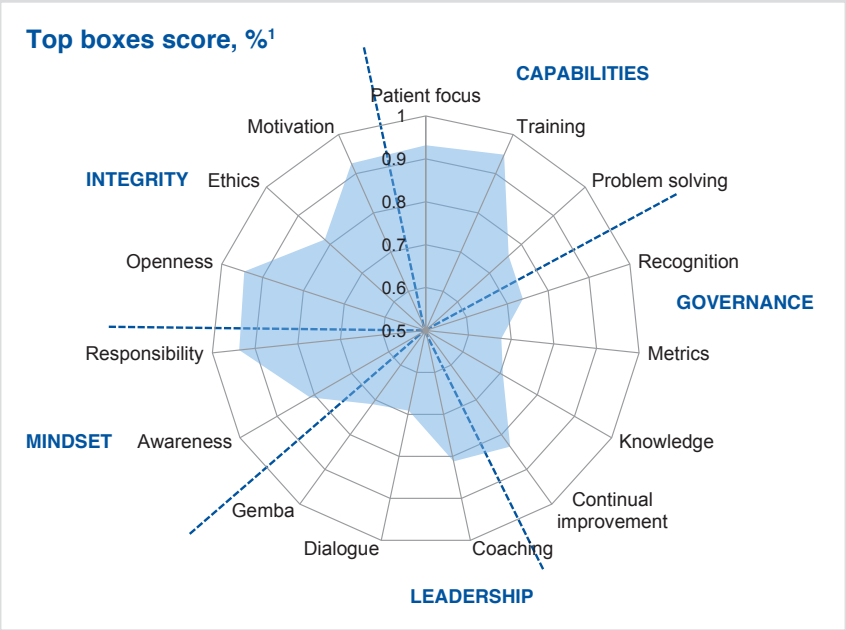
For the quality culture survey, each of the 15 questions could be answered using one of five response options:

- Strongly agree
- Agree
- Disagree
- Strongly disagree
- I can't answer this question

To facilitate data analysis and relationship mapping, the scoring mechanism was based on the "top boxes" approach. For each question, the proportion of "Strongly agree" and "Agree" answers was calculated. Top boxes analysis assigns a 1 (or 100%) if all respondents reply "Strongly agree" or "Agree," and 0 (0%) if all respondents reply "Disagree" or "Strongly disagree."

Figure 15 shows these values plotted in a radar diagram.

Figure 15: Quality culture scores overview



Based on surveys conducted in 82 plants

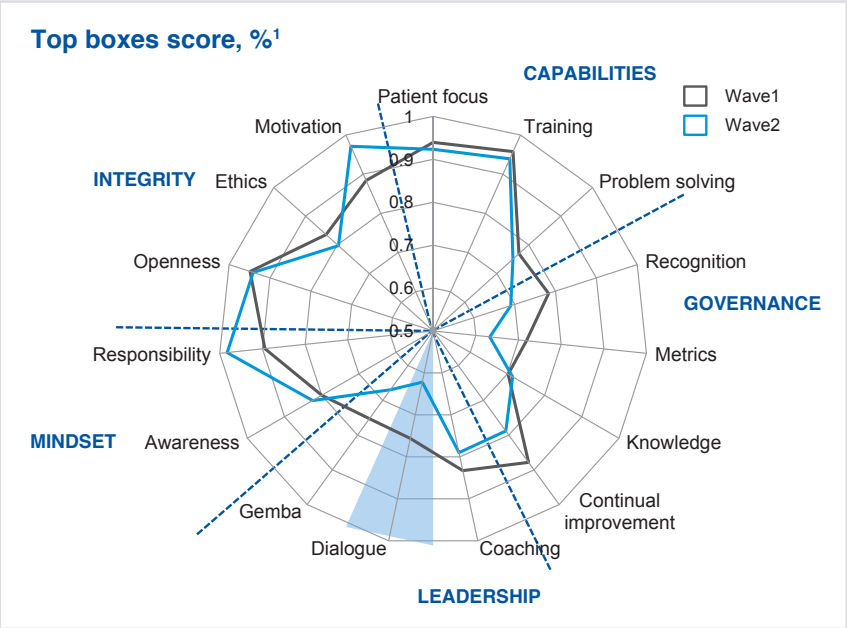
[1] Total score calculated as "top boxes" (share of "agree" and "strongly agree" responses) ratio. 100% = all respondents agree or strongly agree, 0% = nobody agrees or strongly agrees.

High scores in capabilities and integrity show industry-wide strengths in training, patient focus, personal responsibility for quality, open escalation of quality issues, and motivation to ensure quality. Lower scores are observed in governance and leadership, showing industry-wide gaps in metrics visualization and understanding, management presence on shop floor (Gemba), and daily dialogue.

Lower scores are observed in Governance and Leadership showing industry-wide gaps in metrics visualization and understanding, management presence on shop floor (Gemba) and daily dialogue.

Wave 1 and Wave 2 quality culture scores are compared in Figure 16.

Figure 16: Wave 1 and Wave 2 quality culture scores



Based on surveys conducted in 82 plants.

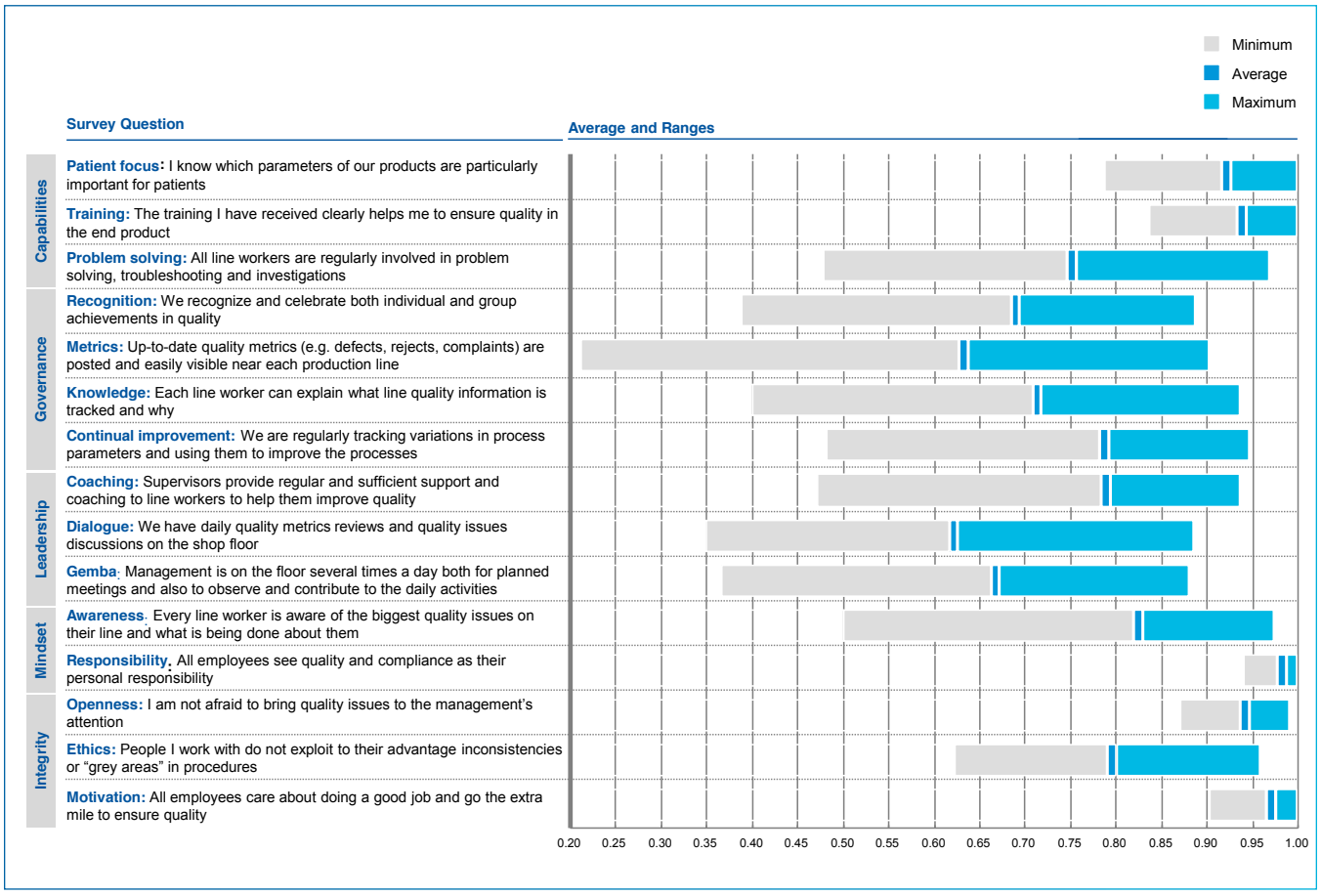
[1] Total score calculated as "top boxes" (share of "Agree" and "Strongly agree" responses) ratio. 100% = all respondents agree or strongly agree, 0% = nobody agrees or strongly agrees.

5

Wave 1 and Wave 2 participants had broadly similar quality culture scores.

Average values and ranges for Wave 2 quality culture scores (36 sites) are given in Figure 17.

Figure 17: Wave 2 quality culture scores



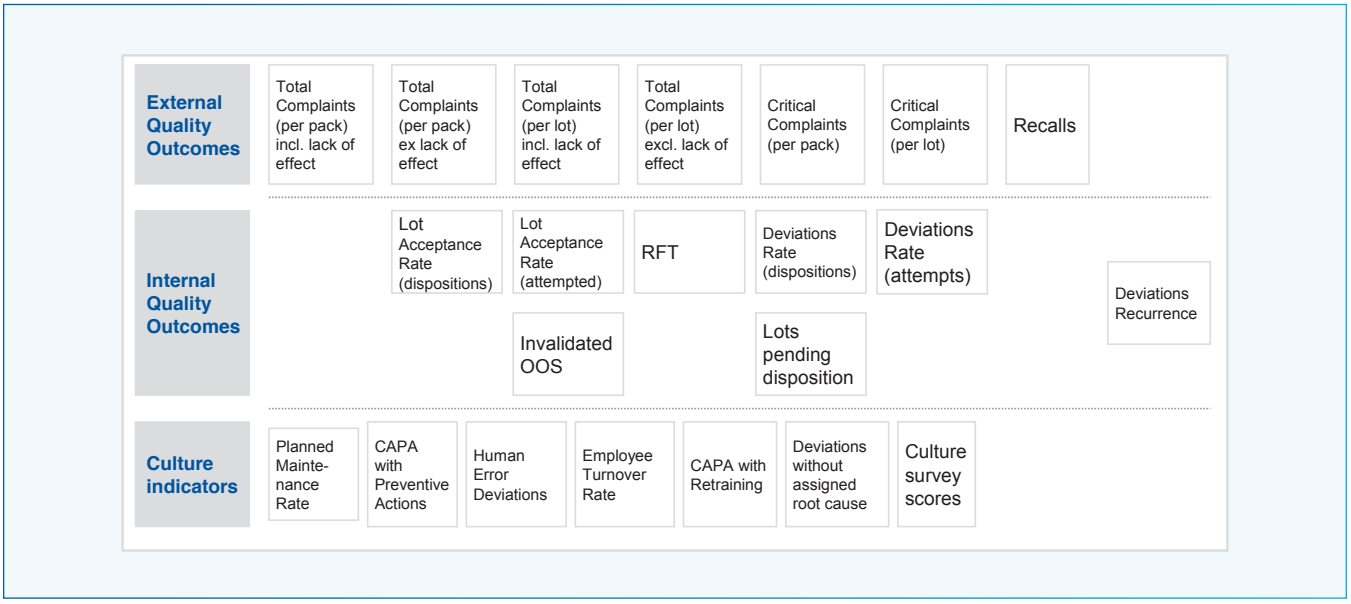
Based on surveys conducted in 36 plants (Wave 2 participants only).

Highest variability in quality culture scores in Wave 2 was observed for metrics, dialogue, and Gemba.

5.3.3 Overview of relationships

As with Wave 1, metrics and culture indicators evaluated in Wave 2 were assigned to groupings as external quality outcomes, internal quality outcomes, or culture indicators, as given in Figure 18.

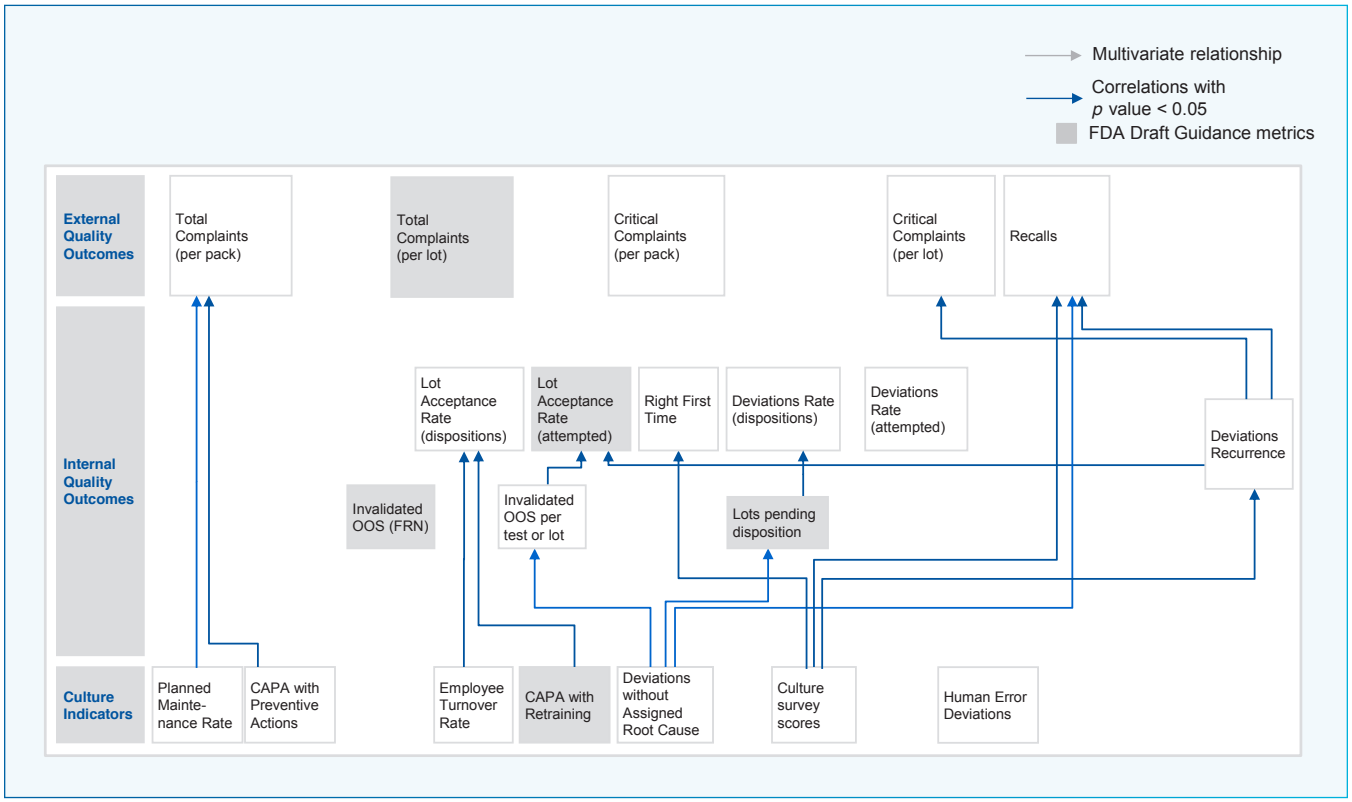
Figure 18: Metrics allocated to quality outcomes or culture indicators



Note: Major outliers excluded

Figure 19 provides an overview of statistically significant relationships identified in Wave 2. A blue line indicates that a relationship could be shown at a *p* value of less than 0.05. Metrics given in FDA Draft Guidance (FRN) and evaluated are shaded in gray. Lots pending disposition is a data point requested by FDA for verification of data validity supporting Lot Acceptance Rate.

Figure 19: Overview of statistically significant relationships

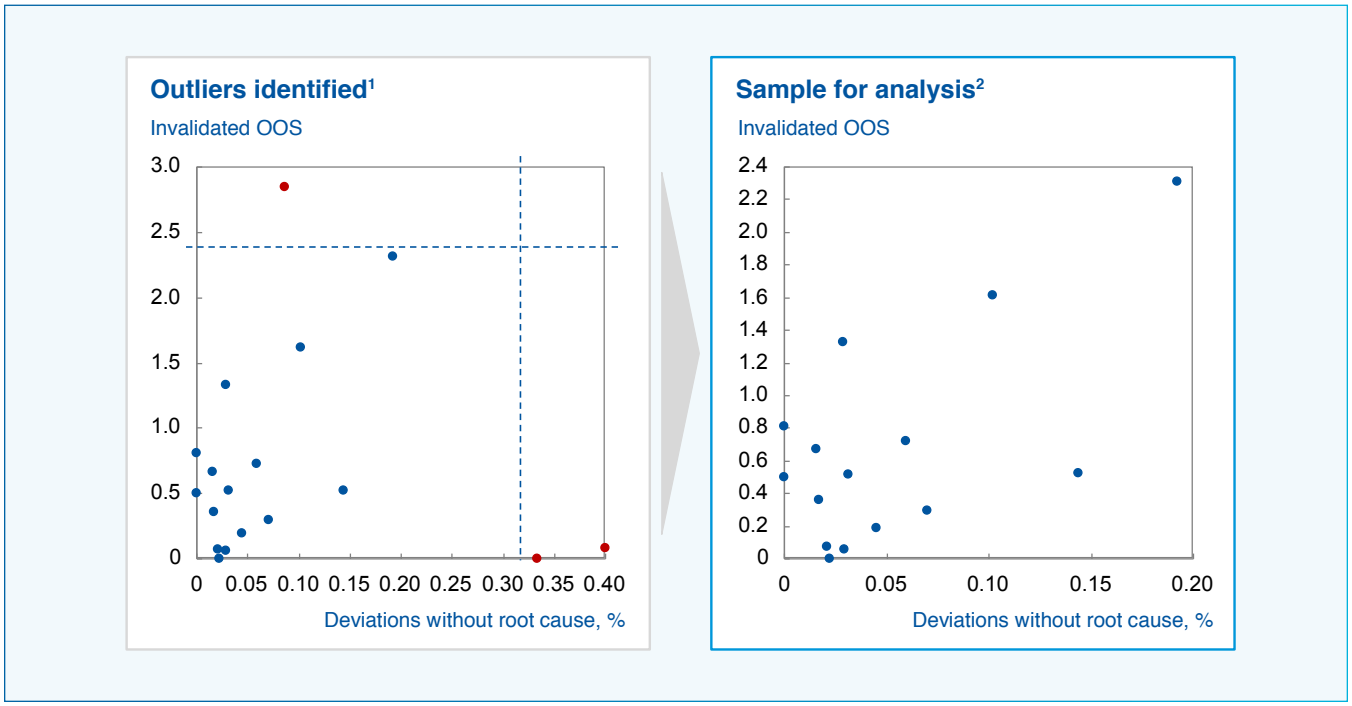


Note: Outliers more than two standard deviations or two interquartile ranges away from sample mean were excluded.
p value is probability that correlation between X and Y is zero, value below 0.05 indicates statistically significant results.

5.3.4 Outlier identification and analysis

Section 4, Data Analysis discusses treatment of outliers in detail. In summary, data points more than two standard deviations from the average or two interquartile ranges from the median for nonnormal distributions have been removed from the analysis. An example of outlier identification and reanalysis is given in Figure 20.

Figure 20: Outlier identification and sample analysis



[1] Data points more than two standard deviations from average or two interquartile ranges from median for nonnormal distributions. No judgments were made regarding specific data points, due to limited context and insights on participating sites
[2] Outliers removed from analysis

In the left hand graph, the red dots are identified as outliers based on them being more than two standard deviations from the average. If these data points are removed from the analysis the resulting scatter plot is given on the right hand side. For completeness, full data sets are given in Appendix 3 and explanation of exclusion of outliers is given in relevant figures.

5.3.5 Analysis of FDA Draft Guidance metrics

5.3.5.1 Analysis of FDA guidance metrics using FDA definitions

This section discusses findings from Wave 2 with each proposed FDA Draft Guidance Metric and a data point in the order given in the FDA Draft Guidance.

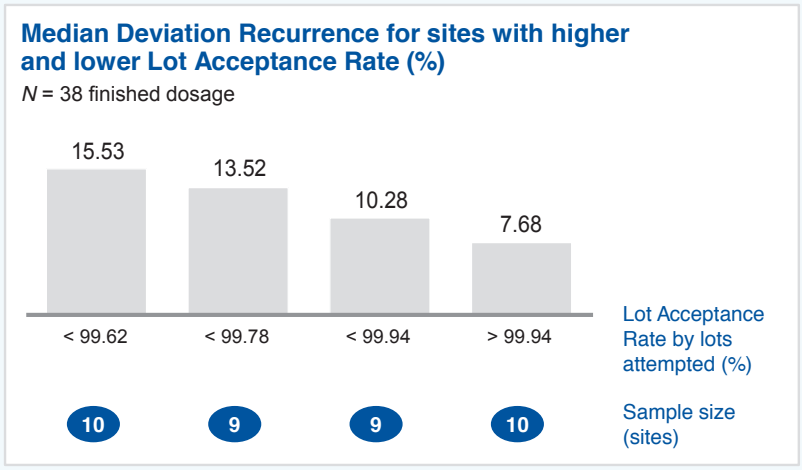
Lot Acceptance Rate

Lot Acceptance Rate by lots attempted does not have a relationship with an external quality outcome or a culture indicator.

It does have a relationship with other internal quality outcomes: Invalidated OOS per lot (not FDA definition) for sterile sites, and Deviations Recurrence.

In Figure 21 the relationship is shown between Lot Acceptance Rate with a denominator of lots attempted as quartiles and Deviations Recurrence Rate. A scatter plot (not shown) indicates a relationship of $p < 0.05$. Lower Deviations Recurrence Rate is linked to higher Lot Acceptance Rate.

Figure 21: Relationship of Lot Acceptance Rate and Deviations Recurrence Rate

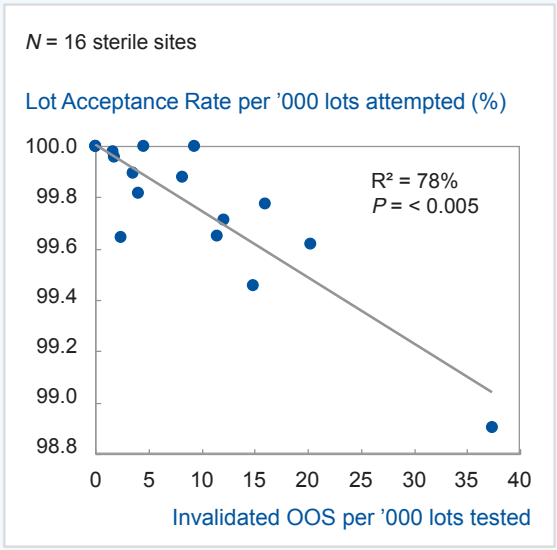


Note: no outliers excluded.

Sites with lower Deviations Recurrence Rate might be expected to have higher Lot Acceptance Rate.

A statistical relationship was found for 16 sterile manufacturing sites between Lot Acceptance Rate with a denominator of lots attempted and Invalidated OOS per lots tested. The Invalidated OOS Rate metric differs from the definition given in the FDA Draft Guidance. The relationship is shown in Figure 22.

Figure 22: Relationship of Lot Acceptance Rate per lot attempted with Invalidated OOS Rate per lot tested



Note: Outliers more than two standard deviations away from sample mean excluded: one US-based CMO site on Lot Acceptance Rate and one European Rx site on Invalidated OOS Rate.
R^2 measures to what extent metric Y (dependent variable) is explained by the variability of metric X (independent variable).
p value is probability that correlation between X and Y is zero, value below 0.05 indicates statistically significant results.

The relationship is hard to understand since Invalidated OOS Rate is a measure of laboratory performance, while Lot Acceptance Rate is related to manufacturing shop floor quality performance.

Product Quality Complaint Rate

Product Quality Complaint Rate ("Total Complaints" in Figure 19) as defined in the FDA Draft Guidance (i.e., divided by number of lots released) does not have a relationship with any internal quality outcome or culture indicator.

Invalidated OOS Rate

Invalidated OOS Rate using the FDA definition (which has two denominators) does not have a relationship with either an external quality outcome or a culture indicator. Furthermore, the rationale for normalization by two factors is hard to understand.

Feedback from laboratory CMOs, however, indicates Invalidated OOS Rate using a single denominator could be valuable as a metric for laboratory quality and/or method robustness.

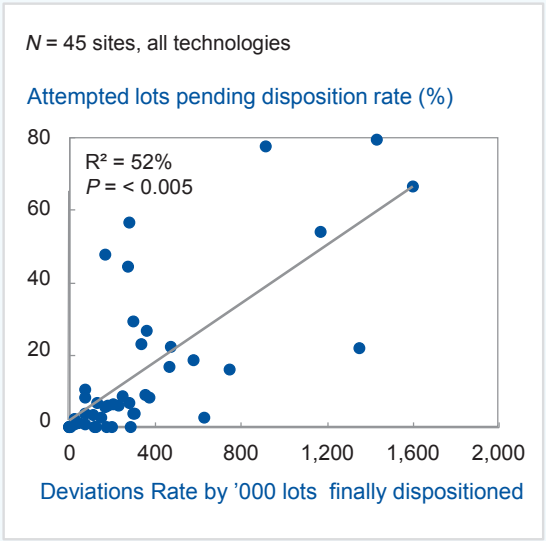
APR/PQR on-time rate

Wave 1 showed that APQ and PQR on-time rate had low differentiating power, being reported as 100% by most of the sites. Determination of APR/PQR on-time rate was not included in Wave 2.

Attempted lots pending disposition for more than 30 days data point

Attempted lots pending disposition for more than 30 days data point is related to the internal quality outcome "Deviations Rate per lots dispositioned," as shown in Figure 23.

Figure 23: Relationship of attempted lots pending disposition for more than 30 days with Deviations Rate



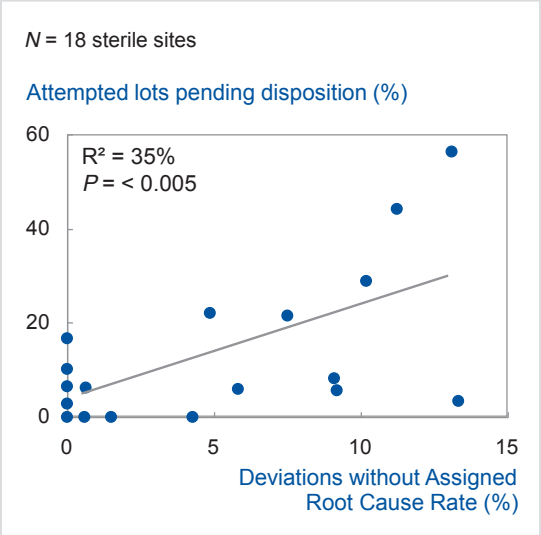
Note: Major outliers excluded—more than two standard deviations away from sample mean—five sites were excluded, all drug substance ones
R^2 measures to what extent metric Y (dependent variable) is explained by the variability of metric X (independent variable).
p value is probability that correlation between X and Y is zero, value below 0.05 indicates statistically significant results.

This could be a logical relationship, since pending disposition is often related to deviations investigations.

Most drug substance sites are outliers on the pending disposition metric due to process specifics, which could be due to extended testing or deviation investigation times.

Attempted lots pending disposition for more than 30 days data point also has a relationship with the culture indicator Deviations without Assigned Root Cause Rate. Figure 24 shows findings for 18 sterile manufacturing sites.

Figure 24: Relationship of attempted lots pending disposition for more than 30 days data point with Deviations without Assigned Root Cause Rate



Note: Outliers more than two standard deviations away from sample mean were excluded: One Puerto Rico Rx site was excluded on lots pending disposition.
R^2 measures to what extent metric Y (dependent variable) is explained by the variability of metric X (independent variable),
p value is probability that correlation between X and Y is zero, value below 0.05 indicates statistically significant results.

This relationship could be considered reasonable since pending disposition for sterile manufacturing sites may be influenced by longer and more challenging investigations where root cause is not confirmed.

As discussed in Section 5.2.1 and shown in Figure 7, the attempted lots pending disposition data point has a relatively high collection burden (1.3 hours per product)

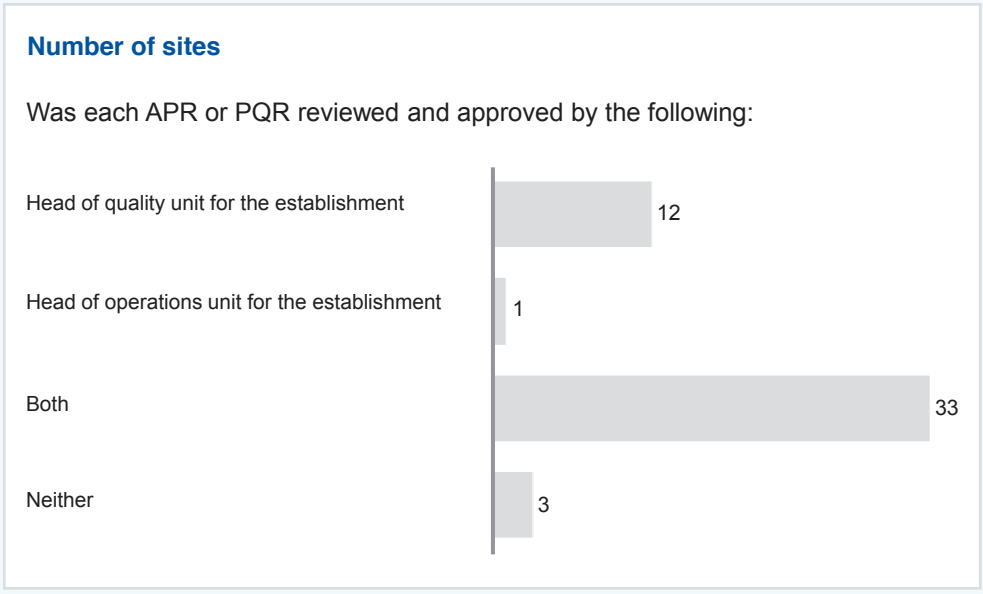
Optional metrics

In their Draft Guidance, FDA requested comments on optional metrics for quality culture (senior management engagement and CAPA effectiveness) and process capability/performance.

Quality culture, senior management engagement

Wave 2 findings for FDA the senior management engagement quality culture indicator are given in Figure 25.

Figure 25: Senior management engagement

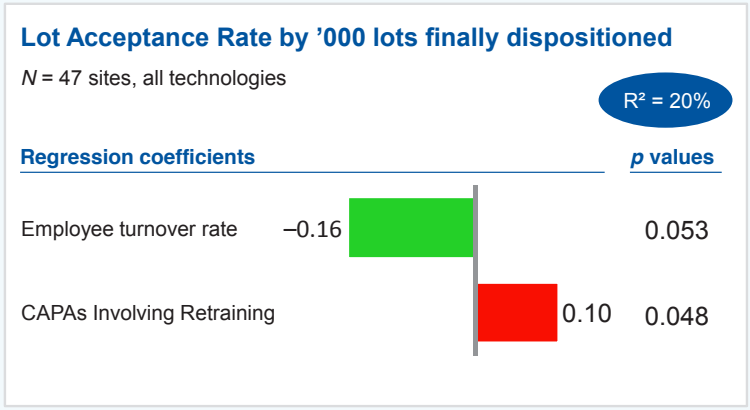


For more than 90% of pilot products, APR/PQR is reviewed and approved by head of the quality unit and in most of those cases the head of the operations unit, as well. As a measure, this shows management engagement, but the extent of this engagement cannot be determined, as indicated in ISPE's response to the Draft Guidance. [4]

Quality culture: CAPA Effectiveness Rate

CAPA effectiveness rate determined by CAPA requiring retraining rate appears to have limited usefulness. CAPA requiring retraining rate showed no direct relationship with other FDA-proposed metrics. It did have a multivariate relationship to Lot Acceptance Rate using lots dispositioned as a denominator, as shown in Figure 26.

Figure 26: Multivariate relationship of CAPA Requiring Retraining Rate and Employee Turnover rate to Lot Acceptance Rate using lots dispositioned



Note: Outliers more than two standard deviations away from sample mean were excluded.
R^2 measures to what extent metric Y (dependent variable) is explained by the variability of given set of independent metrics.
p value is probability that correlation between variables is zero, value below 0.05 indicates statistically significant results.

Higher Employee Turnover Rate and a lower CAPA Requiring Retraining Rate are linked in a multivariate relationship to lower Lot Acceptance Rate with lots dispositioned as denominator. This complex relationship is difficult to understand.

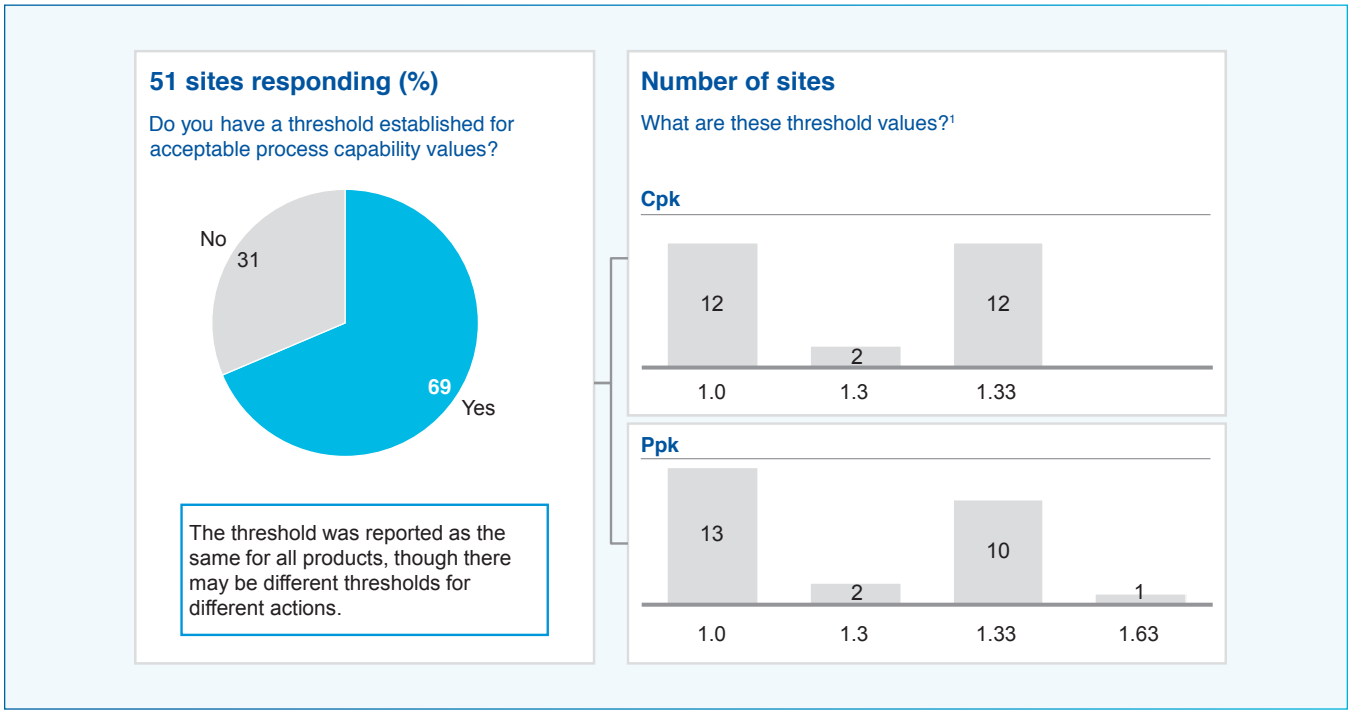
In conclusion, suggested Optional Metrics for quality culture are limited in usefulness.

Process capability/performance

FDA questions relating to process capability/performance are discussed below. Wave 2 questions were wider-ranging than the FDA's, nonetheless; some answers to FDA questions were obtained.

Figure 27 indicates responses related to questions regarding application of thresholds for process capability/performance.

Figure 27: Process capability/performance thresholds



[1] Some sites reported using more than one value, e.g., < 1.0 requires action plan; 1.0 – < 1.33 requires investigation.

A simple response to the first part of FDA Optional Metric 3 first bullet question:

A "yes" or "no" value of whether the establishment's management calculated a process capability or performance index for each critical quality attribute (CQA) as part of that product's APR or PQR.

would be 100% "no." Many CQAs do not lend themselves to calculation of process capability or performance values as shown in responses summarized in Figure 28.

Figure 28: Reasons and examples for excluding CQAs when calculating process capability/performance indices

Reasons and examples behind attributes exclusion
Provided by 12 sites, although 27 acknowledged having exclusions

- Quantitation limit for metrics, such as:
  - Pass/Fail pharmacopoeia tests such as Uniformity of Dosage Units
  - Sterility
  - Endotoxins
  - Identification tests
- Data type, such as:
  - Results are not normally distributed (e.g., only discrete number of different data points are available)
  - Tests produce "complies" vs. "non-complies" values
  - Not enough precision in the data relative to the specification acceptance criteria
- Cpk/PpK are not calculated for all quantitative quality attributes, only for the most representative
- Capabilities are calculated only on intermediate specifications and CQAs, not on CPPs
- Cpk/Ppk are not calculated for in-process tests and final drug potency

In response to the last part of the first bullet question in Wave 2,

… as part of that product's APR or PQR.

34 sites responded and 59% (20) answered "yes"; 41% (14) replied "no." These responses indicate it is not routine to include process capability/performance values in APRs or PQRs.

Responses given in Figure 27 also help answer FDA second bullet question:

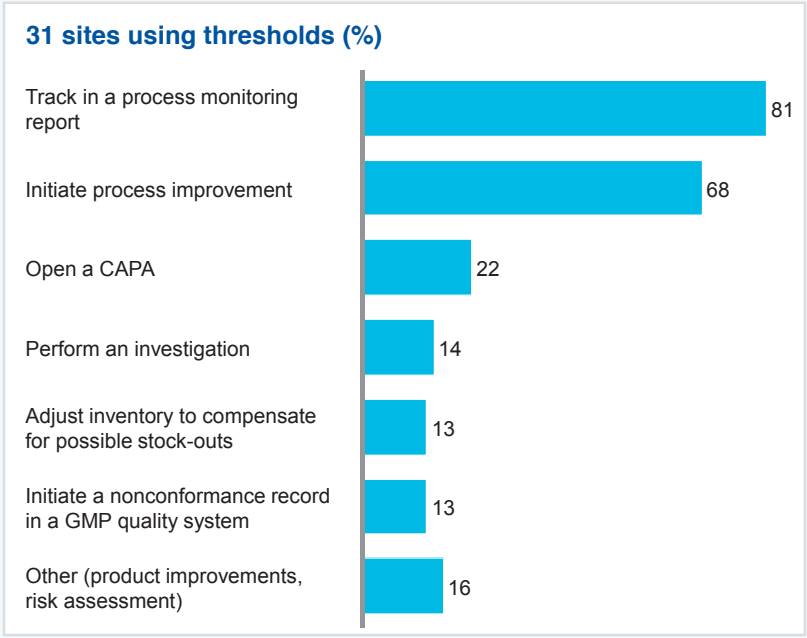
A "yes" or "no" value of whether the establishment's management has a policy of requiring a corrective action or preventive action (CAPA) at some lower process capability or performance index.

From Figure 27, 69% of sites responded that they applied thresholds for acceptable process capability values. Threshold choices varied from 1.0 to 1.63, with approximately equal numbers of sites applying threshold values of 1.0 and 1.33. The threshold was reported as the same for all products, although a different threshold may lead to different actions. Some sites, for example, apply one limit (e.g., 1.0) to an action plan, and the other limit (e.g., 1.33) to an investigation.

These responses are general and not restricted to application of a limit that triggers a CAPA, as required by the FDA second bullet question.

Answers to the FDA second bullet question—relating to application of a CAPA at some process capability or performance index—are given in Figure 29.

Figure 29: Actions taken in response to process capability/performance values being below a threshold



As an answer to second bullet question, only 22% of sites opened a CAPA with process monitoring, and improvement was the most common action when the threshold was exceeded. From Figure 27 it is assumed that threshold values of 1.0 or 1.3 are applied as a partial answer to FDA third bullet question:

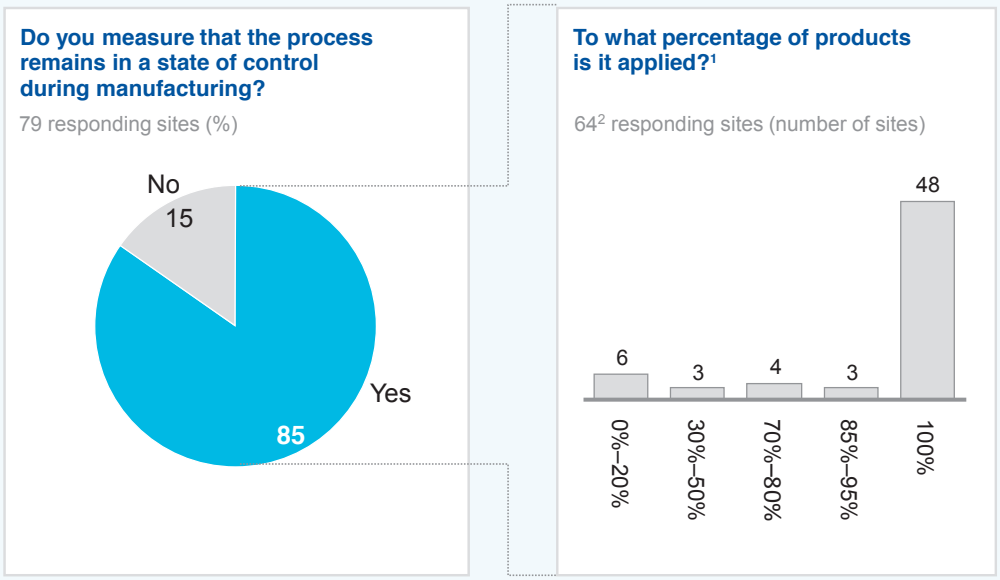
… what is the process capability or performance index that triggers a CAPA?

Wave 2 explored more widely how process capability and performance indices are used in industry.

5

In Figure 30, 85% of sites apply ongoing monitoring during production process, with a majority applying this to all their products.

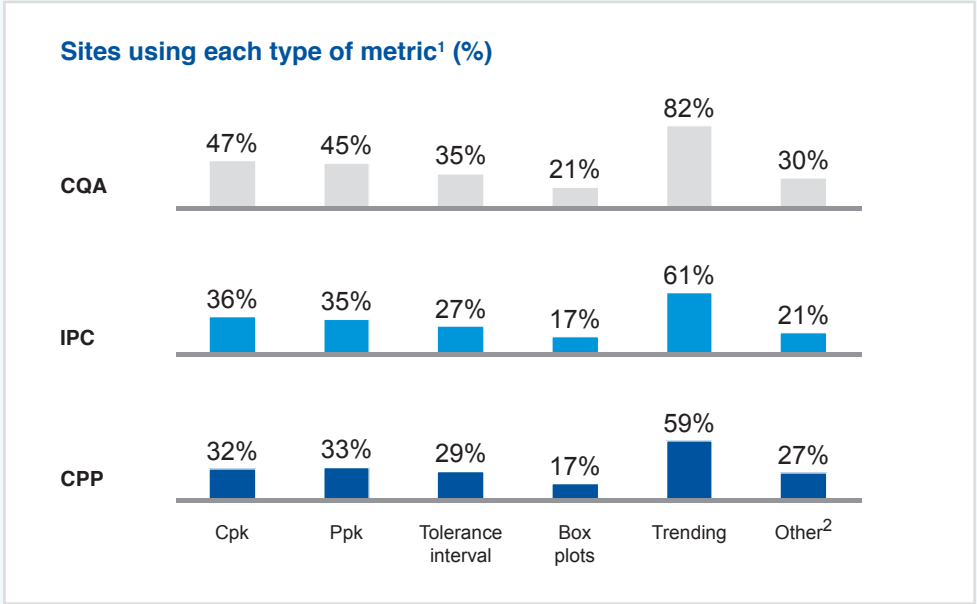
Figure 30: Percentage of sites that apply ongoing monitoring, and to what percentage of products



[1] When only some products are chosen, choice is based on risk approach to customer and importance for business.
[2] Three out of 67 sites did not provide share of products.
Based on 79 responses from Wave 2 participants and sites from Wave 1 that did not participate

The most common monitoring method for CQAs, in-process controls (IPCs) and critical process parameters (CPPs) is trending, as shown in Figure 31. This finding reflects Wave 1 results.

Figure 31: CQA, IPC, and CPP monitoring methods



[1] Out of 67 sites monitoring capability that provided detailed explanation
[2] Other mentioned metrics were Pareto charts; monitoring via excursions trending, I-charts, regression, 3 sigma, run/control charts
Based on responses from Wave 2 participants and sites from Wave 1 that did not participate again

5

Two-thirds of sites calculate Cpk/Ppk for legacy (existing) products quarterly or annually, and compare and/or trend the results. The median minimum number of batches for calculation of a process capability/performance index is 25, with the range being 3 to 30 (Figure 32).

Figure 32: Application of process capability/performance indices to legacy products

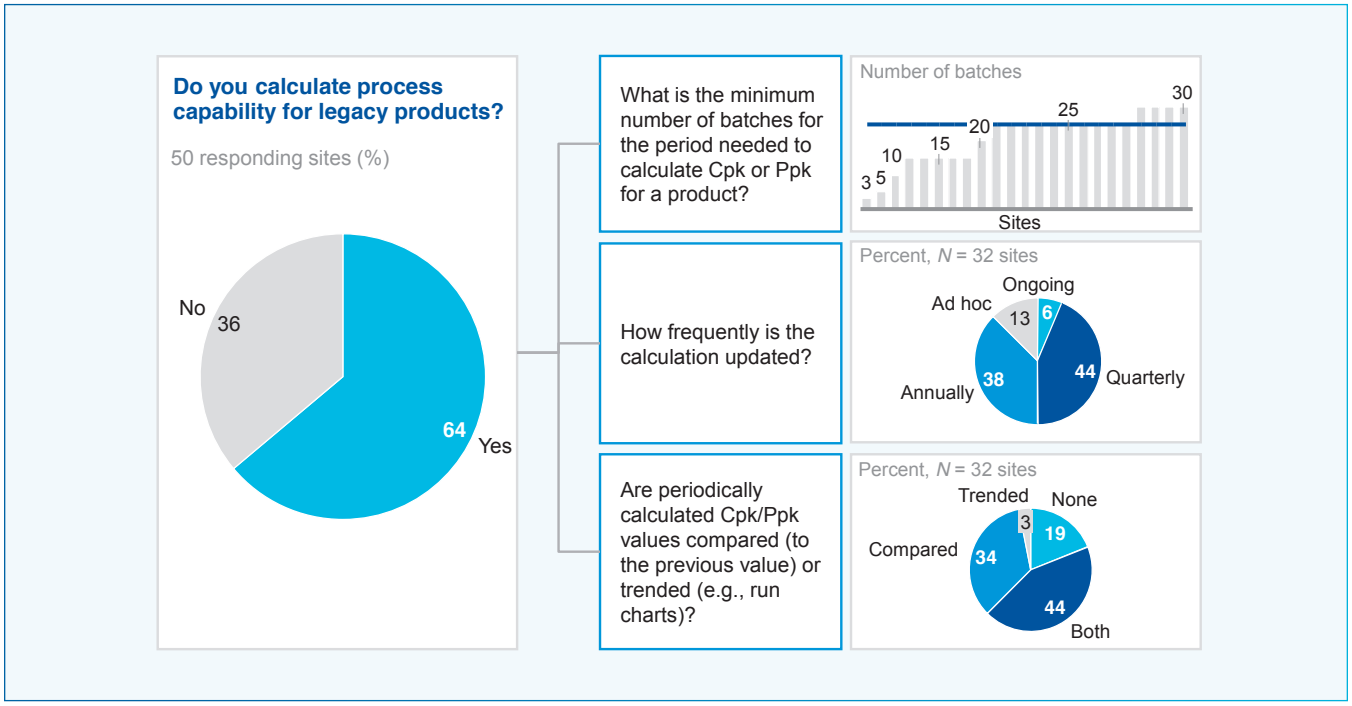
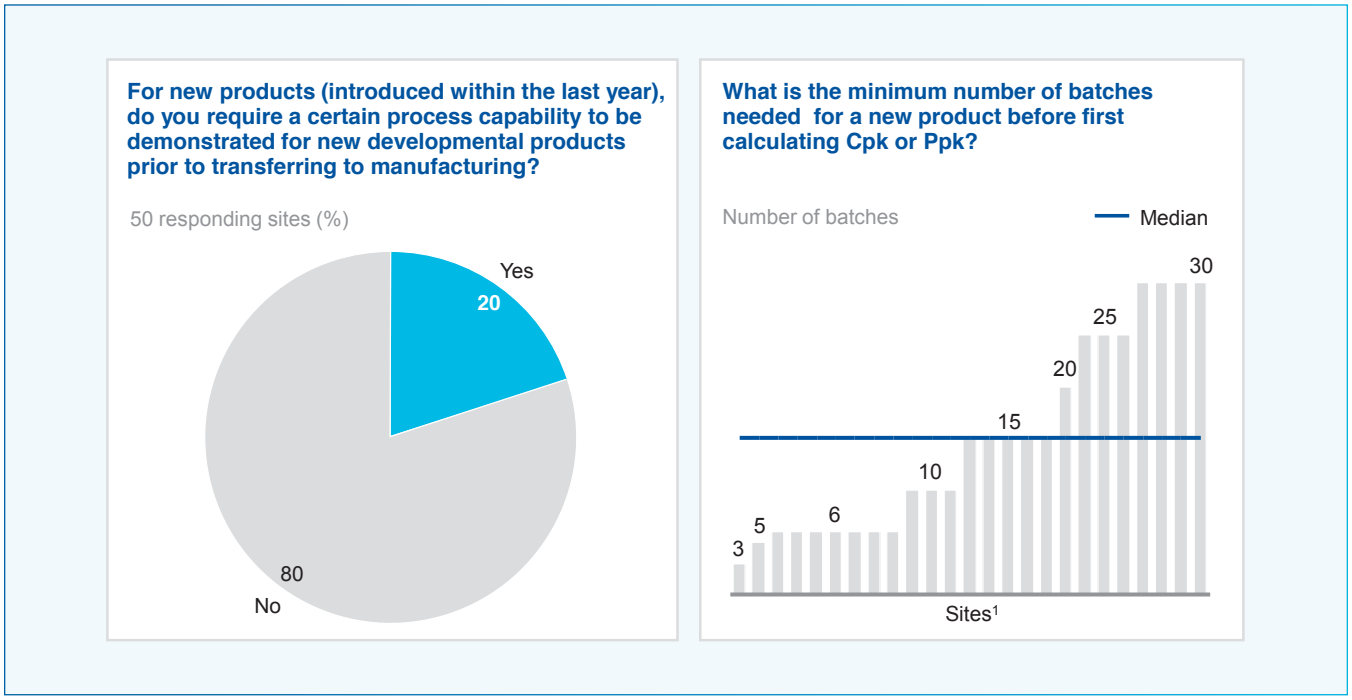


Figure 33 shows that 20% of sites indicated that new products should meet an unspecified Cpk/Ppk threshold level before commercialization.

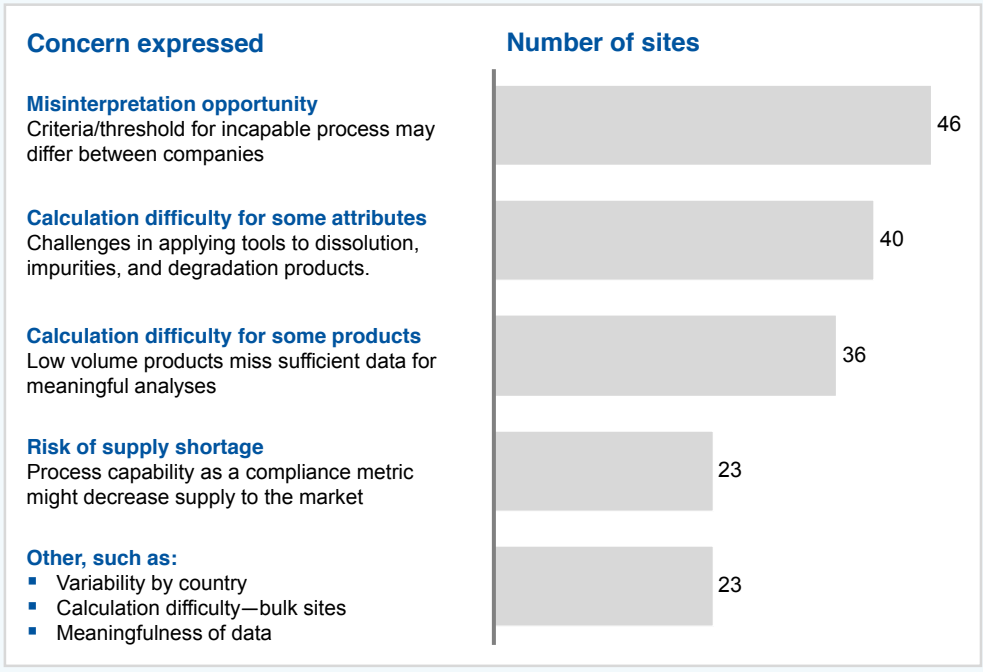
Figure 33: Process capability/performance indices applied during development



[1] Based on 25 sites providing this data, including 21 that do not require certain process capability for commercialization and replied "no" in the chart on left.

Feedback from ISPE's Process Capability team and Wave 1 participants indicated industry reservations regarding formal process capability/performance values and thresholds reporting to FDA. Wave 2 posed a series of questions to further explore these concerns, as shown in Figure 34.

Figure 34: Industry response relating to formal reporting of process capability/performance indices



Many sites expressed concern about reporting process capability outcomes to regulators.

Major concerns were:

- Misinterpretation of data
- Inability to apply to all CQAs
- Inappropriateness of calculation for some (e.g., low-volume) products

In conclusion, process capability/performance indices are widely used in industry to help control processes and identify continual improvement opportunities. These indices are used in a variety of ways, presumably since use is linked to specific products, processes, and situations. Respondents expressed reservations regarding process capability/performance indices as a reportable metrics.

Given that the sample in Waves 1 and 2 has a significant representation from large companies and sites with advanced technology and relatively good compliance, findings from an industry-wide population are likely to show less use of process capability measurements.

Conclusions relating to FDA Draft Guidance metrics

Metrics

Wave 2 pilot confirmed the importance of clear and detailed definitions and the need to check data from different reporting periods for consistency. The three metrics in the FDA Draft Guidance that were evaluated (Lot Acceptance Rate, Product Quality Complaint Rate and Invalidated OOS Rate) did not have relationships with external quality outcomes or culture indicators. Further discussion and some considerations and recommendations are given in the following section when data were collected using alternative definitions.

Optional Metrics

Wave 2 findings related to FDA-proposed Optional Metrics confirmed points made in ISPE's response to the FDA Draft Guidance: Optional Metrics have limited utility and were inconsistently applied. ISPE recommended that Optional Metrics be deferred and considered once the program is established.

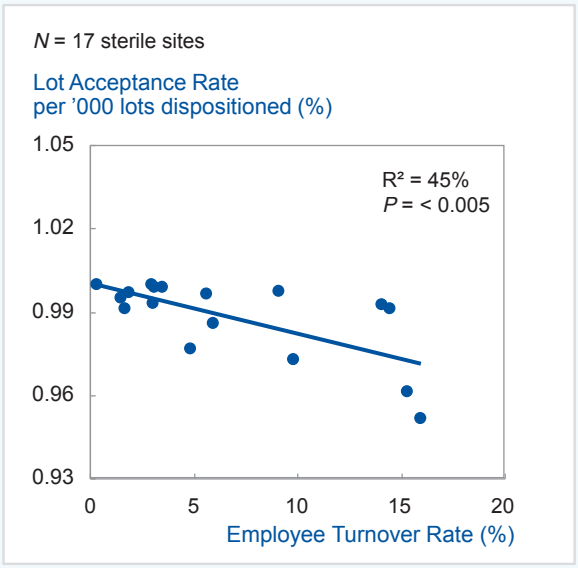
5.3.5.2 Analysis of FDA metrics based on alternative definitions

In general, alternative metrics proposed by ISPE showed better relationships.

Lot Acceptance Rate

Lot Acceptance Rate using a denominator of lots dispositioned does have a univariate relationship with the Employee Turnover Rate culture indicator in 17 sterile manufacturing sites, as shown in Figure 35. It also has a multivariate relationship with CAPA with Retraining Rate and Employee Turnover Rate, as shown in Figure 26.

Figure 35: Relationship of Lot Acceptance Rate and Employee Turnover Rate



Note: Outliers more than two standard deviations away from sample mean were excluded: One on Employee Turnover Rate (US-based Rx site), and one on Lot Acceptance Rate (India-based Rx site).
R^2 measures to what extent metric Y (dependent variable) is explained by the variability of metric X (independent variable).
p value is probability that correlation between X and Y is zero, value below 0.05 indicates statistically significant results.

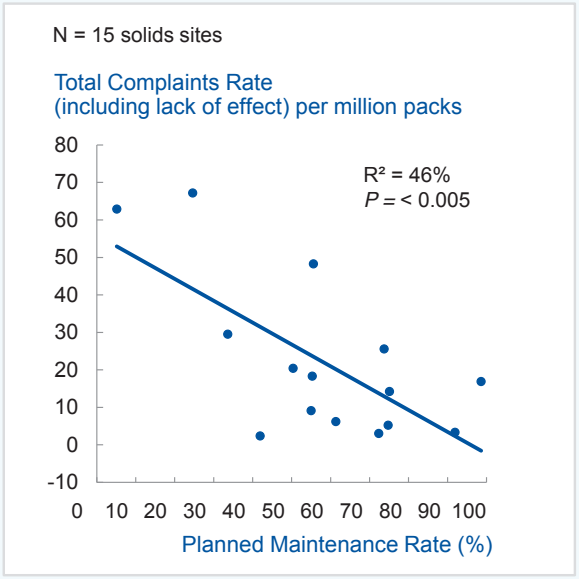
Employee Turnover Rate may be influenced by some of the cultural factors that also underline Lot Acceptance Rate, such as attention to employees' mindsets, capabilities building, and leadership focus on shop floor issues. Employee Turnover Rate could also be affected by external factors such as local unemployment rate and competition from local companies; for this reason, full explanation of the relationship is not possible.

The multivariate relationship of Lot Acceptance Rate by lots dispositioned to Employee Turnover Rate and CAPA Involving Retraining (Figure 26) is complex and hard to understand.

Product Quality Complaint Rate

Product Quality Complaint Rate (Total Complaints in Figure 19) divided by number of packs does have relationships with the Planned Maintenance Rate and Proportion of CAPAs with Preventive Action culture indicators, as shown in Figures 36 and 37.

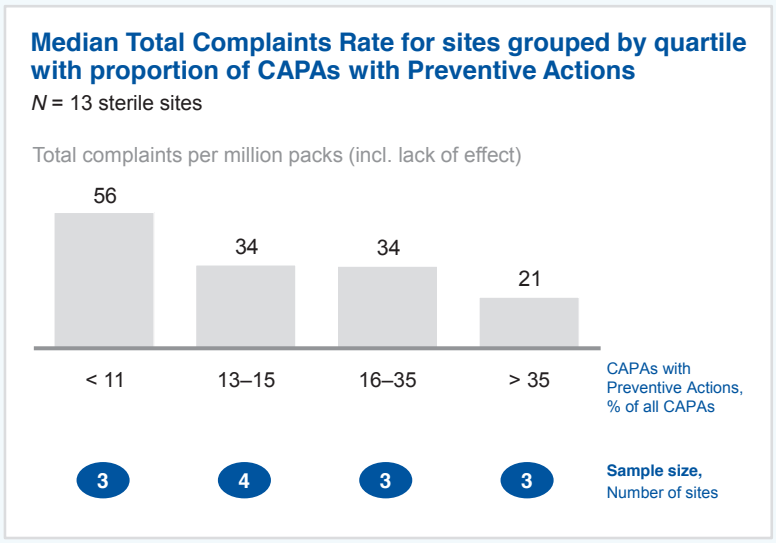
Figure 36: Relationship of total complaints and planned maintenance rate



Note: Outliers more than two standard deviations away from sample mean were excluded: two sites on Total Complaints, both with very high LOE rates.
R^2 measures to what extent metric Y (dependent variable) is explained by the variability of metric X (independent variable)
P- value is probability that correlation between X and Y is zero, value below 0.05 indicates statistically significant results.

This relationship may be related to common factors that influence both Planned Maintenance Rate and Total Complaints, such as focus on prevention, operational excellence, and quality improvement mindset.

Figure 37: Relationship of Total Complaints and Proportion of CAPA with Preventive Actions



Note: Outliers more than two standard deviations away from sample mean were excluded: three sites on total complaints with high LOE rates, as well as one European Rx site that reported 100% CAPAs with preventive action.

Higher proportion of CAPA with Preventive Action is related to lower Total Complaints Rate for these 13 sterile sites. This relationship may be related to the common factors that influence both Proportion of CAPA with Preventive Action and Total Complaints Rate—such as focus on prevention, operational excellence, and quality improvement mindset.

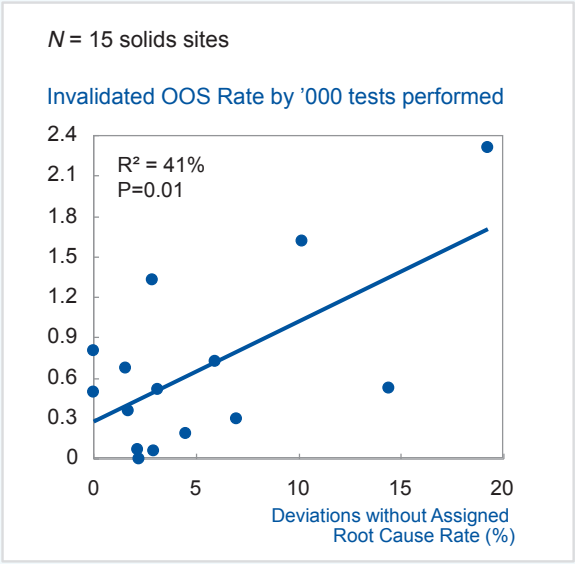
Invalidated OOS Rate

Invalidated OOS Rate should be normalized by a single denominator—either number of tests or number of lots tested.

Invalidated OOS Rate by lots tested does show a relationship with the internal quality outcome Lot Acceptance Rate per lot attempted, as discussed in Section 5.3.5.1 and shown in Figure 22. The relationship is hard to understand since Invalidated OOS Rate is a measure of laboratory performance, while Lot Acceptance Rate is more related to manufacturing shop floor quality performance.

Invalidated OOS Rate by tests performed shows a relationship with the Deviations without Assigned Root Cause Rate culture indicator (Figure 38).

Figure 38: Relationship of Invalidated OOS Rate and Deviations Without Assigned Root Cause Rate

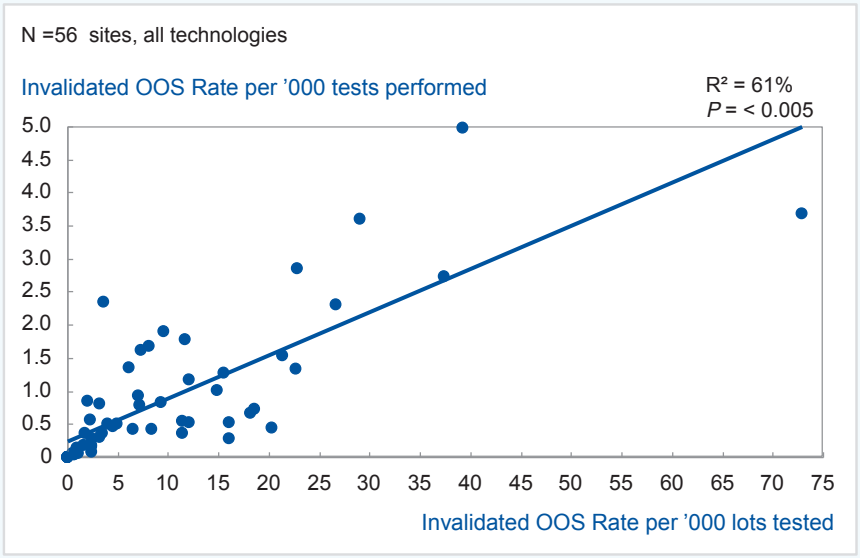


Note: Outliers more than two standard deviations away from sample mean were excluded: two on deviations without root cause (LATAM-based, small-scale sites with < 10 deviations annually), and one on Invalidated OOS (Puerto Rico–based Rx site)
R^2 measures to what extent metric Y (dependent variable) is explained by the variability of metric X (independent variable).
p value is probability that correlation between X and Y is zero, value below 0.05 indicates statistically significant results.

Figure 38 shows a relationship for 15 OSD sites ($P = 0.01$; $R^2 = 41\%$). Invalidated OOS Rate is a measure of laboratory performance, while Deviations without Assigned Root Cause is related to shop floor quality performance. It is a hypothesis that both are influenced by factors such as problem-solving skills, focus on error reduction, and quality improvements, which may explain this relationship. Sites with better shop floor quality may be likely to have better laboratory performance as well.

Invalidated OOS Rate by lot tested and by number of tests are interrelated, as shown in Figure 39.

Figure 39: Relationship of Invalidated OOS Rate by tests and by Lot



Note: Outliers more than two standard deviations away from sample mean were excluded: one lab, one US-based drug substance site, and one Europe-based solids site
R^2 measures to what extent metric Y (dependent variable) is explained by the variability of metric X (independent variable)
p value is probability that correlation between X and Y is zero, value below 0.05 indicates statistically significant results.

This curve suggests either number of tests or number of lots could be used as a denominator in the Invalidated OOS Rate calculation.

Conclusions from FDA Metrics using alternative definitions

Lot Acceptance Rate using a denominator of lots dispositioned does have a relationship with the Employee Turnover Rate culture indicator.

Product Quality Complaint Rate with a denominator of number of packs has relationships with two culture indicators: Planned Maintenance Rate and Proportion of CAPAs with Preventive Actions

Invalidated OOS Rate should be normalized by a single denominator of either number of tests or number of lots tested; both options are interrelated. Invalidated OOS Rate has a relationship the Deviations without Assigned Root Cause culture indicator.

5.3.5.3 Other relationships relevant to FDA metrics

Other Wave 2 findings relevant to the three FDA Draft Guidance metrics evaluated are discussed below in order of listing in the FDA Draft Guidance.

Lot Acceptance Rate

Discussion in sections above suggests that definition of Lot Acceptance Rate would be improved if lots dispositioned is used instead of lots attempted. Lots dispositioned was the data point defined in Wave 1 and also used in Wave 2.

Following completion of the Wave 2 analysis, an alternative to lots dispositioned has been considered. Appendix 4 discusses the pros and cons (positives and negatives) of lots attempted and lots dispositioned.

Conclusions from this discussion are:

- A rationale for using lots attempted based on the US Code of Federal Regulations (CFR) Title 21, Parts 210 and 211 [7] and ICH Q7 [8] is very hard to understand. The phrase is not used in those documents and, therefore, not defined.
- "Disposition" is used in CFR 210 and 211, the preamble to 210 and 211, and in ICH Q7.
- FDA, however, did not like use of "disposition" as an alternative to "release/reject."
- "Disposition" as used in CFR 210 and 211 does not clearly refer to reject/release decisions of a lot of drug product; as used in ICH Q7 it does not clearly refer to release/rejection of a lot of drug substance.
- As used in CFR 210 and 211, ICH Q7, and to some extent in the preamble to 210 and 211, "disposition" is more associated with components, containers, closures, and labeling materials rather than release/reject decisions of drug substances/drug products. It also is associated more with rejection than drug product or drug substance release.
- A word or phrase other than "disposition" should be considered.

A recommended alternative definition to "lots finally dispositioned" is:

Lots released or rejected: Total number of lots for commercial use produced and/or packaged on-site that went through final release/reject decision during the period—i.e., were released for shipping or rejected (for destruction). Rejections should be counted at whatever production stage they occurred. "Release" refers only final release for shipping (whether shipping bulk to another site or final product to market). It excludes lots that have been sent for rework or put on hold/quarantined in this period, and hence are not finally released or rejected. It excludes lots that are not produced or packaged on-site, but released only for CMOs.

Rationale

1. “Release” and “reject” are common phrases used throughout CFR 210 and 211, and ICH Q7 and are understood by practitioners. “Release” is often associated with a release for [something], such as “release for distribution,” “release for manufacture,” or “release for shipping.” Neither word is defined in CFR 210 and 211, or ICH Q7. In this definition, “release” is associated with an appropriate action as given by precedent in CFR 210 and 211.
2. Although not defined as such in CFRs 210 and 211, and ICH Q7, "reject" is used in CFR 210.3, Definitions in sub section (b)(20):

   (20) "Acceptance criteria" means the product specifications and acceptance/ ***rejection*** *criteria, such as acceptable quality level and unacceptable quality level, with an associated sampling plan, that are necessary for making a decision to accept or* ***reject*** *a lot or batch (or any other convenient subgroups of manufactured units).*

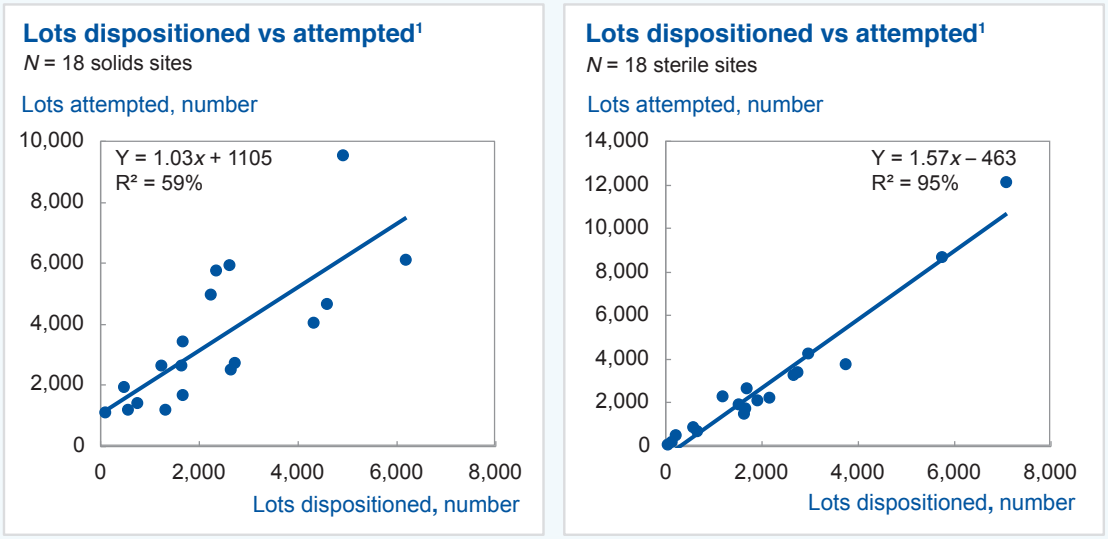
3. The FDA definition of Product Quality Complaint Rate included a reference to “product released,” which gives further precedent for use of a phrase that covers “release” (and “reject”):

   Product Quality Complaint Rate *= the number of product quality complaints received for the product divided by the total number of lots of the product released in the same timeframe.*

Relationship of lots attempted and lots dispositioned

The Wave 2 data points lots dispositioned and lots attempted are highly related, but differ in magnitude, as shown in Figure 40.

Figure 40: Relationship of lots dispositioned and lots attempted



[1] Average annual lots reported
Note: Outliers more than two standard deviations away from sample mean were excluded: one European Gx site in solids sample, one European Rx site in sterile sample.
R^2 measures to what extent metric Y (dependent variable) is explained by the variability of metric X (independent variable).

In Figure 40 there is a relationship of "lots dispositioned" and "lots attempted," especially for 18 sterile manufacturing sites (R^2 = 95%).

The systematic difference in values is explained by examining how and when in a manufacturing process firms assign a lot number. For example relating to definition of a lot attempted, different workflow designs result in variety of practices as to when new lot number is assigned. A single lot number could be associated with a single release step or with multiple release steps or no release step. These differences are a function of work order/electronic batch manufacturing instruction design.

A lot could be considered “attempted” when a lot number is issued for it. In drug substance synthesis, one batch of isolated intermediate may lead to several batches of drug substance, or several batches of isolated intermediate may be used in a single batch of drug substance. Firms often assign unique batch identification numbers for each step in the process. Batch production records clearly reference actual batch use to meet GMP traceability requirements. Such work flows lead to a systematic difference between lots “attempted” and those “dispositioned” when a release decision is made.

Similar examples apply to drug product manufacture where, for example, multiple granulations could be blended to make one lot for tableting, or a single tableting step could lead to multiple coating steps followed by multiple packaging steps.

Sites with continuous processing may have one lot number assigned from end to end.

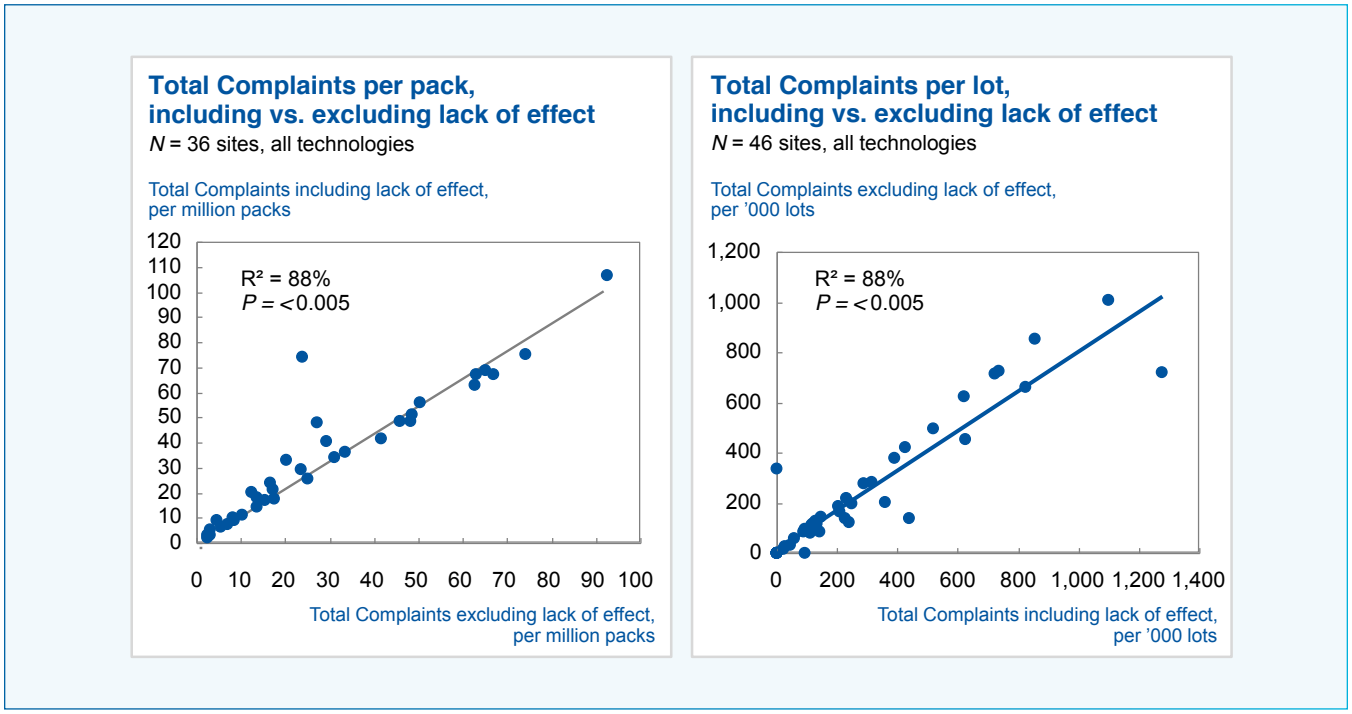
More explanation of the reasons for a systematic difference between lots attempted and lots dispositioned—including the impact on burden—is given in Appendix 6.

In summary, for most of the sample, lots-attempted data values were much higher than lots dispositioned” (up to a factor of 2 or even more for solids), since many sites issue new lot numbers at interim stages. There is, however, a relationship between lots attempted and lots dispositioned, which indicates that lots attempted reflects administrative differences between sites more than it does real structural differences. Given that lots attempted is a high-burden metric, using an alternative definition as a denominator merits consideration.

Product Quality Complaint Rate

The Product Quality Complaint Rate was analyzed by comparing Total Complaints, including and excluding lack of effect, as shown in Figure 41.

Figure 41: Total Complaints including and excluding lack of effect values



Note: Outliers more than two standard deviations away from sample mean were excluded: seven sites for the left (two USA-based, two LATAM-based, and three Europe-based) and seven sites for the right chart: four Rx and three OTC
R^2 measures to what extent metric Y (dependent variable) is explained by the variability of metric X (independent variable)
p value is probability that correlation between X and Y is zero, value below 0.05 indicates statistically significant results.

Relationships (R^2 = 88%) are observed whether the denominator is number of packs or number of lots. This suggests that either lots or packs could be used as denominator, however, some practitioners consider that number of packs is more appropriate, since it reduces variability due to lot size. Furthermore, from Figure 19, Total Complaints per Pack has relationships with some culture indicators whereas Total Complaints per Lot does not, again suggesting that "per pack" is a more appropriate denominator.

Overall conclusion from analysis of FDA metrics

FDA metrics evaluated using FDA-proposed definitions were not found to have relationships with quality outcomes or directly with cultural indicators.

Alternative metric calculations proposed by ISPE showed better relationships: the three FDA Draft Guidance metrics with ISPE-recommended definitions, for example, have relationships with a culture indicator.

Wave 2 findings suggest that definitions should be adjusted for these three FDA Draft Guidance metrics:

- Lot Acceptance Rate should have a different denominator. “Lots dispositioned” is the data point and definition used in Wave 2. The alternate calculation produced more appropriate relationships and the burden to report `lots attempted' data was high. Subsequent consideration has suggested that “lots released or rejected” should be used as an alternative to “lots disositioned.”
- Product Quality Complaint Rate should use the “number of packs” denominator, since this calculation produced more appropriate relationships and reduced variability due to lot size (e.g. from packing).
- Invalidated OOS should be normalized by number of tests or number of lots tested. There was no difference which denominator to use and relationships were found for both. The rationale for the FDA-proposed double normalization is hard to understand and this calculation does not produce relationships.

Lots pending disposition data point has limited usefulness. Lots pending disposition values have a relationship with Invalidated OOS, a relationship, which is hard to understand and rationalize. It is not related to any outcomes and drives high burden. It also showed high variability as a metric, with extreme outliers especially in drug substance sites.

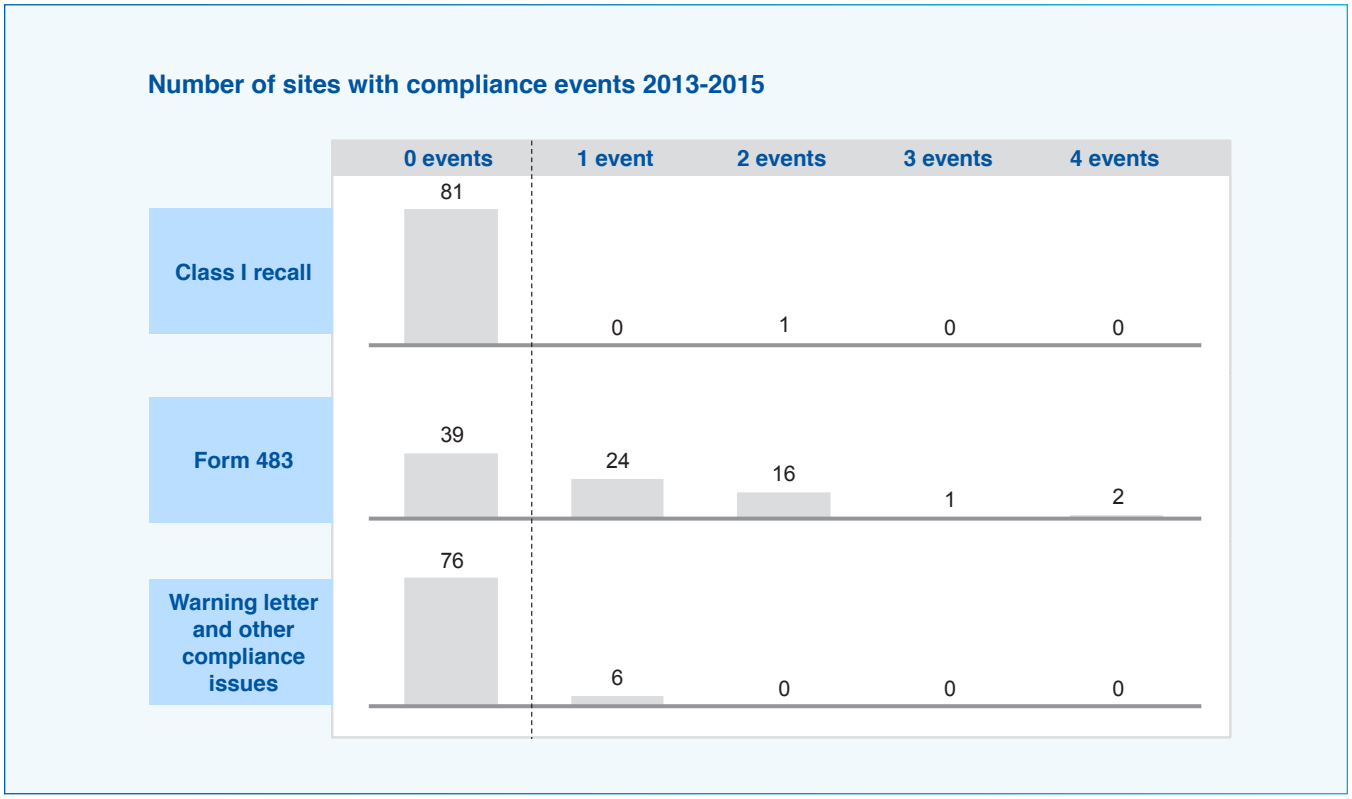
Wave 2 findings related to FDA-proposed Optional Metrics confirmed points made in ISPE’s response to the FDA Draft Guidance: Optional Metrics have limited utility and were inconsistently applied. ISPE recommended that Optional Metrics should be deferred and considered once the program is established.

5.3.5.4 Link to compliance status

Wave 2 data analysis prompted the question “Can metrics be linked to or predict a site’s compliance status?”

To obtain information relatively quickly, a search was made using the FDA website, FDAzilla, and press releases for publicly available information such as warning letters, number of 483s resulting from inspections, and number of Class 1 recalls for all sites in Waves 1 and 2. A summary of resulting information for the period 2013 to 2015 is given in Figure 42.

Figure 42: Publicly available compliance information



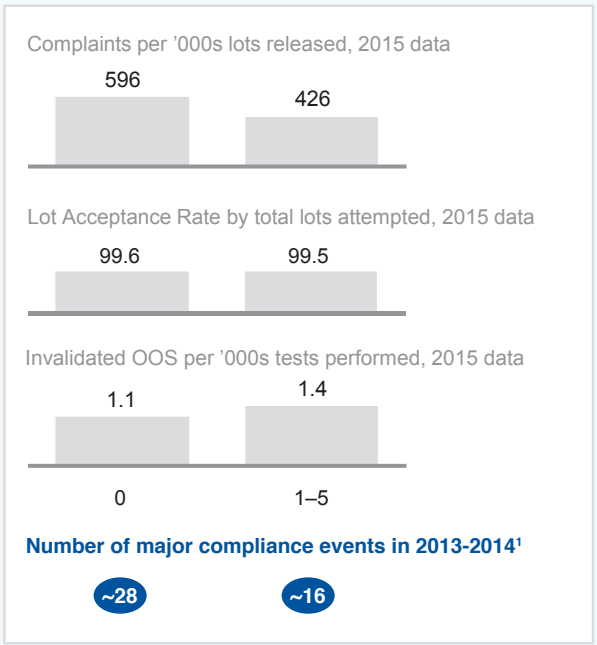
Source: FDA website, FDAzilla, press search, sample of 82 sites

The most common compliance issue was Form 483 notices, which were issued to approximately half the sites in the 3-year period. Class I recalls, warning letters, and other major compliance issues were very rare; one site had two recalls, and six sites received warning letters.

Whatever the numbers, it is impossible to judge the severity of these compliance events or determine if they are related to compliance status. In some cases, for example, a Form 483 may be easily addressed and cause no significant change in site practices; in others it may have implications that extend over a long period of time and could affect production output. Hence, number of 483s may not be related to compliance status.

Wave 2 Pilot performance for each site was evaluated and compared with values of the three FDA Draft Guidance metrics using FDA definitions for Lot Acceptance Rate and Product Quality Complaint Rate, as well as the ISPE-recommended definition for Invalidated OOS by number of tests performed. Findings are given in Figure 43.

Figure 43: FDA Draft Guidance metrics compared to compliance performance, 2013–2014



[1] Warning letters, Form 483 notices, import bans, and Class I recallsrelated to GMP compliance

Source: FDA website, FDAzilla, press search

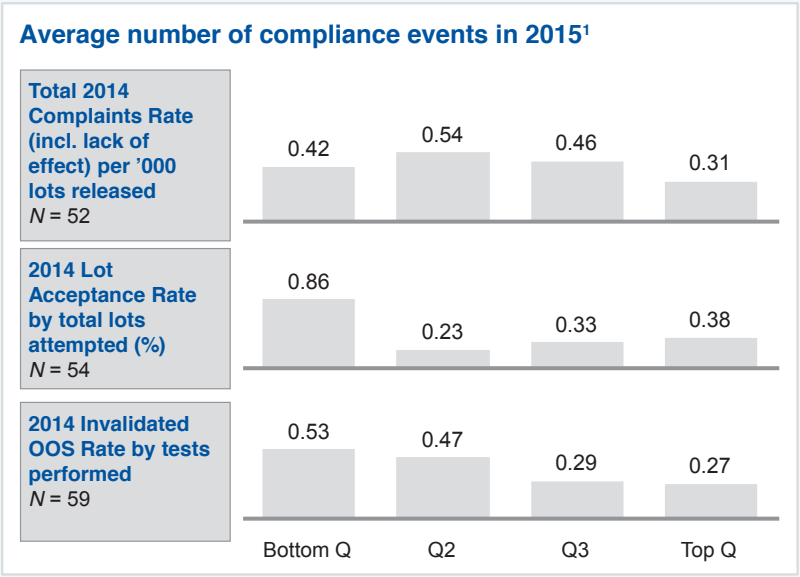
To try to develop some type of analysis all "compliance events" are considered equal (an unlikely assumption) and summed. It is recognized that this analysis may be "lagging" relative to generation of quality metric values. Additionally the grouping of number of "major compliance events" is judgmental.

For sites with one to five compliance events Lot Acceptance Rate was approximately similar to sites with zero events. Product Quality Complaint Rate was lower for sites with one to five compliance events. Only Invalidated OOS Rate showed a slight increase with one to five compliance events. An hypothesis is that the compliance issue itself may lead to a lower Product Quality Complaint Rate, because products with issues are not shipped to market. Similarly, a higher Invalidated OOS Rate for sites with compliance events may be due to more diligent and conservative OOS assignment, which could lead to more false positives.

The analysis given in Figure 43 may be due to inappropriate phasing between compliance events and metric collection. To try to address this issue, compliance events in 2015 were compared with four quartiles of data for the three FDA Draft Guidance metrics evaluated.

Findings shown in Figure 44 plot the average number of 2015 compliance events against four 2014 data quartiles for the three FDA Draft Guidance metrics evaluated.

Figure 44: Compliance events in 2015 compared with 2014 FDA Draft Guidance metrics



[1] Warning letters, Form 483 notices, import bans, and Cass I recalls related to GMP compliance.
Source: FDA website, FDAzilla, press search

Lot Acceptance Rate and Product Quality Complaint Rate data do not show a clear trend, with both good and poor performers showing fewer compliance events than medium performers. This may be related to:

- Sample issues
- Lack of differentiation in compliance-event severity based on publicly available information
- Product Quality Complaint Rate value "noise"
- Lack of differentiation in Lot Acceptance Rate data

The Invalidated OOS Rate does seem to follow an expected trend, with a better Invalidated OOS Rate associated with lower number of compliance events. Without knowing the details of the Form 483 observations, however, it is not possible to indicate if this trend is associated with better laboratory performance.

It is also not possible to determine relationships or absence of relationships from data in Figure 43 and Figure 44; this could be due to data variability across the sample.

An attempt was made to examine data from two case study sites with apparently poor relative compliance status in the sample. Findings are given in Figure 45.

Figure 45: Comparing compliance status of two sites with FDA Draft Guidance metric performance

		Performance on FRN metrics		
		Total complaints	Lot Acceptance Rate	Invalidated OOS Rate
Influence of past compliance events	OTC site with 2–3 years' history of serious compliance issues	Median	Top quartile	Top quartile
Predictive of future compliance events	Rx site with no past events, but serious compliance issues occurring in late 2015	Median	Bottom quartile	Bottom quartile

Source: FDA website, FDAzilla, press search

While “serious compliance issue” is hard to define, it is judged that compliance findings at these sites had major implications. The first case study compared compliance status of an OTC site with 2–3 years’ history of serious compliance issues to FDA Draft Guidance metrics data from 2015. The site was in the top quartile for Lot Acceptance Rate and Invalidated OOS Rate, and in the median quartile for Product Quality Complaint Rate. No link to compliance status is apparent, leading to the hypothesis that historical presence of a compliance event may improve compliance status as measured by FDA Draft Guidance metrics.

In the second case study, an Rx site with no past serious compliance events had an issue in late 2015. FDA Draft Guidance Lot Acceptance Rate and Invalidated OOS Rate metric data were in the bottom quartile, and in the median quartile for Product Quality Complaint Rate. Poor results for Lot Acceptance Rate and Invalidated OOS Rate could have been connected to the impending 2015 compliance event; much more understanding is required.

In summary, a post–Wave 2 review comparing 2015-only performance on the three FDA Draft Guidance metrics evaluated for all Wave 2 sites with different levels of compliance issues in 2013–2014 shows no clear trends. It is possible that the compliance issues themselves may affect performance.

Attempting to link 2014 performance of the three FDA Draft Guidance metrics to future results showed no clear trends. One Rx site with compliance events in late 2015 may have had predictive metric data; more information is required to understand this observation, however.

Invalidated OOS Rate is the only metric that may have a link to compliance outcomes, but more detail is required to determine if this is related to laboratory performance or to some other factors (e.g., site culture).

In summary, a post–Wave 2 data analysis attempting to compare compliance status with evaluated FDA Draft Guidance metrics was inconclusive.

5.3.6 Conclusions

FDA metrics evaluated using FDA-proposed definitions were not found to have relationships with quality outcomes or with culture indicators.

Alternative metric calculations proposed by ISPE showed better relationships. All three FDA Draft Guidance metrics with ISPE-recommended definitions evaluated in Wave 2 showed a relationship with a culture indicator.

Findings from Wave 2 suggest that definitions should be adjusted for the three FDA Draft Guidance metrics evaluated:

- **Lot Acceptance Rate** should have a different denominator. Using lots dispositioned, the data point and definition used in Wave 2, produced more appropriate relationships. The burden to report lots attempted data was high. Subsequent consideration has suggested that lots released or rejected should be used as an alternative to "lots dispositioned."
- **Product Quality Complaint Rate** should use number of packs as a denominator, since this calculation produced more appropriate relationships and reduces variability due to lot size (e.g., from packing).
- **Invalidated OOS Rate** should be normalized by number of tests or number of lots tested. There was no difference which denominator to use and relationships were found for both. The rationale for the FDA-proposed double normalization is hard to understand, and this calculation does not produce relationships.

The lots pending disposition data point has limited usefulness. Lots pending disposition values have a relationship with Invalidated OOS, which is hard to understand and rationalize. It is not related to any outcomes and creates a high burden. It also showed high variability as a metric, with extreme outliers, especially in drug substance sites.

Wave 2 findings related to FDA-proposed Optional Metrics confirmed points made in ISPE's response to the FDA Draft Guidance: Optional Metrics have limited utility and were inconsistently applied. ISPE recommended that Optional Metrics should be deferred and considered once the program is established.

A post–Wave 2 data analysis comparing compliance status with evaluated FDA Draft Guidance metrics was inconclusive.

5.4 Culture findings and relationships

An objective of the Wave 2 Pilot study was to explore the effect of quality culture on quality performance in greater depth, and to examine any links to quality metric data. Refer to Figure 3 for a list of culture indicators, and to Figure 18 to see how they are relate to internal and external quality outcomes. In addition to determination of culture indicators, personnel at participating sites also completed a quality culture survey.

Figure 19 provides an overview of findings from Wave 2 showing relationships of culture indicators to both internal and external quality outcomes. Some of these relationships have been discussed already; for the sake of completeness, however, a full review of relationships is given in this section.

5.4.1 Quality culture survey

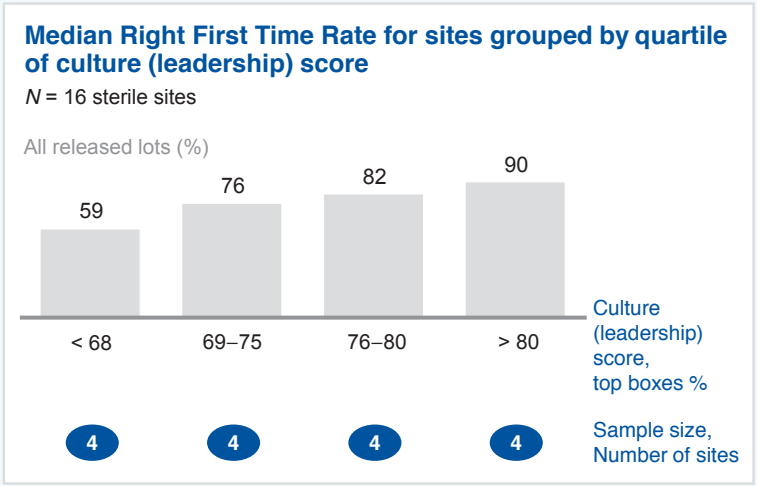
Scores from the quality culture survey gave statistically significant relationships to:

- Right First Time Rate
- Deviations Recurrence Rate
- Recalls

5.4.1.1 Relationship of quality culture scores to Right First Time Rate

Findings for the relationship of quality culture scores to the internal quality outcome Right First Time Rate are given in Figure 46.

Figure 46: Relationship of quality culture scores to Right First Time Rate



Note: Outliers more than two standard deviations away from sample mean were excluded: one on culture scores and one on RFT. The two excluded outliers are US-based Rx sterile sites with no known specifics to explain the reason for the extreme values.

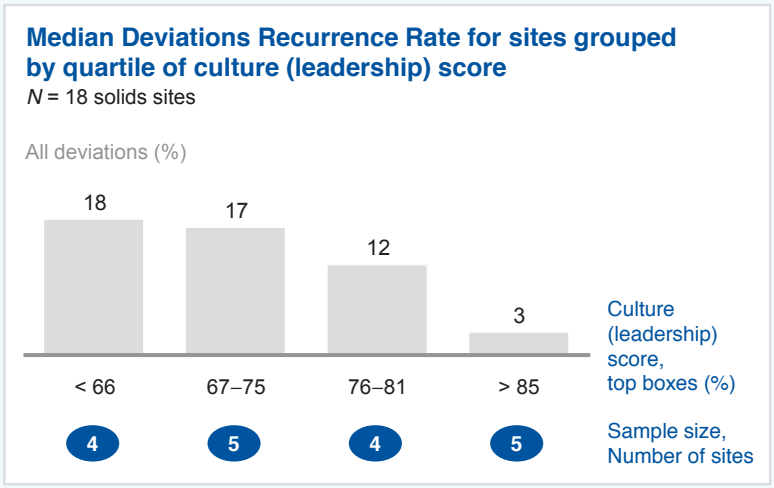
For these 16 sterile manufacturing sites higher culture leadership scores (related to coaching, daily dialogue, management presence on the shop floor) were linked to higher share of lots released without deviations (Right First Time Rate).

While RFT could be influenced by multiple factors, including product and process capability, it is logical that a site focus and attention to shop floor performance is likely to exert a positive influence on both the survey score and the RFT outcomes.

5.4.1.2 Relationship of quality culture scores to deviation recurrence

Findings for the relationship of quality culture scores to the Deviations Recurrence Rate internal quality outcome are given in Figure 47.

Figure 47: Relationship of quality culture scores to Deviations Recurrence Rate



Note: Outliers more than two standard deviations away from sample mean were excluded: one on culture scores and one on RFT.
The two excluded outliers are US- and Europe-based solids sites with no known specifics to explain the reason for the extreme values.

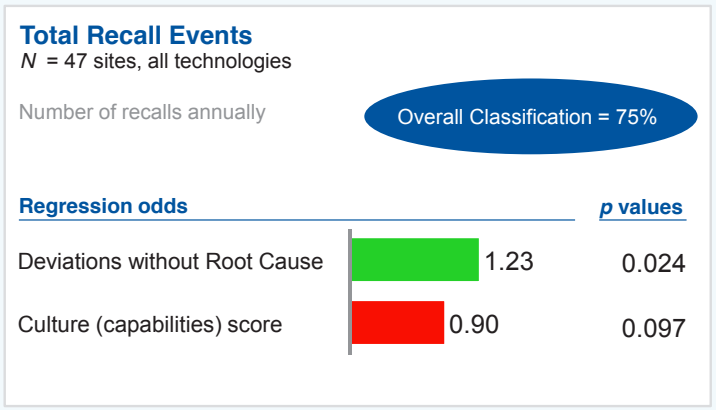
For 18 solid manufacturing sites, higher culture leadership scores (related to coaching, daily dialogue, management presence on the shop floor) were linked to lower Deviations Recurrence Rates.

While the Deviations Recurrence Rate is influenced by multiple factors, including systems tracking capability, it is logical that more coaching and attention to shop floor execution will influence both the survey score and (especially) recurrence of deviations related to human error positively.

5.4.1.3 Relationship of quality culture scores to recall events

Findings for the relationship of quality culture scores to the external quality outcome number of recalls is given in Figure 48.

Figure 48: Relationship of quality culture scores to recall events



Regression odds: the ratio of the probability of having the event vs. not having the event. Range is from 0 to infinity.
Classification: Tells us how well our model correctly classifies cases (predicts outcomes) by comparing observed vs. predicted cases in category 1 and observed vs. predicted cases of category 2 to get a percentage of correct cases per category, then averaging both rates to get the overall classification rate.

Higher culture capabilities scores (related to training and problem solving skills) and a lower Deviations without Assigned Root Cause Rate are related to the likelihood of recalls in a complex relationship. Both these factors move in understandable directions: An increase in Deviations without Root Cause is associated with higher recalls, and a higher quality culture capability score is associated with lower recalls.

5.4.2 Culture indicators

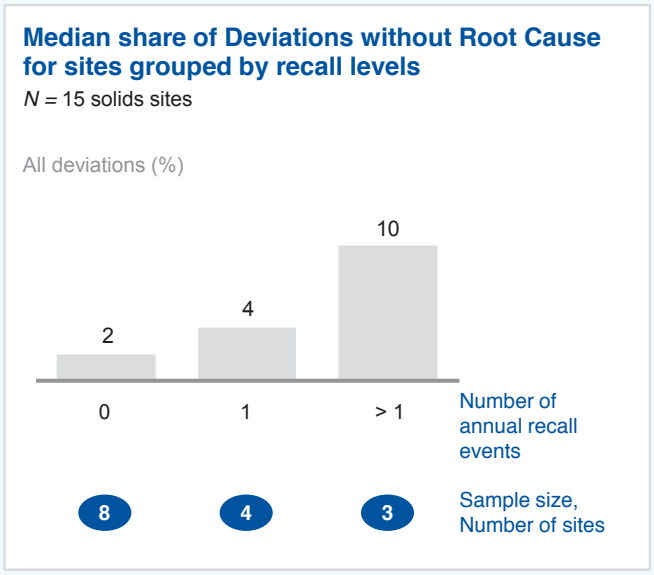
In Table A, relationships between culture indicators (other than quality survey scores) and internal and external quality outcomes are summarized, with links to both data and discussion.

Table A: Culture indicator and quality outcome relationships

CULTURE INDICATOR	OUTCOME		REFERENCE TO DATA AND DISCUSSION
	Internal/External	Metric	
Planned Maintenance Rate	External	Total Complaints (per pack)	Section 5.3.5.2, Figure 36
CAPA with Preventive Actions	External	Total Complaints (per pack)	Section 5.3.5.2, Figure 37
Employee Turnover Rate	Internal	Lot Acceptance Rate (per dispositions)	Section 5.3.5.2, Figure 35
CAPA with Retraining	Internal	Lot Acceptance Rate (per dispositions)	Section 5.3.5.1, Figure 26
Deviations without Assigned Root Cause	Internal	Invalidated OOS rate per test or lot	Section 5.3.5.2, Figure 38
	Internal	Lots Pending disposition data point	Section 5.3.5.1, Figure 24
	External	Recall Events	Below

Figure 49 shows the relationship between the Deviations without Assigned Root Cause culture indicator and Recall Events external quality outcome.

Figure 49: Relationship of deviations without assigned root cause and external quality outcome recall events



Note: Outliers more than two standard deviations away from sample mean were excluded: two on Deviations without Root Cause (LATAM-based small-scale sites with < 10 deviations annually), and two on recalls (US- and India-based Rx sites).

In Figure 49, median share of Deviations without Assigned Root Cause is plotted against number of annual Recall Events, subdivided into 0, 1, and > 1 event categories. A higher share of Deviations without Assigned Root Cause is linked to a higher level of recall events.

Recalls are influenced by multiple factors, including product and process capability, operational excellence, and quality-systems maturity. A high share of unidentified root causes, however, suggests problem-solving concerns and limited issue-resolution abilities on-site. It is a hypothesis that these relatively poor abilities could influence the likelihood of recalls.

In conclusion, every culture indicator except human error deviations has a relationship to either an internal or external quality outcome.

5.4.3 Conclusions

Wave 2 confirmed findings from Wave 1 that quality culture is important.

Quality culture scores affect two internal quality outcomes (Right First Time and Deviations Recurrence Rate) and one important external quality outcome (Recall Events).

Every culture indicator except human error deviations has a relationship to either an internal or external quality outcome.

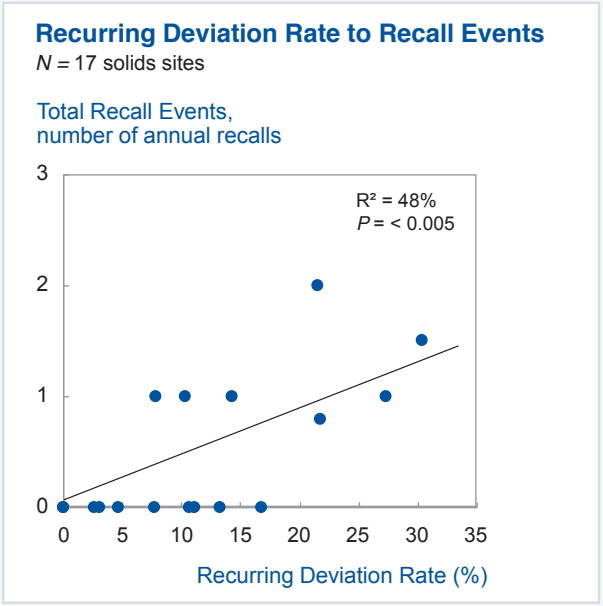
5.5 Internal quality outcomes

Another Wave 2 objective was to explore any internal quality outcomes that could help predict external quality outcomes.

Examination of these relationships, shown in Figure 19, indicates that Deviations Recurrence Rate is related to the external quality outcomes of Recall Events and Critical Complaints Rate per lot.

Figure 50 shows the relationship of Deviations Rate to Recall Events.

Figure 50: Relationship of Deviations Recurrence Rate to Recall Events



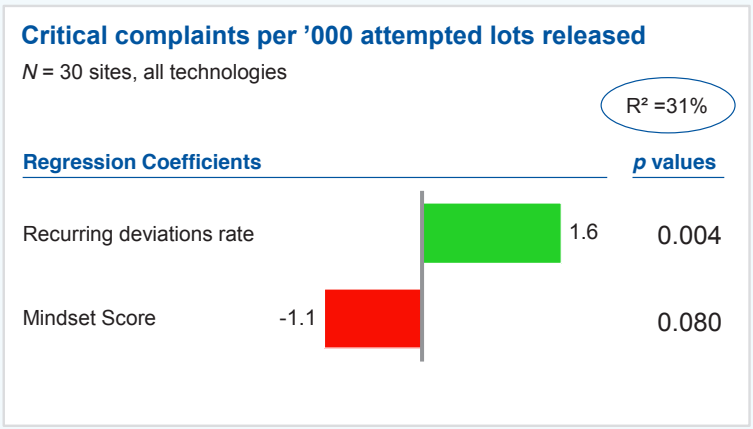
Note: Outliers more than two standard deviations away from sample mean were excluded: two on Deviations without Root Cause (LATAM-based, small-scale sites with < 10 deviations annually), and one US-based Rx site on recurrence.
R^2 measures to what extent metric Y (dependent variable) is explained by the variability of metric X (independent variable)
p value is probability that correlation between X and Y is zero, value below 0.05 indicates statistically significant results.

Deviations Recurrence Rate has a relationship with Recall Events. The scatter plot suggests that an increased Deviations Recurrence Rate could be associated with a higher number of Recall Events.

Recall events are influenced by multiple factors, including product and process capability, operational excellence, and quality systems maturity. A high Deviations Recurrence Rate indicates that there may be lower problem-solving ability and limited issue-resolution abilities on-site. These factors could influence the likelihood of recalls.

In Figure 51 an interdependent relationship is shown between Recurring Deviations Rate and the quality culture survey mindset score and critical complaints per lot released.

Figure 51: Relationship for Recurring Deviations rate and Cultural Survey score for Mindset on Critical Complaints per lot released



Note: Outliers more than two standard deviations away from sample mean were excluded.
R^2 measures to what extent metric Y (dependent variable) is explained by the variability of given set of independent metrics
p value is probability that correlation between variables is zero, value below 0.05 indicates statistically significant results.

A lower Deviations Recurrence Rate and higher mindset score are linked to lower Critical Complaints Rate with a denominator of lots released.

5.5.1 Conclusion

Deviations Recurrence Rate has relationships to two external quality outcomes: Critical Complaints Rate per lot and Recall Events. Additionally, it has a relationship to the FDA Draft Guidance metric Lot Acceptance Rate by lots attempted. These findings parallel those found in Wave 1 and make a good case for Deviations Recurrence Rate as a leading metric. Feedback from Wave 1 participants (Figure 9 in Wave 1 Report [1]) and from the ISPE Definitions team indicates that Deviations Recurrence Rate is hard to define clearly.

5.6 Drug shortages

Responses to drug shortages questions are extremely hard to interpret, possibly due to inaccurate question drafting and interpretation. One participant, for example, reported “no” overall and “yes” on an individual product sheet. That participant may have misinterpreted how a drug shortage is defined—perhaps not an event reportable to FDA, but a disturbance in supply.

There were two reports of a drug shortage being associated with a participating site. Neither site reported that any Wave 2 metric was useful in predicting the shortage; one site reported that the shortage was supplier related.

In summary, Wave 2 did not produce data that could relate metrics evaluated to drug shortages.

5.7 Information technology systems

To help participants implement a quality metrics program, the Wave 2 Pilot collected information about site IT systems and what proportion of processing effort was manual.

Figure 52 presents IT systems used to collect Wave 2 data points. Blue squares indicate systems used by more than a third of the sites to collect that data point.

Figure 52: IT systems

IT systems used by responding sites[1] (%)

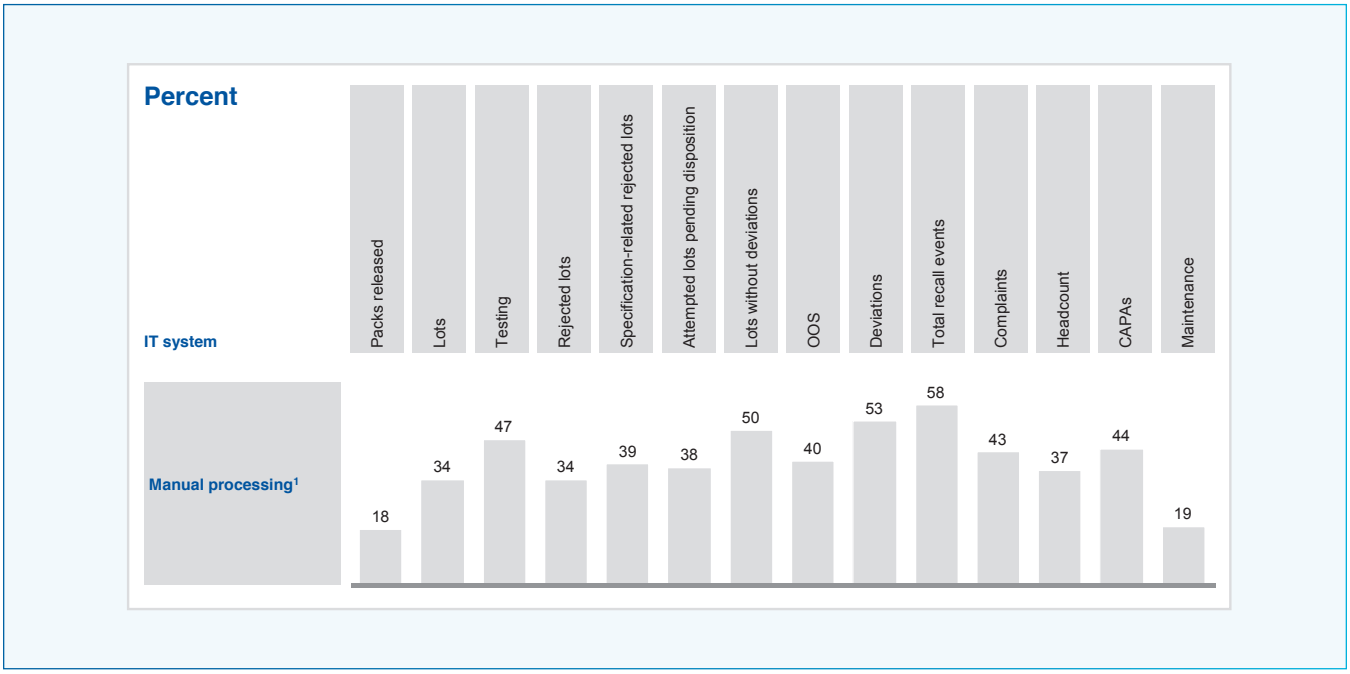
Used by more than one-third of the sites

IT system	Sites having each system (%)	Packs	Lots	Testing	Total rejects	Specification-related reject	Lots pending disposition	Lots without deviations	OOS	Deviations	Total recalls	Complaints	Headcount	CAPAs	Maintenance
ERP[2]	94	77	93	47	86	59	75	57	7	3	5	7	34	0	55
LIMS[3]	55	2	11	52	7	21	9	4	23	0	0	0	0	0	0
TrackWise	74	0	2	2	4	30	2	48	64	69	26	52	3	69	8
Excel	65	11	23	32	21	29	18	30	25	31	23	30	26	19	11
Manual	85	16	27	29	23	21	30	36	29	39	40	21	18	32	15
Other	69	5	11	13	9	14	11	16	13	18	16	23	50	19	37

1 There are overlaps in systems used to track each of enlisted data points
2 Enterprise Resource Planning: software suite used to collect, store manage and interpret data from business activities
3 Laboratory Information Management System: software used to support laboratory's operations

Enterprise Resource Planning (ERP) systems and Trackwise are most widely used systems. Despite availability of multiple IT systems, a significant number of data points across the sample were collected manually. Figure 53 explores manual collection efforts further.

Figure 53: Manual data point collection



[1] Excel or other manual solution, including supplementary role to other systems

Despite high levels of automation, between one-third and one-half of sites surveyed use manual processing as main or supplementary means to process Wave 2 data.

5.7.1 Conclusion

Most participants used existing systems to source some or most of the data points. Between 75% and 95% of participants used ERP and Trackwise systems to manage data.

Despite using automated systems, one-third to one-half of the sites used manual processing as a primary or supplementary method to process Wave 2 data. This included data resorting and adjustments to ensure conformance to definitions.

6 Output and Lessons learned from ISPE Pilot Program Wave 2

Wave 2 achieved essentially all its main objectives, more discussion being given later in the section. For example:

- Samples sizes increased across all technologies giving good representation in all technologies
- Some Wave 1 relationships were confirmed
- FDA Draft Guidance metrics were evaluated
- Alternative definitions applied to FDA Draft Guidance metrics were also evaluated
- More understanding of the use of Process Capability/Performance measures was obtained
- Quality culture was explored more deeply

The sample size was insufficient to allow meaningful trend analysis of metric data with time.

Companies reported great value in receiving their data and participating in the confidential benchmarking exercise.

6.1 Findings relevant to FDA Draft Guidance

Findings relating to FDA Draft Guidance were:

- The three FDA Draft Guidance metrics evaluated, Lot Acceptance Rate, Product Quality Complaint Rate and Invalidated OOS Rate did not have relationships to External Quality Outcomes and did not have direct relationships to any Culture Indicator or Quality Survey scores
- Definitions should be re-considered for these three evaluated FDA Draft Guidance metrics.
- Optional metrics have limited utility and were inconsistently applied
- The data point of Attempted Lots Pending Disposition for 30 Days has limited usefulness and has a relatively high burden to collect
- Estimates of effort to collect data as given in the FDA Draft Guidance have been made and are approximately 3 times higher than estimates given in the FRN. This is considered an underestimate, especially for OTC companies and companies with complex supply chains.

6.2 ISPE Recommendations relevant to FDA Draft Guidance

All three evaluated FDA Draft Guidance metrics calculated using ISPE-recommended definitions in Wave 2 have a relationship with a Culture Indicator. These findings suggest that definitions should be adjusted for the three evaluated metrics in the FDA Draft Guidance:

- Lot Acceptance Rate should have a different denominator. 'Lots dispositioned' is the data point and definition used in Wave 2 and using this in an alternate calculation produced more appropriate relationships. The burden to report 'lots attempted' data was high. Subsequent consideration has suggested that 'lots released or rejected' should be used as an alternative to 'lots dispositioned'.
- Product Quality Complaint Rate should have a denominator of number of packs since this calculation produced more appropriate relationships and reduces variability due to lot size variation (e.g. from packing)
- Invalidated OOS should be normalized by number of tests or number of lots tested. There was no difference which denominator to use and relationships were found for both. The rationale for the FDA-proposed double normalization is hard to understand and this calculation does not produce relationships.

Using the ISPE-recommended calculations for the three FDA Draft Guidance metrics evaluated is estimated by ISPE as 1/3 the effort of collecting data according to FDA calculations.

6.3 Other findings

Findings from Wave 1 were largely confirmed, for example:

- McKinsey continued to provide substantial effort
  - Checking submitted data for consistency
  - Providing assistance to new sites when commencing Wave 2
  - Answering queries from participants
- Many Wave 1 statistically significant relationships were confirmed:
  - Deviations Recurrence Rate to Recall events
  - Quality Culture scores to Deviations Recurrence Rate
- Some Wave 1 statistically significant relationships were not found, probably due to wider variability in the sample (an objective) for example;
  - Lot Acceptance Rate to Critical Complaints
- Wave 1 finding that quality culture was important was confirmed. Some Culture Indicators new to Wave 2 such as Deviations without Assigned Root Cause Rate had relationships to Internal Quality Outcomes (Invalidated OOS Rate per lot or test) and an External Quality Outcome (Recall events). Other Culture Indicators had statistically significant relationships to External Quality Outcomes (Planned Maintenance rate and CAPA with Preventive Actions) and others to Internal Quality Outcomes (Employee Turnover Rate). Relationships do not imply causation and many further studies are required to confirm or deny these relationships attempt to determine underlying causes.
- Determination of quality culture using simple-to-collect metrics, for example the three proposed FDA metrics, is confirmed as extremely difficult and may be not possible.
- Process Capability/Performance measures are extensively used by companies to assist in controlling processes and identifying continual improvement opportunities. Use of these indices varies between companies.
- An internal metric, Deviations Recurrence Rate was identified which could be used by companies to assist predicting External Quality Outcomes.
- A post data analysis attempt to compare compliance status with evaluated FDA Draft Guidance metrics was inconclusive.

7 Proposals

Preliminary findings from Wave 2 were included in ISPE's response to the FDA Draft Guidance and FRN. Full review has not altered relevant high-level messages:

ISPE supports FDA's effort to implement a Quality Metrics program

ISPE continues to support introduction of a quality metrics program to help FDA develop risk-based scheduling of drug manufacturers in the near term, and potentially to help reduce post-approval manufacturing change categories in the longer term.

Wave 2 findings did indicate that the three FDA Draft Guidance metrics were valuable if alternative definitions were applied. Findings also indicated that much further learning is required to implement an effective and efficient program. Some suggestions and examples are given below.

Start with a small, targeted approach

- Clear and detailed definitions are important
  - FDA definitions required multiple clarifications and support for lots attempted, lots pending disposition, and specifications-related rejected lots
- ISPE-recommended definitions should be tested more widely
- Data from different reporting periods should be checked for consistency; most repeat participants needed to revise submissions of repeated data points
- Understand data characteristics and data analysis, for example:
  - Outlier identification and treatment
  - Statistical analysis method
- Understand how metrics are used with other information to assist with risk-based inspection scheduling. ISPE suggests estimates could be made on a carefully designed program with a selected sample (equivalent to a development study experiment) rather than involving the whole industry
- Minimize the burden

Phased introduction by segment

- Minimizes the burden
  - Higher burden for OTC sites: 60% more time to collect data vs. Rx/Gx
  - Significant outliers in some metrics like Total Complaints (Product Quality Complaint Rate), which has high variability, especially for sterile products
  - Coordination, review, and confidentiality challenges related to reporting CMO data by product
- This is an option for "starting small'

Start with the three FDA proposed metrics

- The three FDA Draft Guidance metrics evaluated have value, but they should be adjusted in calculation/normalization approach—Invalidated OOS Rate, Product Quality Complaint Rate, and Lot Acceptance Rate.
- Wave 1 findings show that APR/PQR on-time rate has limited value
- Complementary data points such as lots pending disposition and Optional Metrics such as CAPA Requiring Retraining and process capability/performance indices have limited usefulness as potential quality performance metrics.

Effort is underestimated

- Estimated effort at least 3 times higher than FDA projections is probably still an underestimate. In the Pilot, the measured effort covered simple supply chains, had self-selection bias, unofficial submission context, and only some data points requested in the FDA Draft Guidance.
- Industry effort using the FDA-proposed method of calculation is approximately three times higher than the industry effort to calculate the same metrics using ISPE-recommended method of calculation of Lot Acceptance Rate by site differentiated by product, Product Quality Complaint Rate by product and Invalidated OOS Rate by site

8 Conclusions

ISPE Pilot Program Wave 2 met its objectives and confirmed findings from Wave 1. Main findings were:

- Effort to collect FDA Draft Guidance metrics was approximately three times that given in the FRN. This value is considered an underestimate, especially for OTC companies and companies with complex supply chains.
- ISPE-recommended calculations for the three FDA Draft Guidance metrics evaluated are estimated to require one-third the effort of collecting data according to FDA calculations.
- When calculated using FDA definitions, the three FDA Draft Guidance metrics evaluated did not exhibit relationships with either external quality outcomes or culture indicators. Lot Acceptance Rate (using the FDA-recommended lots attempted as the denominator) did show relationships with two other internal quality outcomes: Invalidated OOS per test or lot and Deviations Recurrence.
- When calculated using alternative calculations as defined in the Wave 2 Pilot, the same three FDA metrics did show relationships with culture indicators.
- Consistent with the ISPE response to the FDA Draft Guidance, the Wave 2 Pilot study confirmed that alternative definitions of the three FDA Draft Guidance diagnostic metrics evaluated should be considered.
- The importance of quality culture was confirmed with some further relationships identified. Determining quality culture using simple-to-collect metrics, such as the three proposed FDA metrics, for example, is confirmed as extremely difficult and may be not possible.
- Process capability/performance measures are used extensively by companies to help control processes and identify continual improvement opportunities. Use of these indices varies between companies.
- Deviations Recurrence Rate, an internal metric, was identified as one that companies could use to help predict external quality outcomes.
- Preliminary findings from Wave 2 were used to develop ISPE's response to the FDA Draft Guidance and FRN. Final analysis confirmed those points.
- Participating companies reported that they received great value from the metrics data they received and in participating in the confidential benchmarking exercise.

9 References

1. International Society for Pharmaceutical Engineering. “ISPE Quality Metrics Initiative: Wave 1 Report.” June 2015. www.ISPE.org.
2. US Food and Drug Administration. Federal Register Notice. “Request for Quality Metrics; Notice of Draft Guidance Availability and Public Meeting; Request for Comments.” Docket No. FDA-2014-D-2537. 28 July 2015. http://federalregister.gov/a/2015-18448.
3. US Food and Drug Administration. “Guidance for Industry: Request for Quality Metrics.” Draft guidance. July 2015. www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm455957.pdf.
4. International Society for Pharmaceutical Engineering. “ISPE Response to FDA Federal Register Notice on Quality Metrics [Docket No. FDA-2014-D-2537] and Draft Guidance, Request for Quality Metrics, 27 July 2015.” 24 November 2015. www.ispe.org/ispe%20comments%20fda%20draft%20request%20for%20quality%20metrics.pdf.
5. US Food and Drug Administration. “2015 Annual Report on Inspections of Establishments in FY 2014.” www.fda.gov/downloads/RegulatoryInformation/Legislation/SignificantAmendmentstotheFDCAct/FDASIA/UCM432287.pdf.
6. Bizjak, Tara Gooen. “FDA Quality Metric Program Update: Policy and Implementation.” PDA Pharmaceutical Quality Metrics Conference, held 9 November 2015, Bethesda, Maryland.
7. US Government Publishing Office. Electronic Code of Federal Regulations. Title 21: “Food and Drugs.” www.ecfr.gov/cgi-bin/text-idx?SID=3ee286332416f26a91d9e6d786a604ab&mc=true&tpl=/ecfrbrowse/Title21/21tab_02.tpl.
8. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonized Tripartite Guideline Q7: “Good Manufacturing Practice Guide for Active Pharmaceutical Ingredients.” 10 November 2000. www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q7/Step4/Q7_Guideline.pdf.

Appendix 1

Definitions

External Quality Outcomes

METRIC	DEFINITION
Total Complaints Rate per million packs excluding lack of effect = Total complaints excluding lack of effect divided by total number of packs divided by 10^6 for a site	▶ Total complaints excluding lack of effect = All complaints received by the site in the reporting period, related to the quality of products manufactured in the site (i.e., complaints involving any possible, including actual, failure of a drug product to meet any of its specifications designed to ensure that any drug products conform to appropriate standards of identity, strength, quality, and purity), regardless whether subsequently confirmed or not. All complaints received by the site should be counted, even if a complaint affects more than one site, or if eventually the root cause analysis attributes the issue to another site. Please exclude lack of effect complaints. Drug substance sites should also report end customer complaints that were received for investigation by their site (i.e., the complaints related to API or DS issues). ▶ Total number of packs = Total number of packs (final product form that leaves the plant, one level less than tertiary packs, most usually it is secondary packaging unit e.g. pack of blisters or bottle in carton pack) released in the period. *API and Drug substance did not complete this data point.*
Total Complaints Rate per million packs including lack of effect = Total complaints including lack of effect divided by total number of packs divided by 10^6 for a site	▶ Total Complaints including lack of effect = Total complaints as defined above, plus all complaints related to lack of effect ▶ Total number of packs is defined in the cell above.
Total Complaints Rate per thousand attempted lots released excluding lack of effect = Total complaints excluding lack of effect per product divided by number of attempted lots released divided by 1,000 for a site	▶ Total complaints excluding lack of effect is defined in the cell above. ▶ Total number of attempted lots = Total number of lots intended for commercial use produced and/or packaged on site that were initiated during the period (i.e., manufacturer has charged API (for finished drug sites) or primary starting materials (for API sites)) and assigned an individual number. When a lot is dispositioned after a production stage and assigned a new number for the next stage, count each new assigned number as a new attempted lot. For example, if one formulation lot is split into three packaged lots this should be counted as four attempted lots. If a lot was sent for rework/reprocessing and received a new number after the rework/reprocessing, it should be counted as two attempted lots. Excludes lots that are not produced or packaged on site, but just released for CMOs. ▶ Total number of attempted lots released = The number of lots attempted per above definition, which are released for distribution or for the next stage of manufacturing of the product.

Appendix 1

METRIC	DEFINITION
Total Complaints Rate per thousand attempted lots released including lack of effect = Total complaints including lack of effect per product divided by number of attempted lots released divided by 1,000 for site	▶ Total complaints including lack of effect = all complaints in total. Complaints are defined in the cell above. Complaints affecting more than one site should be counted only once. ▶ Total number of attempted lots released is defined in the cell above.
Critical Complaint Rate per million packs = Number of critical complaints divided by the total number of packs produced per site divided by 10^6	▶ Number of critical complaints = Complaints received by the site which may indicate a potential failure to meet product specifications, and may impact product safety and could lead to regulatory actions, up to and including product recalls. Critical complaints include those that potentially could lead to FDA notification (e.g., Field Alert Reports, Biological Product Deviation Reports). Critical (or expedited) complaints are identified upon intake, **whether subsequently confirmed or not**, based on the description provided by the complainant, and include, but may not be limited to: i. Information concerning any incident that causes the drug product or its labelling to be mistaken for, or applied to, another article. ii. Information concerning any bacteriological contamination, or any significant chemical, physical, or other change or deterioration in the distributed drug product, or any failure of one or more distributed batches of the drug product to meet the specification established for it in the application. ▶ Total number of packs produced is defined in the first cell above
Critical Complaints Rate per thousand attempted lots released = Number of critical complaints divided by the number of attempted lots released divided by 1,000	▶ Number of critical complaints is defined in the cell above. ▶ Total number of attempted lots released is defined in the third cell above.
Total Recall Events per year per site = Total number of recalls in the reporting period	▶ Total market recalls (including non-US ones if the site is reporting data for the full site).

Appendix 1

Internal Quality Outcomes

METRIC	DEFINITION
Lot Acceptance Rate per finally dispositioned lots = number of lots released divided by the number of lots dispositioned in the reporting period multiplied by 100	▶ Number of lots released = the number of finally dispositioned lots less the number of finally rejected lots. ▶ Lots finally dispositioned = Total number of lots intended for commercial use produced and/or packaged on site that were initiated during the period (i.e., manufacturer has charged API (for finished drug sites) or primary starting materials (for API sites)) and assigned an individual number. When a lot is dispositioned after a production stage and assigned a new number for the next stage, count each new assigned number as a new attempted lot. For example, if one formulation lot is split into three packaged lots this should be counted as four attempted lots. If a lot was sent for rework/reprocessing and received a new number after the rework/reprocessing, it should be counted as two attempted lots. Excludes lots that are not produced or packaged on site, but just released for CMOs. ▶ Finally rejected lots = Out of all lots that were finally dispositioned during the period, the total number of full lots that were rejected for quality reasons. Rejected means intended for destruction or experimental use, not for rework or commercial use. Rejections should be counted regardless at what production stage the rejection occurred. “Cancelation of a batch” at the early stages (i.e., before charging API (for finished drug product sites) or primary starting materials (for API sites)) is not a quality disposition and is not counted as a reject.
Lot Acceptance Rate per attempted lots = number of lots of product released divided by number of attempted lots multiplied by 100	▶ Number of lots released is defined in the third cell above. ▶ Number of lots attempted is defined above in the third cell.
Invalidated OOS Rate per thousand lots tested = Number of Invalidated OOS divided by number of lots tested divided by 1,000	▶ Number of Invalidated OOS = The total number of OOS Invalidated (identified as lab errors) by establishment or by a contractor, related to finished product lot release finished product and stability testing only. ▶ Number of lots tested = Total number of lots used for commercial production that are tested and dispositioned out of the lab in the period, i.e., have a QC pass or fail decision on them. Includes only: – Finished product lot release testing (counted as one lot tested, even if sampled separately for chemical and microbiological testing, or for in-process analytical testing in lab or on shop floor). Lot release testing includes all finished product tests, all real time release tests, and all in-process tests that act as a surrogate for finished product lot release. Finished product covers FDF (finished dosage form) and API. – Lots undergoing stability testing in that period (counted as one per each time point and condition sampled per the approved stability protocol).

Appendix 1

METRIC	DEFINITION
Invalidated OOS Rate per thousand tests performed = Number of Invalidated OOS divided by number of tests performed divided by 1,000	▶ Number of Invalidated OOS results is defined in the cell above. ▶ Number of tests performed = Total number of individual tests performed during the period against the specifications or acceptance criteria established in drug applications, drug master files (DMFs), official compendia, formulary or applied by the manufacturer. Includes finished product lot release testing (i.e., all finished product tests, all real time release tests, and all in-process tests that act as a surrogate for finished product lot release) and stability testing. If you run multiple tests of the same type on one lot but each individual test result is compared against the specification, each should be counted separately. If only their average is compared against the specification, they should be counted as one.
Invalidated OOS Rate per total OOS per number of tests performed = The number of Invalidated OOS divided by the total number of OOS test results divided by the total number of tests performed by the establishment in the same timeframe.	▶ Invalidated OOS test result is defined in the cell above. ▶ All OOS results = The number of OOS results for the product, including stability testing. Include (1) finished product lot release and stability test results only and, (2) all finished product and stability test results that initially appear as OOS, even if invalidated by a subsequent laboratory investigation. ▶ OOS results are all test results that fall outside the specifications or acceptance criteria established in drug applications, drug master file, official compendia, or by the manufacturer. ▶ Lot release testing includes all finished product tests, all real time release tests, and all in-process tests that act as a surrogate for finished product lot release. Finished product covers FDF (finished dosage form) and API. ▶ Number of tests performed is defined in the cell above
Right First Time Rate per released lots attempted = Number of lots without deviations divided by the number of attempted lots released	▶ The number of lots without deviations = Out of all attempted lots that were released at each stage during the period, how many did not have any deviations. A deviation is any major or minor unplanned occurrence, problem, or undesirable incident or event representing a departure from approved processes or procedures, also includes OOS in manufacturing or laboratory or both. Do not count Invalidated OOS as deviations for reporting of this metric since they have no impact on product or material., which are released during the period for distribution or for the next stage of manufacturing of the product. ▶ The number of released lots attempted is defined in the third cell

Appendix 1

METRIC	DEFINITION
Deviations Rate per thousand finally dispositioned lots = The total number of deviations divided by the total number of lots dispositioned divided by 1,000	▶ The total number of deviations = Any major or minor unplanned occurrence, problem, or undesirable incident or event representing a departure from approved processes or procedures, also includes OOS in manufacturing or laboratory or both. Please count only deviations that have been closed/resolved in the period. Deviations from one period, for which the investigation was closed in the next period, should be counted in the latter period. Do not count Invalidated OOS as deviations for reporting of this metric since they have no impact on product or material. ▶ Number of lots finally dispositioned = Total number of lots for commercial use produced and/or packaged on site that went through **final** disposition during the period, i.e., were released for shipping or rejected (for destruction). Rejections should be counted as final disposition regardless at what production stage the rejection occurred. Release is only final release for shipping (whether shipping bulk to another site or final product to market). Excludes lots that have been sent for rework or put on hold/quarantined in this period and hence are not finally dispositioned. Excludes lots that are not produced or packaged on site, but just released for CMOs.
Deviations Rate per thousand attempted lots = The total number of deviations divided by the total number of lots attempted divided by 1,000	▶ The total number of deviations is defined in the cell above. ▶ The total number of lots attempted is defined in the third cell above.
Recurring Deviations Rate = Total number of recurring or repeat deviations divided by the Total number of deviations multiplied by 100	▶ The total number of recurring deviations = Number of recurring or repeat deviations. A recurring/repeat deviation has occurred at least once more (same or very similar issue) during the preceding period (at minimum 12 months back, but could be longer depending on process involved) with the same root cause within the same or similar process and/or work area. – Include deviations regardless whether a CAPA has previously been opened for them. – Consider the true root cause, not just general categorization as “human/operator error” or “method error,” but at least a level deeper. – If similar processes or equipment exist, please consider deviation events common to the grouping/work center as recurring. For example, if a deviation for missing desiccant occurs twice, on two separate packaging lines with comparable equipment/systems, it should be counted as recurring if the same root cause is identified (i.e., as two “same” deviations, rather than one “different” for each line). Or you could have line clearance issue on two different packaging lines, and in both cases the operator completely neglected to clear the line prior to starting the next batch. ▶ Total number of deviations is defined in the cell above

Appendix 1

Culture Indicators

METRIC	DEFINITION
CAPAs with preventive actions = The number of CAPAs with preventive actions divided by the total number of CAPAs multiplied by 100	▶ Total number of CAPAs = All CAPAs opened during the period. CAPA is a corrective or preventive action. Corrective action is an action taken to eliminate the cause(s) of a detected nonconformity or other undesirable situation in order to prevent recurrence. Preventive action is an action taken to eliminate the cause(s) of a potential nonconformity or other undesirable potential situation to prevent an occurrence. You should count all CAPAs as tracked in your systems. CAPAs may be generated from the following activities (this list is not meant to be exhaustive, you may have more types and all should be counted). – Any investigations, deviations and/or nonconformities – Product complaints, and vendor/supplier complaints related investigation – Laboratory investigations – Action limit excursion – Investigations associated with equipment/instruments that are founded out of tolerance and are used in manufacturing or testing materials/products – Internal audits – Partner/vendor audits – Health authority inspections – Continual improvement projects ▶ Out of all CAPAs in the above bullet, how many included at least one preventive action—an action taken to eliminate the cause(s) of a potential nonconformity or other undesirable potential situation to prevent an occurrence (as defined above)?
Planned maintenance rate = The number of planned/routine maintenance orders divided by the total number of maintenance orders multiplied by 100	▶ Total maintenance orders = Total maintenance work orders opened during the period related to production and laboratory operations. ▶ Planned maintenance orders = Out of all orders in the above bullet how many were planned or routine, rather than non-routine or emergency orders?
Employee Turnover Rate = Number of employee separations/ departures divided by the average of beginning and end headcount multiplied by 100	▶ Number of employee separations/departures = Number of full time employees who left the site during the period (voluntary or non-voluntary). Excludes contract and temporary employees. ▶ Beginning site headcount = total employees in the beginning of the reporting period. Excludes contract and temporary employees ▶ End site headcount = total employees at the end of the reporting period. Excludes contract and temporary employees

Appendix 1

METRIC	DEFINITION
Human Error Deviations Rate = Number of deviations with human error root cause divided by the total number of closed deviations multiplied by 100	▶ Number of deviations with human error root cause = Out of all closed deviations, how many were assigned a root cause "human error" (as defined by the site) without more specific details on the root cause? ▶ Total number of closed deviations = Any major or minor unplanned occurrence, problem, or undesirable incident or event representing a departure from approved processes or procedures, also includes OOS in manufacturing or laboratory or both. Please count only deviations that have been closed/resolved in the period. Deviations from one period, for which the investigation was closed in the next period, should be counted in the latter period. Do not count Invalidated OOS as deviations for reporting of this metric since they have no impact on product or material.
Deviations with No Assigned Root Cause Rate = Number of Deviations with No Assigned Root Cause divided by the total number of closed deviations multiplied by 100	▶ Number of deviations with no assigned root cause = Out of all closed deviations, how many were not assigned a most probable or confirmed root cause (i.e., the investigation did not result in an identified root cause)? ▶ Total number of closed deviations is defined in the cell above.
CAPA Requiring Retraining Rate = Number of CAPAs requiring retraining divided by Total Number of CAPAs multiplied by 100	▶ Total number of CAPAs is defined above in the first cell of this section. ▶ Out of all CAPAs from above, how many involved retraining of personnel (i.e., addressing a root cause of lack of adequate training)

Appendix 2

Survey Questions

Quality culture questions

		STRONGLY AGREE	AGREE	DISAGREE	STRONGLY DISAGREE	I CAN'T ANSWER THIS QUESTION
Capabilities	**Patient focus:** I know which parameters of our products are particularly important for patients	☐	☐	☐	☐	☐
	Training: The training I have received clearly helps me to ensure quality in the end product	☐	☐	☐	☐	☐
	Problem Solving: All line workers are regularly involved in problem solving, troubleshooting and investigations	☐	☐	☐	☐	☐
Governance	**Recognition:** We recognize and celebrate both individual and group achievements in quality	☐	☐	☐	☐	☐
	Metrics: Up-to-date quality metrics (e.g. defects, rejects, complaints) are posted and easily visible near each production line	☐	☐	☐	☐	☐
	Knowledge: Each line worker can explain what line quality information is tracked and why	☐	☐	☐	☐	☐
	Continual Improvement: We are regularly tracking variations in process parameters and using them to improve the processes	☐	☐	☐	☐	☐
Leadership	**Coaching:** Supervisors provide regular and sufficient support and coaching to line workers to help them improve quality	☐	☐	☐	☐	☐
	Dialogue: We have daily quality metrics reviews and quality issues discussions on the shop floor	☐	☐	☐	☐	☐
	Gemba: Management is on the floor several times a day both for planned meetings and also to observe and contribute to the daily activities	☐	☐	☐	☐	☐
Mindsets	**Awareness:** Every line worker is aware of the biggest quality issues on their line and what is being done about them	☐	☐	☐	☐	☐
	Responsibility: All employees see quality and compliance as their personal responsibility	☐	☐	☐	☐	☐
Integrity	**Openness:** I am not afraid to bring quality issues to the management's attention	☐	☐	☐	☐	☐
	Ethics: People I work with do not exploit to their advantage inconsistencies or 'grey areas' in procedures	☐	☐	☐	☐	☐
	Motivation: All employees care about doing a good job and go the extra mile to ensure quality	☐	☐	☐	☐	☐

Appendix 2

Process capability

Capabilities—Section 1

QUESTION	RESPONSE	COMMENTS
1. Do you have a threshold established for acceptable process capability values?	[Y/N]	
2. Is the threshold the same for all products?	[Y/N]	
3. What are these threshold value(s)?		
– For Cpk	[#]	
– For Ppk	[#]	
4. If you have a threshold, what do you do when an attribute or a parameter falls below the acceptable threshold?		
– Open a CAPA	[Y/N]	
– Initiate process improvement	[Y/N]	
– Track in a process monitoring report	[Y/N]	
– Adjust inventory to compensate for possible out of stocks due to process failures	[Y/N]	
– Initiate a nonconformance record in a GMP quality system	[Y/N]	
– Other, please specify	[Y/N]	
– Please specify details if you selected “Other”	[Y/N]	
5. Do you include the calculated results in the product APR?	[Y/N]	
6. For new products (introduced within the last year), do you require a certain process capability to be demonstrated for new developmental products prior to transferring to manufacturing?	[Y/N]	
7. What is the minimum number of batches needed for a new product before first calculating Cpk or Ppk?	[#]	

Appendix 2

QUESTION	RESPONSE	COMMENTS
8. Do you calculate process capability for legacy products (introduced more than a year earlier)? If yes, answer the questions below.	[Y/N]	
– What is the minimum number of batches (in manufacturing volume for the relevant period) needed to calculate Cpk or Ppk for a product?	[#]	
– How frequently is the calculation updated?	<Text>	
– Are periodically calculated Cpk/Ppk values compared (to the previous value) or trended (e.g. run charts)?	<Text>	
9. Are there quantitative quality attributes for which Cpk/Ppk is not calculated even when sufficient number of batches is manufactured? Please give reasons/examples (e.g., limit of quantitation, risk-based decision, etc.)	<Text>	
10. What are your concerns regarding reporting process capability values as 'official' to FDA?		
– Opportunity for misinterpretation—criteria/ threshold for an incapable process may differ between companies, the scale of Cpk/Ppk values doesn't convey risks of specification failure in a linear fashion	[Y/N]	
– Difficult to calculate for some attributes. These tools can be applied to assay and uniformity measurements, however, there are challenges applying to dissolution, impurities and degradation products, and it is not possible to apply to subjective qualitative tests e.g. appearance	[Y/N]	
– Difficult to calculate for some products, which are not manufactured in sufficient volume to accurately estimate the process true mean and standard deviation	[Y/N]	
– Process capability as a compliance metric has the potential to decrease supply to the market	[Y/N]	
– Other	[Y/N]	
– Please specify if you selected "Other"	<Text>	

Appendix 2

Capabilities—Section 2

Please complete the section below if you did not participate in the Wave 1 Pilot. Please note that this section covers broad state of control measurements, not Cpk/Ppk.

If you participated in Wave 1 and your situation has changed (e.g., you started recently measuring state of control) or you didn't complete it at the time, you may complete this section now.

QUESTION	RESPONSE						
Do you measure that the process remains in a state of control (the validated state) during commercial manufacturing?	[Y/N]						
If yes, please answer questions below:	Please indicate which metric(s) you use for ongoing monitoring and the parameters to which you apply them						
	Cpk	Ppk	Tolerance interval	Box plots	Trending	Other	Comments
Applied to CQA's (critical quality attributes tested in the lab)	☐	☐	☐	☐	☐	☐	
Applied to IPC (in-process control) checks	☐	☐	☐	☐	☐	☐	
Applied to CPP's (critical process parameters)	☐	☐	☐	☐	☐	☐	
For what % of products are they applied (based on your total number of products as reported in "Data by site")—excluding packaging operations?	[%]						
If not applied on 100% of products, how do you choose/ segregate/prioritize on which products to apply these metrics?	<Text>						

Appendix 2

Drug shortages

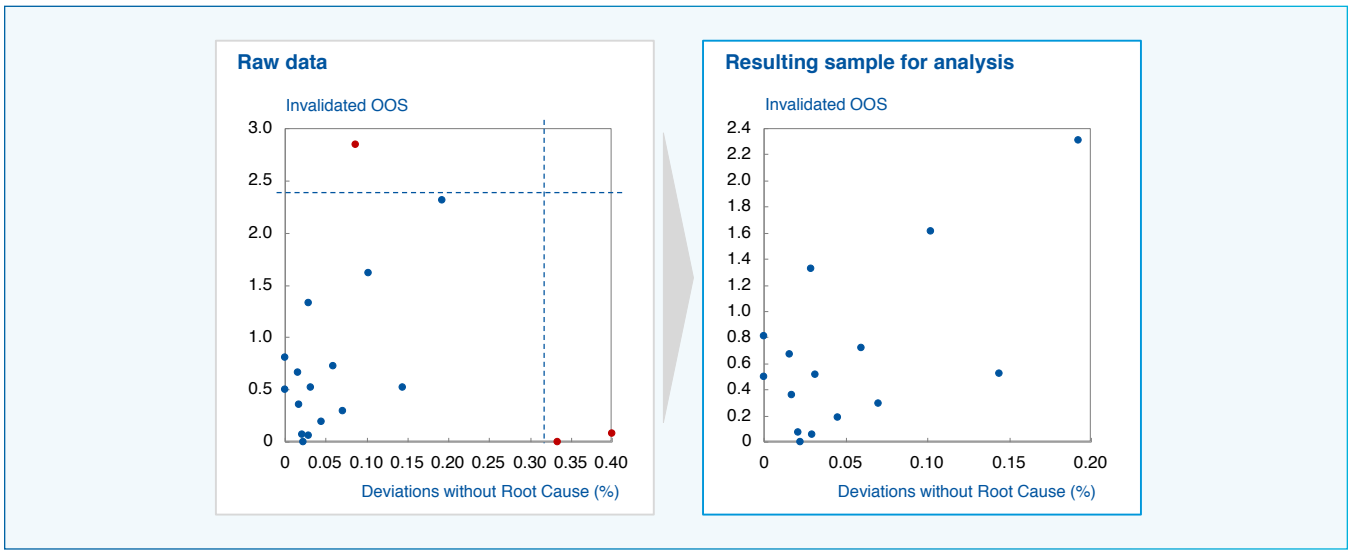
Please answer the questions below in the context of your full US supply portfolio, not just the 2-5 products you have chosen for the rest of the template.

DRUG SHORTAGES	RESPONSE	COMMENTS
1. Have you observed an actual shortage or potential for a shortage of a drug product that you needed to report to the Office of Drug Shortage?	[Y/N]	
2. If the answer to question 1 is "yes", did any metrics you track assist you in predicting or mitigating the potential drug shortage?	[Y/N]	
3. If the answer to question 2 is "yes" for those metrics that you have found helpful in predicting or mitigating the shortage: – Please list any of these metrics that are among the ones currently included in the ISPE Quality metrics pilot	<Text>	
– Please list any of these metrics that are not part of the ISPE pilot (including possibly process capability metrics)	<Text>	
4. Please list any quality metrics that you believe could assist in predicting drug shortages, whether you currently track them or not.	<Text>	

Appendix 3

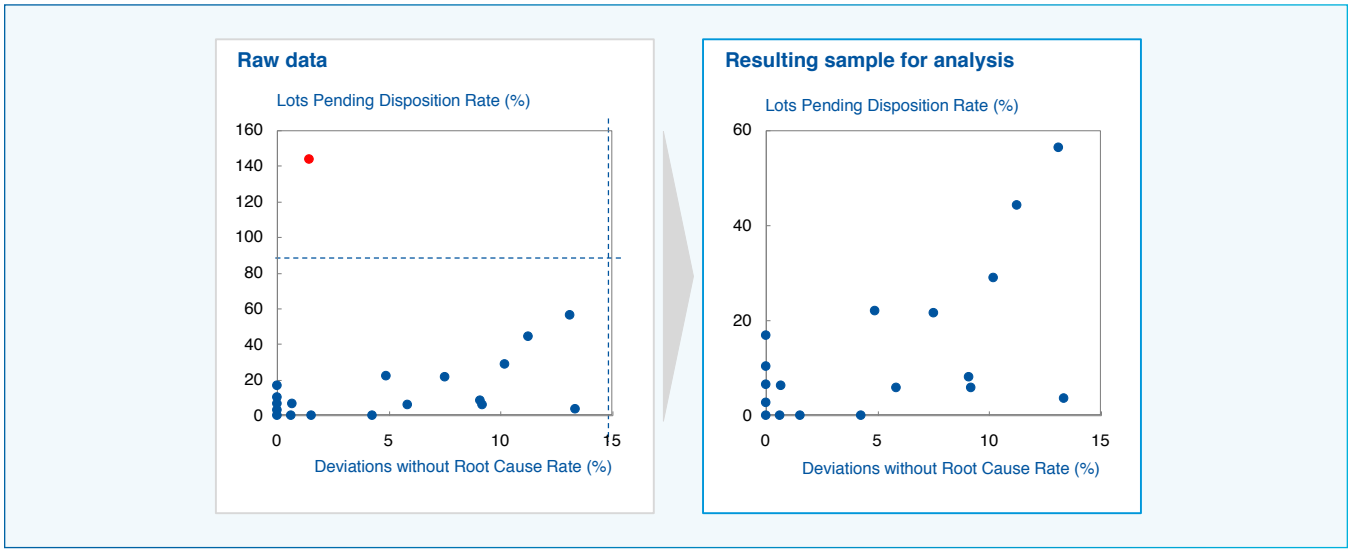
Full Data Sets before Outlier Analysis

Deviations without Confirmed Root Cause and Invalidated OOS



Note: Outliers more than two standard deviations away from sample mean were excluded: two on Deviations without Root Cause (LATAM-based, small-scale sites with < 10 deviations annually), and one on Invalidated OOS (Puerto Rico–based Rx site).

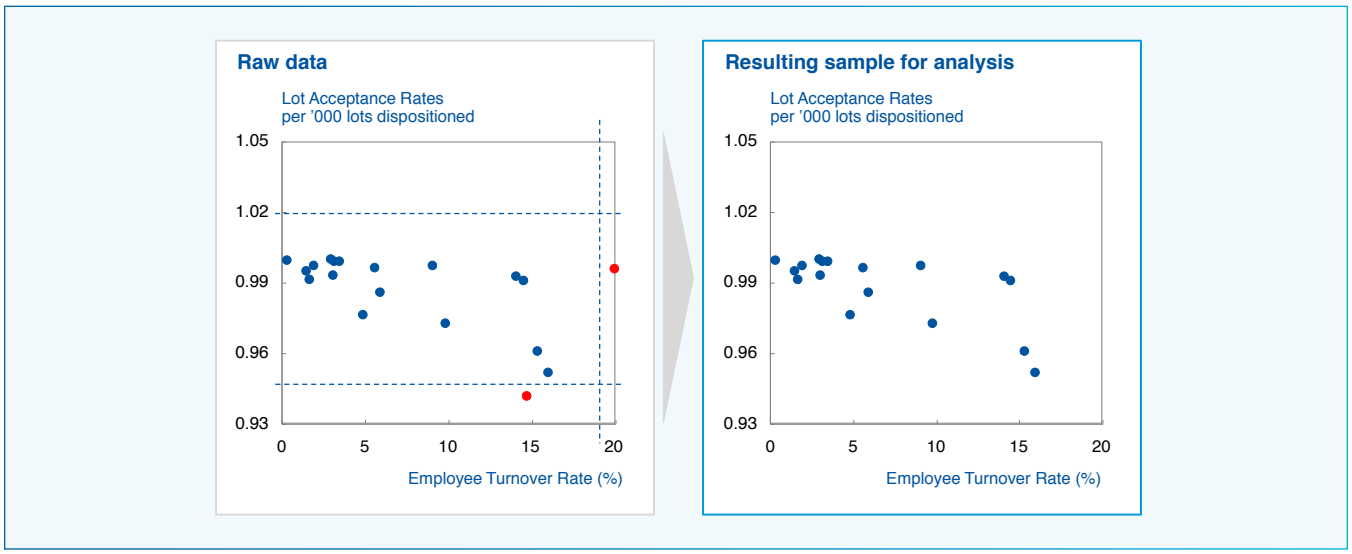
Lots Pending Disposition Rate and Deviations without Root Cause



Note: Outliers more than two standard deviations away from sample mean were excluded: one Puerto Rico Rx site on Lots Pending Disposition.

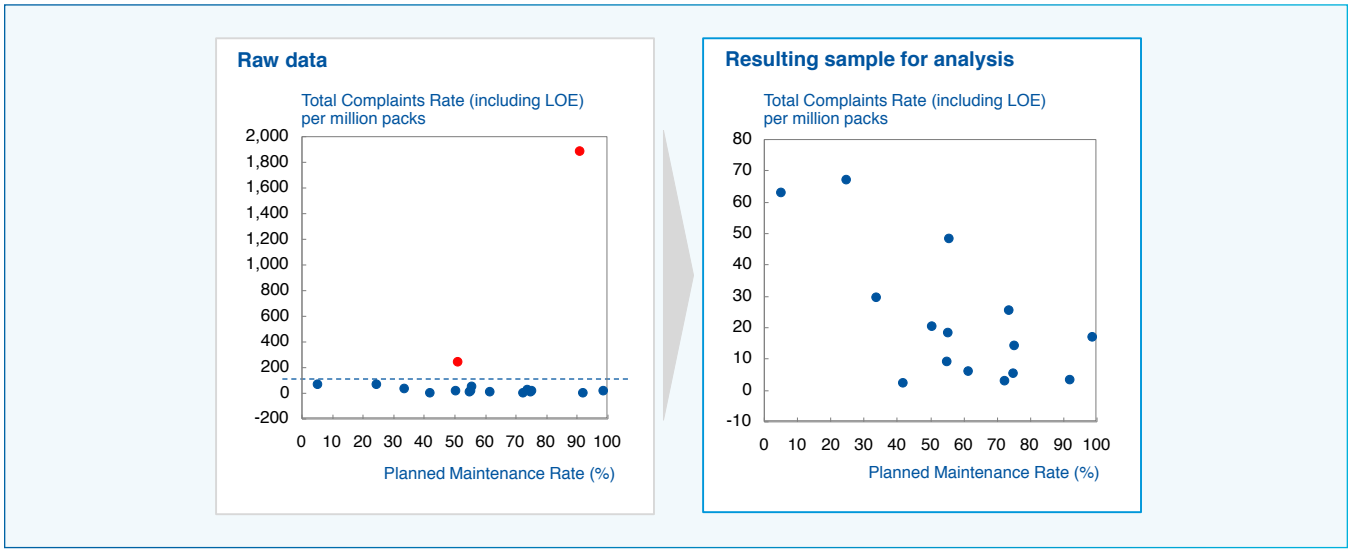
Appendix 3

Employee Turnover Rate and Lot Acceptance Rate



Note: Outliers more than two standard deviations away from sample mean were excluded: one on Employee Turnover Rate (US-based Rx site), and one on Lot Acceptance Rate (India-based Rx site).

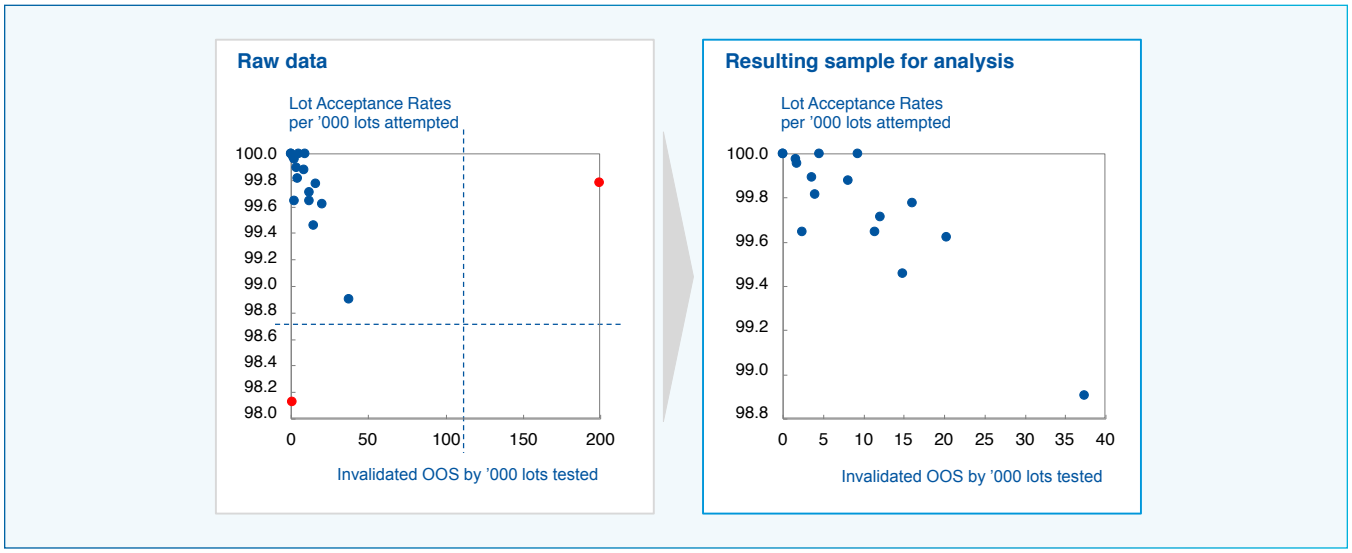
Planned Maintenance Rate and Total Complaints Rate



Note: Outliers more than two interquartile ranges away from sample median were excluded: two sites Total Complaints, both with very high LOE rates, one Europe-, one US-based, both Rx.

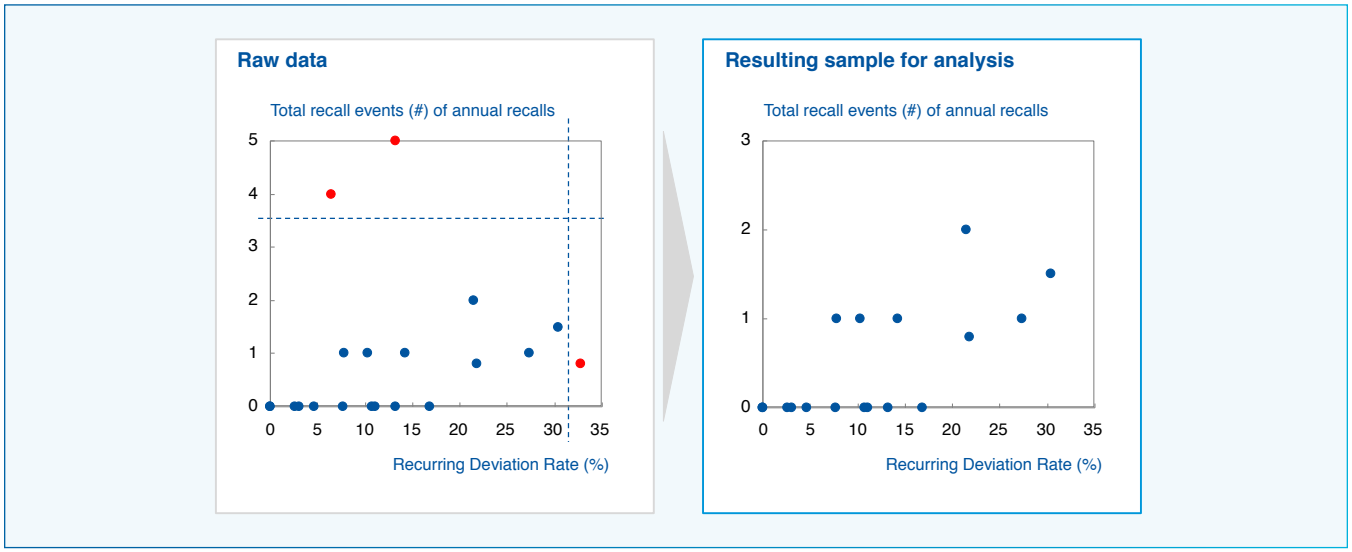
Appendix 3

Invalidated OOS and Lot Acceptance Rate



Note: Outliers more than two standard deviations away from sample mean were excluded: one US-based CMO site on Lot Acceptance Rate, and one European Rx site on Invalidated OOS

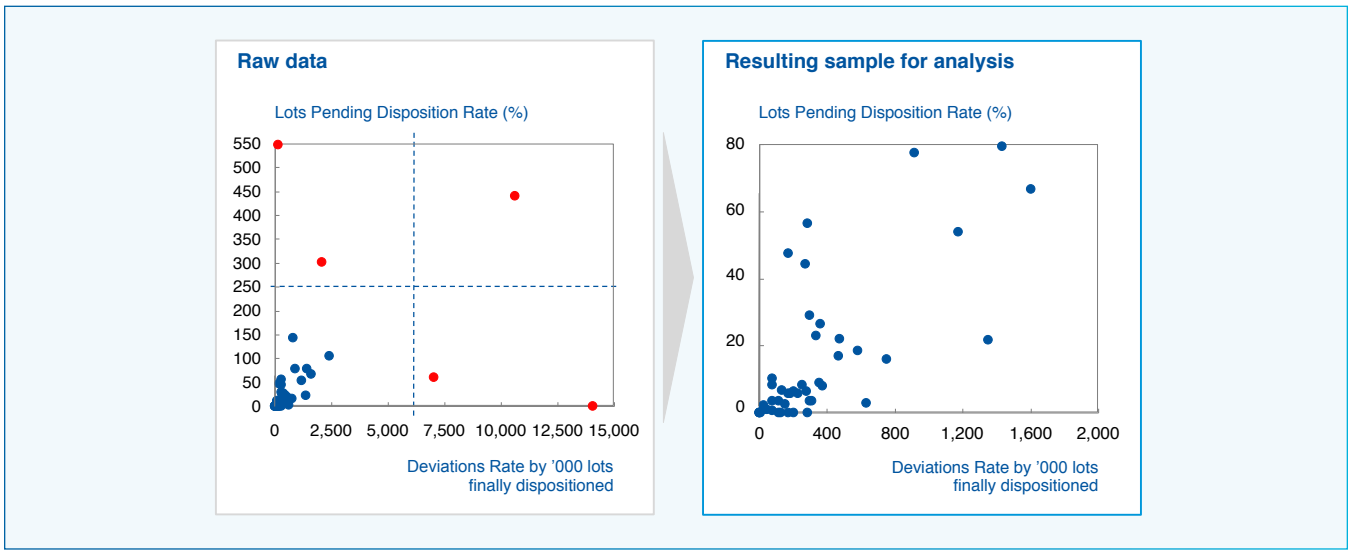
Deviations Recurrence and Recalls



Note: Outliers more than two standard deviations away from sample mean were excluded: two on Deviations without Root Cause (LATAM-based, small-scale sites with < 10 deviations annually), and one US-based Rx site on Recurrence.

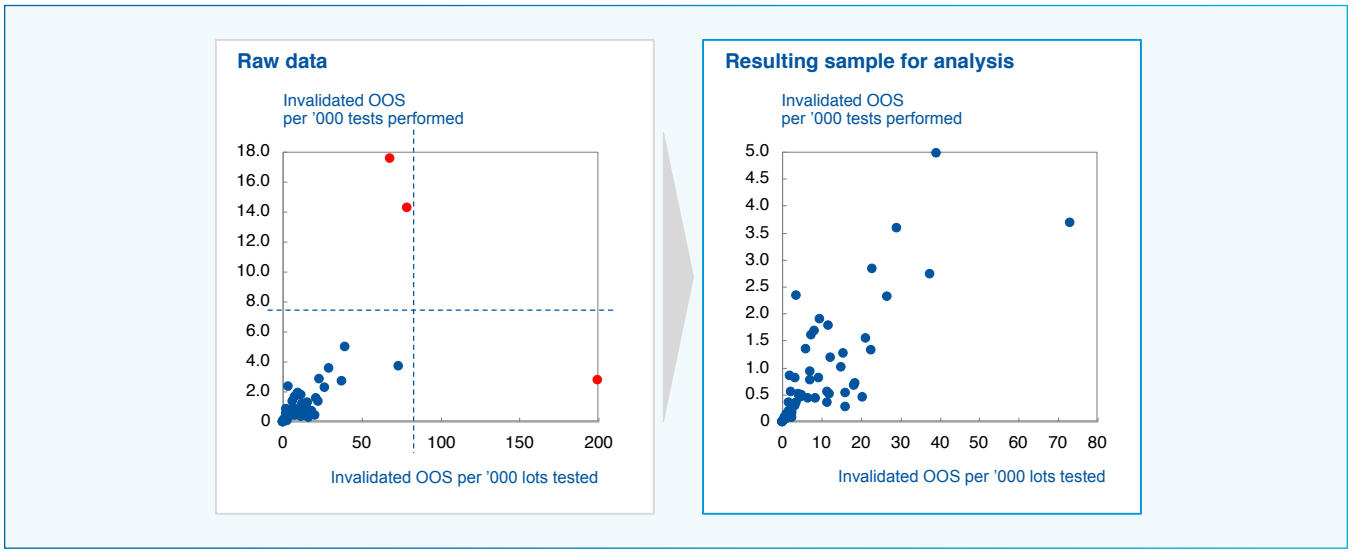
Appendix 3

Lots Pending Disposition and Deviations Rate



Note: Outliers more than two standard deviations away from sample mean were excluded: five sites were excluded, all drug substance sites.

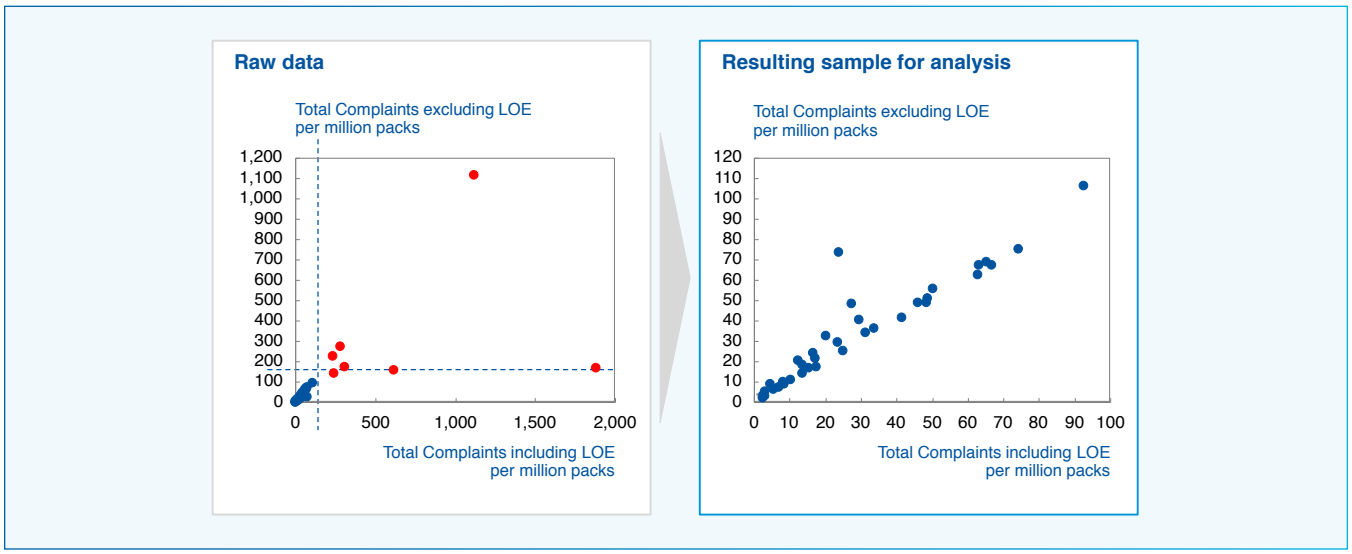
Invalidated OOS by test and by lot



Note: Outliers more than two standard deviations away from sample mean were excluded: one lab, one US-based drug substance site, and one Europe-based solids site

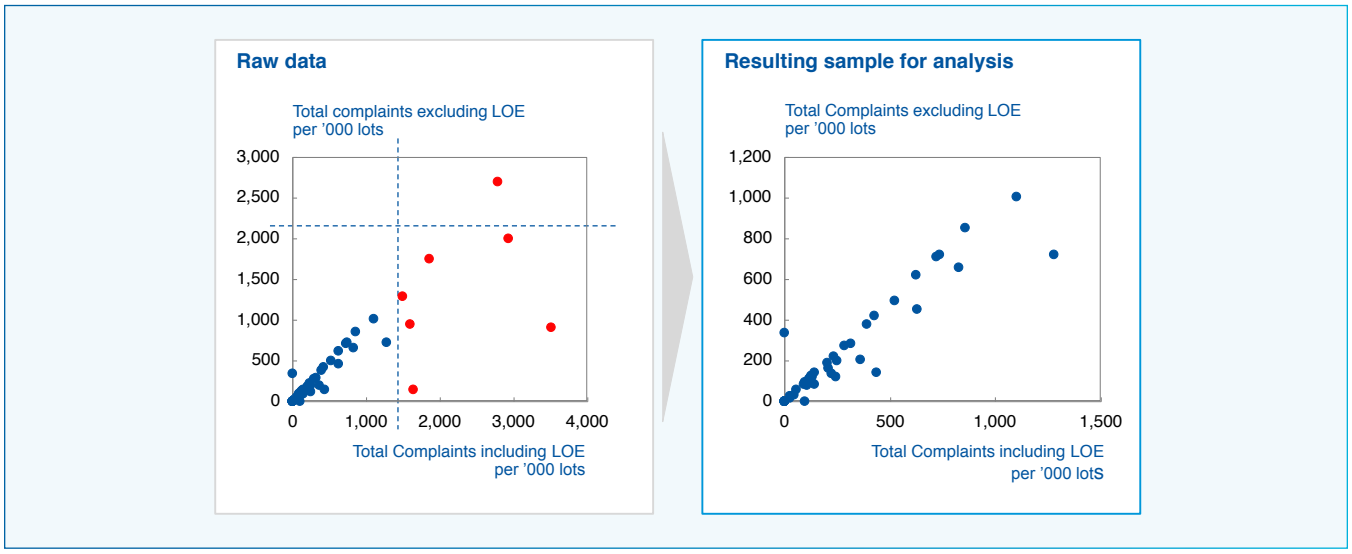
Appendix 3

Total Complaints per pack, including or excluding LOE



Note: Major outliers more than two interquartile ranges away from sample median were excluded: seven sites (two US-, two LATAM-, and three Europe-based).

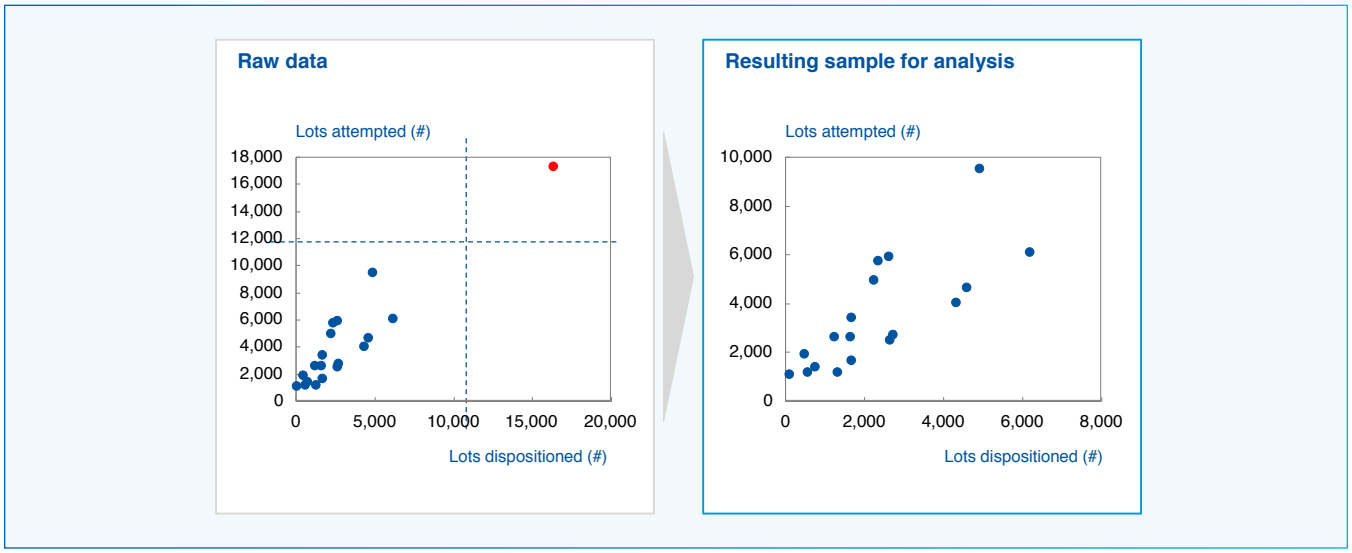
Total Complaints per lot, including or excluding LOE



Note: Major outliers more than two interquartile ranges away from sample median were excluded: seven sites (four Rx and three OTC),

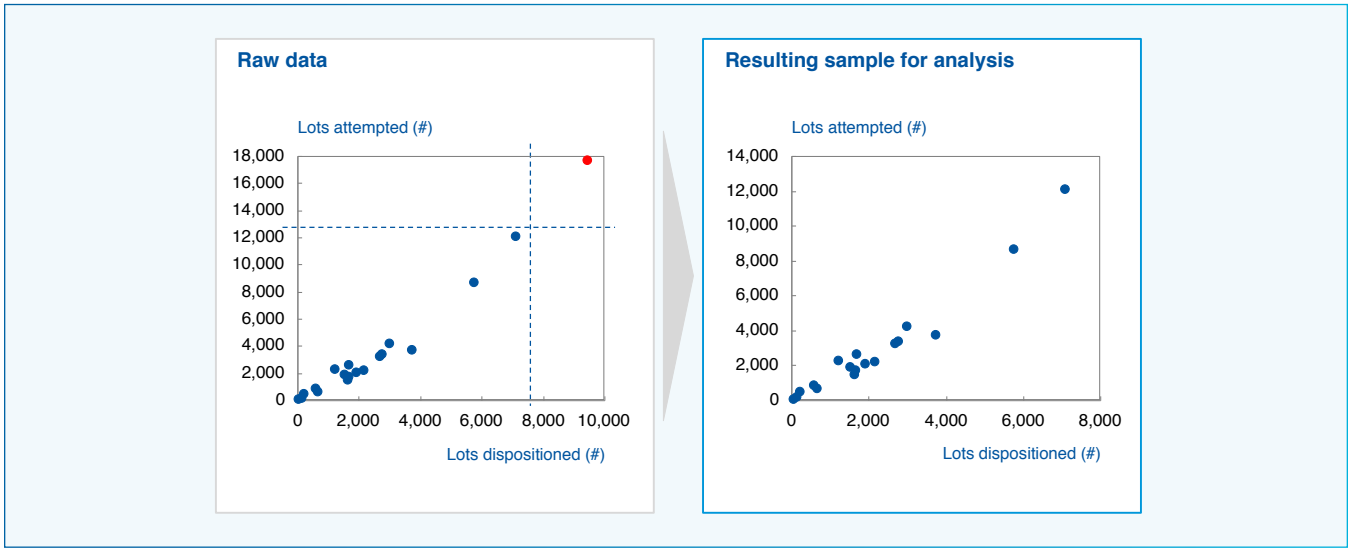
Appendix 3

Lots attempted and dispositioned: solids



Note: Outliers more than two standard deviations away from sample mean were excluded: one Europe Gx site

Lots attempted and dispositioned: steriles



Note: Outliers more than two standard deviations away from sample mean were excluded: one Europe Rx site

Appendix 4

Lots Finally Dispositioned vs. Lot Attempted

Definitions

ISPE definition of "lots finally dispositioned": *Total number of lots for commercial use produced and/or packaged on site that went through final disposition during the period, i.e. were released for shipping or rejected (for destruction). Rejections should be counted as final disposition regardless at what production stage the rejection occurred. Release is only final release for shipping (whether shipping bulk to another site or final product to market). Excludes lots that have been sent for rework or put on hold/quarantined in this period and hence are not finally dispositioned. Excludes lots that are not produced or packaged on site, but just released for CMOs.*

FDA Definition of "lot attempted": *A lot intended for commercial use for which the manufacturer has issued a lot number and charged API (for finished drug manufacturers) or primary starting materials (for API manufacturers).*

Lot attempted values are used in the following calculation:

Lot Acceptance Rate = 1– x

Where x = the number of specification-related rejected lots in a timeframe divided by the number of lots attempted by the same establishment in the same timeframe).

Pros and cons: Lots finally dispositioned

Pros

1. "Disposition" is a term used in CFRs 210 and 211 [21 Code of Federal Regulations, Parts 210 and 211 Current Good Manufacturing Practice for Finished Pharmaceuticals, 1978, www.FDA.gov] albeit "disposition" rather than "dispositioned." "Disposition" is mentioned in CFRs 210 and 211 6 times—211.42 (c)(2), 211.80 (d), 211.184 (e), 211.204 twice and 211.208.
2. "Disposition" is used 13 times in the preamble to CFR 210 and 211. A relevant positive comment from the preamble is:

Comment 208: *Some comments requested a clarification of the word "disposition" in 211.80(d) that would distinguish between a simple transfer of the material and use of the material for a particular purpose.*

The Commissioner believes that the word "disposition" appropriately covers any use or change in control status of the lot, including both those cited in the comments.

Cons

1. In the CFR "disposition" refers most frequently to components, containers, closures, and labeling materials, and in returned drug products and drug product salvaging sections (211.204 and 211.208).
2. In ICH Q7 "disposition" is also used in relation to rejection of materials (section 4.14, 14.10 and 18.36) and to components. In section 7.24 "disposition" is used in relation to recording of use of materials in batch manufacture.
3. "Disposition" in Webster's dictionary does refer to "final arrangement" and "transfer to care or possession of another." These phrases **are not** clearly associated with reject/release decisions in pharmaceutical manufacture. More common uses of "disposition" are related to moods, behavior and temperament.
4. Relatively negative comments relating to "disposition" in the preamble to 210 and 211 are given below.

Comment 203: *Comments were received on 211.80(a) recommending replacement of the phrase "approval or rejection" with the word "disposition."*

The Commissioner disagrees with this suggestion. Written procedures must spell out the criteria for approval or rejection in view of such material's intended use.

Comment 254: *Several comments suggested that this section be expanded to deal with the subsequent disposition of rejected materials.*

The Commissioner notes that the criteria for reprocessing rejected materials are adequately covered in other sections of this part. It is not necessary to deal with other methods of disposition because they are varied, are within the manufacturer's discretion, and may include destruction, return to the supplier, or use in other products where specifications are met. The Commissioner believes the major concerns of FDA are that rejected materials are not inadvertently used in a product for which they are not acceptable and that any such materials that are reprocessed and found suitable for reuse meet specifications, standards, and characteristics for the intended use.

Comment 505: *Several comments argued that the recordkeeping provisions of 211.204 are unnecessary, redundant, and unduly costly for many operations.*

The Commissioner does not agree with this position. He does not believe such recordkeeping to be unduly burdensome. This section does not require separate records for returned goods containing all the information required by this section, but rather requires firms to be able to identify which, if any, drug products have been returned and for what reason, and to be able to determine their disposition. The section would not prevent, for example, the disposition portion of the records on returned goods from being a part of normal distribution records if the lot involved were reshipped. The Commissioner does not agree that the requirements for returned goods are unnecessary. For example, a portion of a lot of drug product may be returned because of unusual shipping conditions and the rest of the lot remain in normal trade channels. If the returned portion of the lot were destroyed and no record were made, there would be an incomplete record of distribution for the lot. In the case of recall, for example, a part of the lot could not be accounted for.

In summary, FDA did not like "disposition" being used for reject or rejection decisions (comment 203). Comments 254 and 505 are associated in part with rejection-type decisions.

Appendix 4

Pros and cons: lot attempted

Pros

Reference 46 in the FDA Quality Metrics Draft Guidance relates to 21 CFR 211.101, "Charge-in of components." There is no clear reference that this section refers to a lot attempted, although this step could be inferred (See Appendix 5 for a full transcript.).

Cons

"Attempt" and "attempted" whether associated with "lot" or otherwise **do not** appear in CFR 210 and 211 for drug product or in ICH Q7 for drug substance or in the preamble to 210 and 211.

Reference 46 in the FDA Quality Metrics Draft Guidance relates to CFR 211.101, "Charge-in of components." This section **does not** specifically require a lot number to be assigned and used throughout manufacture:

If a component is removed from the original container to another, the new container shall be identified with the following information:

1. *Component name or item code*
2. *Receiving or control number*
3. *Weight or measure in new container*
4. *Batch for which component was dispensed, including its product name, strength, and lot number*

There could be some inference that one lot number may be used throughout batch manufacture; this is not clear, however. Based on findings from Wave 2, industry practice is that lots attempted does not always correlate with batches finally released or rejected. ISPE's response to the FDA Draft Guidance particularly comments on proposed definition of "lot attempted":

For example, relating to definition of a lot attempted, different work flow designs result in variety of practices as to when new lot number is assigned. A single lot number could be associated with a single release step or with multiple release steps or no release step. These differences are a function of work order/electronic batch manufacturing instruction design.

Similarly, ICH Q7 for drug substance in section 6.5, “Batch Production Records (Batch Production and Control Records)” and section 6.51 states:

These records should be numbered with a unique batch or identification number, dated and signed when issued. In continuous production, the product code together with the date and time can serve as the unique identifier until the final number is allocated.

There may be an inference that the same unique batch identification number should be used through manufacture of a batch of drug substance. One batch of intermediate, however, may lead to several batches of drug substance, or several batches of an intermediate may be used in a single batch of drug substance. In addition, firms often assign a unique batch identification number to a step that produces an isolated intermediate, and a different batch identification number to a subsequent step. Clear reference is given in batch production records of actual batch use to meet GMP traceability requirements.

In summary, based on CFR 210 and 211, ICH Q7, and the preamble to 210 and 211, it is hard to understand the choice of lot attempted as the denominator.

Conclusion

A rationale for use of lot attempted based on CFR 210 and 211 and ICH Q7 is very hard to understand. The phrase is not used and, therefore, not defined.

“Disposition” is used in CFR 210 and 211, in ICH Q7, and the preamble to 210 and 211. FDA, however, did not like use of “disposition” as an alternative to “release/reject,” nor is “disposition,” used in CFR 210 and 211 to refer to reject/release decisions of a lot of drug product or in ICH Q7 to release/rejection of a lot of drug substance.

“Disposition” in CFR 210 and 21, in ICH Q7, and to some extent in the preamble to 210 and 211, is more associated with components, containers, closures, and labeling materials and their use than release/reject decisions of drug substances/drug products. It also is associated more with rejection rather than drug product or drug substance release.

Another word or phrase for “disposition” should be considered.

Appendix 4

Recommendation

A recommendation for an alternative definition to lots finally dispositioned is:

Lots released or rejected: Total number of lots for commercial use produced and/or packaged on site that went through final release/reject decision during the period, i.e. were released for shipping or rejected (for destruction). Rejections should be counted regardless at what production stage the rejection occurred. Release is only final release for shipping (whether shipping bulk to another site or final product to market). Excludes lots that have been sent for rework or put on hold/quarantined in this period and hence are not finally released or rejected. Excludes lots that are not produced or packaged on site, but just released for CMOs.

Rationale

1. "Release" and "reject" are used throughout CFRs 210 and 211 and ICH Q7, and are understood by practitioners. "Release" is often associated with a release for [something], such as release for distribution, release for manufacture or release for shipping. Neither word is defined in CFR 210 and 211 or ICH Q7.

   In this suggested definition, "release" is associated with an appropriate action as given by precedent in CFR 210 and 211.

2. Although not defined as such in CFR 210 and 211 or ICH Q7, "reject" is used in CFR 210.3, subsection (b)(20):

   (20) "Acceptance criteria" means the product specifications and acceptance/rejection criteria, such as acceptable quality level and unacceptable quality level, with an associated sampling plan, that are necessary for making a decision to accept or reject a lot or batch (or any other convenient subgroups of manufactured units).

3. The FDA definition of Product Quality Complaint Rate included a reference to "product released," which gives further precedent for use of a phrase covering "release" (and "reject"). The FDA definition of Product Quality Complaint Rate is:

   ***Product Quality Complaint Rate** = the number of product quality complaints received for the product divided by the total number of lots of the product released in the same timeframe.*

Appendix 5

CFR 211.10

Transcript of CFR 211.101: Charge-in of components

Written production and control procedures shall include the following, which are designed to assure that the drug products produced have the identity, strength, quality, and purity they purport or are represented to possess:

(a) The batch shall be formulated with the intent to provide not less than 100 percent of the labeled or established amount of active ingredient.

(b) Components for drug product manufacturing shall be weighed, measured, or subdivided as appropriate. If a component is removed from the original container to another, the new container shall be identified with the following information:

- (1) Component name or item code;
- (2) Receiving or control number;
- (3) Weight or measure in new container;
- (4) Batch for which component was dispensed, including its product name, strength, and **lot number**.

(c) Weighing, measuring, or subdividing operations for components shall be adequately supervised. Each container of component dispensed to manufacturing shall be examined by a second person to assure that:

- (1) The component was released by the quality control unit;
- (2) The weight or measure is correct as stated in the batch production records;
- (3) The containers are properly identified.

(d) Each component shall be added to the batch by one person and verified by a second person.

Appendix 6

Lots Attempted vs. Lots Dispositioned

Process step vs. disposition decision

Not every process step in manufacturing has a specific positive disposition decision. Although product can be rejected at any stage due to an in-process test or quality reason, there may not be an affirmative disposition decision on in-process steps before going to the next step. In many companies, a disposition decision occurs only at the final-packaged-form step and possibly the final-bulk-form step.

For OSD, this might look like the following:

- Milling, dispensing, sifting, granulation, and compression would all occur without a specific disposition decision; that is, they would progress at risk, albeit with some in-process testing such as moisture and hardness, with quality control testing of the final compressed bulk tablets.
- The disposition decision may only occur on the final compressed bulk tablets. At many companies, however, the tablets might go to packaging at risk, i.e., without waiting for the test results; in that case, there would be one disposition of the final dosage form that would encompass all the in-process and final-release testing.
- The final compressed bulk tablets would go to packaging, where there would be a specific disposition decision made on the final dosage form.

This means that for lots attempted, as FDA has defined it, there would be at least six lots associated with the above scenario, assuming unique lot numbers were assigned at each process step. The number would be greater than six, typically, because often there is more than one lot produced at each given step (e.g., multiple API intermediates combined into one granulation or multiple granulation lots combined into one compressed lot).

By comparison, in the above scenario, there would be one or two lots dispositioned, depending on whether or not the company waits to disposition final compressed bulk tablets prior to packaging.

Based on this simple scenario, there would be at least a three- to six-fold difference in the number of lots attempted vs. lots dispositioned.

Appendix 6

Effect on Lot Acceptance Rate calculation

Example calculation:

- A site dispositions only after final packaging; it has 1,000 lots dispositioned
- This same site assigns lot numbers at three process steps: blending, tableting, and final packaging.
- Its lots attempted would be around 3,000, assuming no splitting or aggregating
- If the site has 50 rejects during the same period, the Lot Acceptance Rate metric using lots dispositioned will be 95%, and using lots attempted 98.3%.

This difference in outcome is driven not by quality or performance differences, but by an administrative choice on how to calculate Lot Acceptance Rate. Additionally, site decisions regarding when to assign new lot numbers during processing and when to assign dispositions will contribute to vary across sites.

Increased burden

Lots dispositioned are generally easier to collect because there is a disposition decision that is tracked, usually in a searchable electronic system, for production steps having a particular decision point.

Lots attempted are more difficult to collect, especially in cases where in-process lots are documented on paper batch records and not in searchable electronic systems. This means that personnel must read through every batch record and possibly investigations to count in-process lots, making this highly burdensome.