ISPE Quality Metrics Initiative

A Report from the Pilot Project
Wave 1

June 2015

ISPE
600 N. Westshore Blvd., Suite 900
Tampa, FL 33609 USA

Tel: 813-960-2105 – Fax: 813-264-2816 – ask@ispe.org

www.ISPE.org

Table of Contents

1 Executive Summary

ISPE commenced its Quality Metrics Initiative in June 2013 after the US Food and Drug Administration (FDA) announced its Quality Metrics Program in a February 2013 Federal Register notice [15]

To assist in the evaluation of product manufacturing quality, FDA is exploring the broader use of manufacturing quality metrics.

Through an extensive series of engagements with industry and other key stakeholders over the past two years, FDA has further indicated that "an objective set of quality metrics" would be reportable to support their risk-based inspection program, as given in sections 704 – 706 of the US FDA Safety and Innovation Act (FDASIA) [9]. The FDA Quality Metrics Program is also intended to move both industry and the agency toward the desired state [18] for pharmaceutical manufacturing. The FDA Quality Metrics Program, including the set of metrics selected, is expected to be published for public comment in 2015.

In a white paper delivered to FDA in December 2013 [12], ISPE recommended that a pilot program should be conducted within industry to further understand the implementation opportunities, challenges and benefits available from such a quality metrics program. ISPE, in cooperation with McKinsey and Company, undertook this project. The result was the ISPE Quality Metrics Pilot Project—Wave 1. Designed and developed by the ISPE Quality Metrics Core Team, the project drew on the knowledge and experience of cross-functional industry representatives, ex-regulators and academicians, with further insight gained in detailed discussions with a variety of industry associations at many industry meetings.

The ISPE Wave 1 Pilot ran from June through November 2014 and included:

- Data collected at 44 sites from 18 participating companies
- Data was collected retrospectively for 12 months and prospectively for 3 months at each site.
- A Wave 1 set of quantitative quality metrics
- Nearly all metrics collected were reported at site level; three were collected at product level within each site.

1.1 Main Findings: Summary

The Wave 1 Pilot met its overall objectives. 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 in addition to their established programs.
- Understanding organizational 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.

The success of the Wave 1 Pilot can be traced to the following factors:

- Using a standardized set of metrics with clear and specific definitions provided for each of the metrics measured.
- Excellent collaboration between all stakeholders.
- Frequent and direct interaction for guidance and query resolution between the McKinsey support team and participating sites.
- Leveraging McKinsey's experience and capability in metrics program delivery.
- Leadership from the ISPE Quality Metrics team and the sponsorship from the leaders at the participating sites.
- Ongoing dialog and trust-building with FDA throughout the pilot period.

The Wave 1 Pilot also identified several challenges that are present in rolling out a centrally reported standardized metrics program. These include:

- The industry and its sectors have not traditionally shared a common definitions. Consequently, definitions of each metric must be specific, clearly understood and meaningful across the range of organizations under consideration to ensure the program's success. This will require detailed up-front design and ongoing operational support.
- The level of effort required for data collection and submission on behalf of the industry, and data analysis and support on behalf of the agency cannot be underestimated. Refer to Section 5 and Section 5.14 of the report for estimates of the effort involved. However, these estimates may be considered conservative because they do not include several factors, such as:
  - A "good enough" situation was applied to data submission in the Wave 1 Pilot. Submission to FDA would require more thorough and complete data collection, additional management review and data verification.
  - Pilot participants had the flexibility to provide their most pragmatic data set (e.g., all products at the site or only those for the US market); this would not be the case in a formal submission process.
  - Pilot participants typically had mature systems and capabilities and were from developed countries; this would not be the case for all sites under a centralized reporting initiative.
- Understanding the variation in ranges for interpretation of the data will require longer timeframes to assess than those examined in the Wave 1 Pilot.

1

Details of the statistical analysis, main outcomes and recommendations arising from the Wave 1 Pilot can be found in Section 5 and Section 6 of this report. A summary of these findings is as follows:

- Even with 44 sites reporting in this initial phase of the ISPE Quality Metrics Initiative, the sample sizes allowed statistical analysis with some limitations; these are outlined in more detail in Section 5.3 of this report.
- The analysis of the Wave 1 Pilot data identified a number of statistically significant (less than 5% likelihood of a coincidence) relationships between the metrics collected and overall quality outcomes at the sites.
- These initial findings of statistically significant relationships do not imply causation, therefore this report does not attempt to draw conclusions from this phase of the data analysis. Instead, this report is intended to share the findings, identify recommendations and put forward proposals for the next phase of the study.
- To meet one of its objectives, this pilot targeted a set of metrics collected at both the site and the product level to understand the current capacity to collect metrics. Product-level metrics require an understanding of the challenges of aggregating metrics across the supply chain for multisite products. In addition, the definition of "product" as related to an "application number" presented issues for over-the-counter (OTC) products. This element of the pilot has led to some key learnings and recommendations for both industry and FDA, and these are included in the report.
- Quality metrics reporting alone should not be the basis for action (either positive or negative) without understanding the context of the data and the originating company.
- Choosing an appropriate metric set will help identify continual improvement opportunities.
- The knowledge gained from the Wave 1 Pilot has been leveraged to develop a revised set of starting metrics that ISPE now proposes for further analysis in a Wave 2 pilot program. Details of this proposal can be found in Section 7 of this report.
- Learnings from the Wave 1 Pilot have also been shared with FDA for consideration in the design of agency's final set of objective metrics.

1.2 Next Steps for ISPE's Quality Metrics Initiative

Based on the findings from the Wave 1 Pilot, ISPE now recommends the following set of starting metrics:

1. Lot acceptance rate (normalized by lots dispositioned), collected at site level
2. Lot acceptance rate (normalized by lots dispositioned), collected at product level within a site
3. Critical complaints (normalized by packs released), collected at product level by each product application, not broken down by site
4. Critical complaints (normalized by packs released), collected at site level, undifferentiated by product
5. Deviations rate at site level

Following the 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 is a continuing appetite within industry for additional learning with respect to quality performance measures.

ISPE has therefore initiated planning for a Wave 2 Pilot, to commence in the second half of 2015. This second phase will test the starting set of metrics on an extended sample and time period to increase the range and duration of the knowledge base and enable more in-depth statistical data analysis to examine correlations and dependences. The Wave 2 Pilot will also explore the inclusion of other potential metrics of interest and further study of the assessment of quality culture at participating companies.

It is hoped that continuing this work will enable the pharmaceutical industry to undertake the "quality revolution" [16] proposed by Dr. Janet Woodcock at the ISPE Quality Metrics Summit to truly enhance the future state of pharmaceutical manufacturing.

Sincere gratitude is extended to the participating companies and their staff for the excellent input, support and enthusiasm given throughout this Wave 1 Pilot.

2 Background

This section describes the:

- ISPE Quality Metrics Initiative background
- Quality Metrics Pilot Program, a major component of this project
- Wave 1 Pilot background, rationale, and key milestone dates

The FDA's vision for twenty-first century manufacturing in the pharmaceutical industry is often quoted as the "desired state" and is:

A maximally efficient, agile, flexible pharmaceutical manufacturing sector that reliably produces high quality drugs without extensive regulatory oversight.

There have been many guidelines issued by FDA [1][2][3][4] and the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use—Q8 [5], Q9 [6], Q10 [7] and Q11 [8]. These provide a regulatory framework that allow industry and regulators to move toward this desired state. Despite the introduction of the new guidance, recent acknowledgements confirm that more work is required by both the industry and the regulatory community to attain the desired state.

To help FDA and industry work toward the twin goals of ensuring product quality in a global supply chain and reducing drug shortages, FDASIA [9] (July 2012) gave the agency new authority to enhance the safety of the drug supply chain and created legislative mandates affecting current good manufacturing practices (CGMPs). FDASIA also required FDA to implement a risk-based inspection program of pharmaceutical manufacturing sites rather than the current two-year inspection cycle in effect.

Of relevance to a risk-based inspection program are sections 704, 705 and 706 of FDASIA relating to advanced provision of information (e.g., quality metrics).

- Section 704 "… enables FDA personnel to search the database by any field of information submitted in a registration …"
- Section 705 requires "risk-based schedule for drugs" and lists "risk factors" as:

  (A) The compliance history of the establishment.

  (B) The record, history, and nature of recalls linked to the establishment.

  (C) The inherent risk of the drug manufactured, prepared, propagated, compounded, or processed at the establishment.

  (D) The inspection frequency and history of the establishment, including whether the establishment has been inspected pursuant to section 704 within the last 4 years.

  (E) Whether the establishment has been inspected by a foreign government or an agency of a foreign government recognized under section 809.

  (F) Any other criteria deemed necessary and appropriate by the Secretary for purposes of allocating inspection resources.

Section 706 requires "… records or other information … be provided ... in advance or in lieu of an inspection ..." These "records or other information" are interpreted as provision of quality metrics data as part of other information which could potentially be requested.

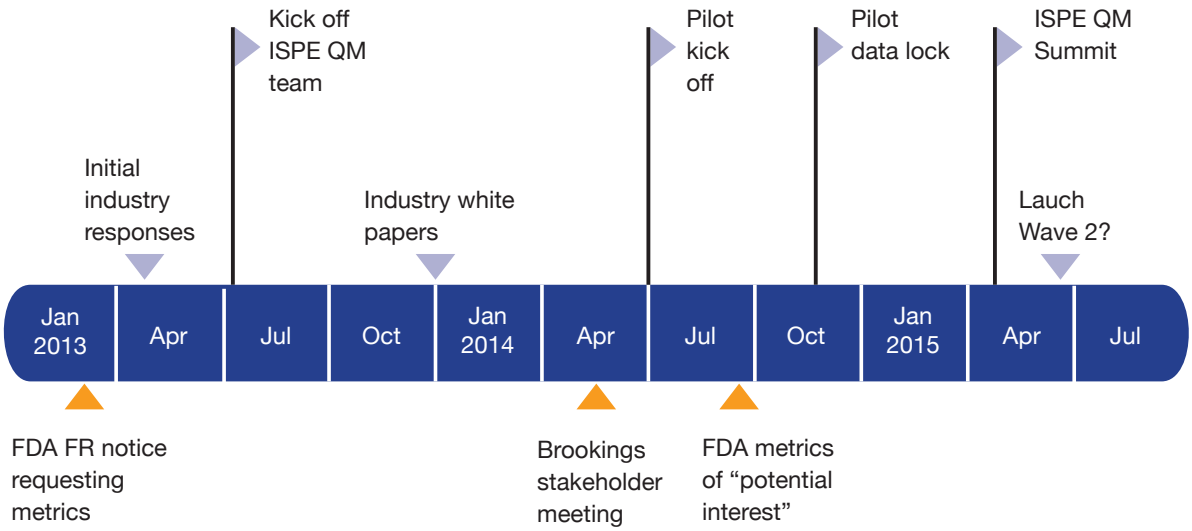
In response to the FDA initiative on quality metrics, ISPE established a Product Quality Lifecycle Implementation® (PQLI) – sponsored Quality Metrics project with a team that consisted of representatives from a variety of pharmaceutical companies.

Guiding principles established at the commencement of this project stipulated that any proposed metrics would be:

- Clearly defined to allow consistent reporting across sites
- Objective and meaningful
- Easy to capture
- Easy to report
- Normalized by factors such as process differences and technical complexity
- Able to drive acceptable, not unwanted behaviors

This team started work at a well-attended session of the ISPE–FDA CGMP conference in Baltimore on 12 June 2013, at which the both FDA and industry were represented. The initial objective for ISPE's project team was to analyze and use the output from the discussion at this meeting as input to a white paper to be issued for discussion with the FDA. A summary of main milestones for the Quality Metrics Pilot Program is shown in Figure 1.

Figure 1: Major Milestones for the ISPE Quality Metrics Project



2

The ISPE white paper, published in December 2013, proposed a list of metrics acceptable to industry that could be reportable to FDA to support a risk-based inspection program. The white paper's main recommendations were to:

- Conduct a pilot to flesh out standard definitions and approach.
- Initiate with site metrics collection, with the potential to move to product metrics later.

Based on these recommendations, ISPE announced its intention to conduct a Quality Metrics Pilot Program on 12 March 2014.

During this period, FDA expressed a desire for industry input on the development of FDA's quality metrics program. FDA then participated in discussions with industry at the Measuring Pharmaceutical Quality through Manufacturing Metrics and Risked-Based Assessment meeting held 1–2 May 2014 and hosted by the Engelberg Center for Health Care Reform at the Brookings Institution. [10] A next step identified at this meeting was:

> *That the pilot quality metrics programs currently under development by the International Society for Pharmaceutical Engineering … may yield important lessons for FDA as it moves forward with its own program.*

In preparation, ISPE explored the options of conducting the Quality Metrics Pilot Program in cooperation with a suitable independent partner that could provide operational expertise and assure participant confidentiality during the pilot. ISPE subsequently agreed to partner with McKinsey and Company, due to their experience of conducting industry-benchmarking programs, specifically the Pharma Operations Benchmarking of Solids (POBOS) series of programs, [11] which have been operating since 2004. It was recognized that benchmarking programs require significant expertise to succeed, such as:

- Development of templates to allow for ease of input of data.
- Structured data submission with detailed guidance on how to report the data.
- Ability to comment on data points to enable interpretation.
- Experienced dedicated support for questions and clarifications during and throughout the data collection.
- Built-in data validity checks and joint review to ensure data consistency and accuracy.
- Development and operation of supporting IT systems.
- Relevant high-level statistical expertise to assist in data interpretation.
- Autonomy and confidentiality in data collection, review and analysis.

ISPE announced the launch of the Quality Metrics Pilot Program in partnership with McKinsey and Company at the ISPE–FDA CGMP conference on 2 June 2014. It was intended that the pilot project would have phases: Wave 1 and Wave 2. This report summarizes the Wave 1 Pilot and proposes recommendations for programs and metrics for consideration in a future Wave 2 Pilot.

2

Wave 1 Pilot was intended to demonstrate the feasibility and value of standard quality metrics. Some important primary objectives were to:

- Test the harmonization of definitions for a set of industry metrics that represent both leading and lagging indicators.
- Test the feasibility of centralized data collection across companies at different maturity levels within their own internal metrics programs.
- Explore industry practices in the areas of quality culture and process capability.
- Inform continued industry input to FDA.

Industry participants were intended to gain the benefits of:

- Influencing the output from the ISPE Quality Metrics Pilot Program in terms of choice of metrics, definitions and ease of data collection based on actual experience.
- Receiving a blinded comparison or benchmark to the participating site industry average and to similar technology platform peers (provided sufficient sample size is achieved).
- Having an opportunity to develop or enhance internal procedures for metric collection along with a set of metric definitions.
- Gaining insight into the implications of external metric reporting.

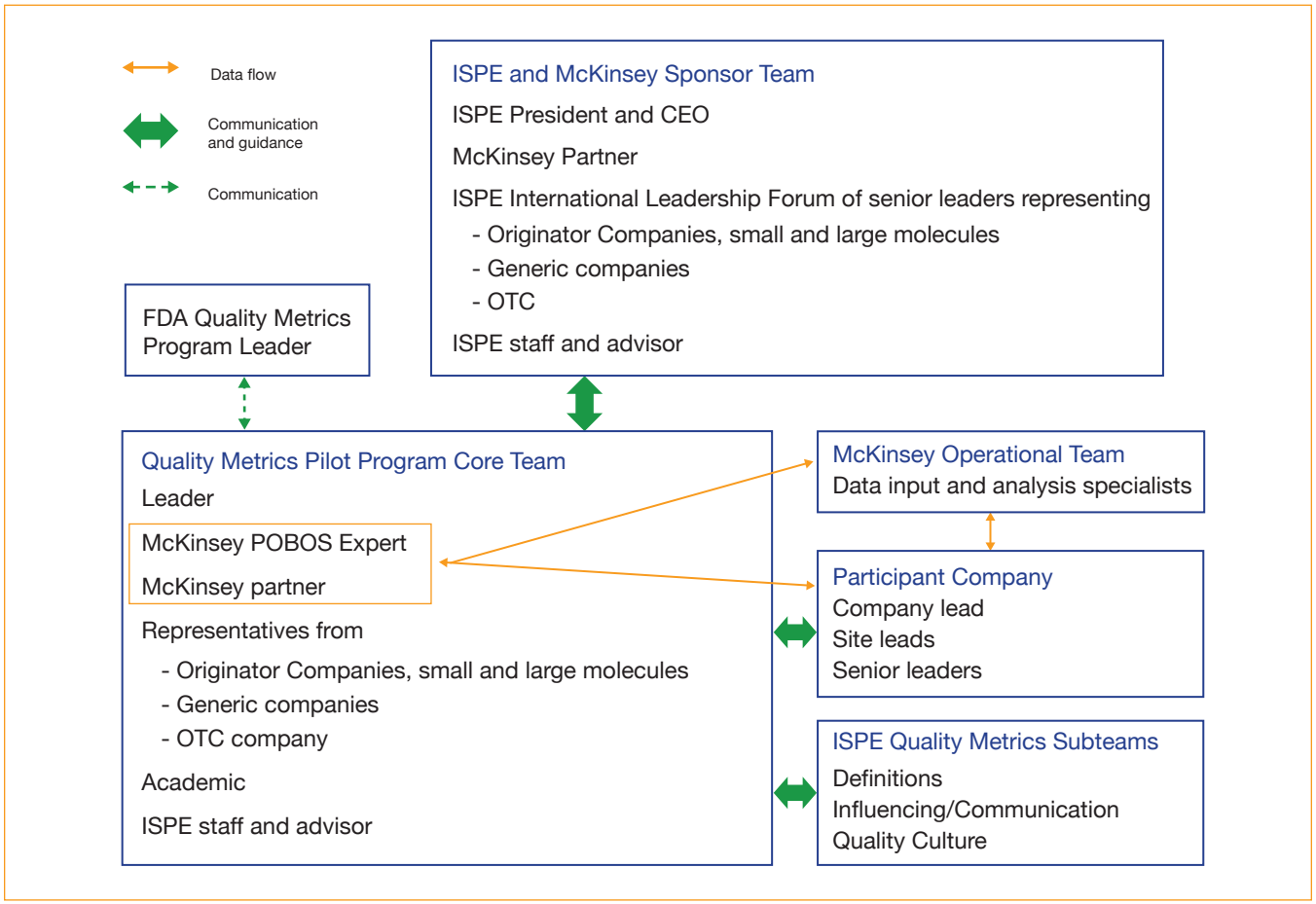
During the period from the white paper's issue to initiation of the Wave 1 Pilot, considerable attention was given to choice of metrics to be included. Consideration was given to discussion from the Brookings meeting, as well as input from FDA and from other industry associations. The final Wave 1 Pilot metrics chosen were selected to measure objective quality performance of a site. They include all the metrics identified by FDA at the Brookings meeting, two technology-specific metrics and two surveys, one on quality culture and one on use of process capability. More discussion on choice of metrics and their associated definitions is given in Section 3.

3 Pilot Design

3.1 Project Governance Model

To manage the Wave 1 Pilot project, ISPE and McKinsey established a project governance model, which is shown diagrammatically in Figure 2.

Figure 2: ISPE and McKinsey Project Governance Model



Key features of this project governance model are:

- Data from individual companies are seen only by McKinsey personnel
- ISPE project team has access only to aggregated data across all companies or to subsets of companies where numbers are sufficient to maintain anonymity
- The ISPE Quality Metrics Core Team, representing a broad spectrum of the pharmaceutical business, meet regularly—usually weekly.
- The ISPE Sponsor Team consists of ISPE's president and CEO, senior leaders of pharmaceutical companies in ISPE's International Leadership Forum and a McKinsey partner.
- The ISPE Core Team seeks communication, guidance and decisions from the ISPE Sponsor Team at about approximately monthly intervals.
- Subgroups of the ISPE Core Team held regular teleconferences with participant company leads and site leads to:
  - Brief them on progress
  - Provide an overview of the data analysis for their review
  - Seek their input
- Subgroups of the ISPE Core Team held informal meetings with the FDA Quality Metrics Program leader to:
  - Seek input to choice of metrics and overall design of the Wave 1 Pilot
  - Provide update on progress of the Wave 1 Pilot
  - Provide early readouts of summary Wave 1 Pilot results
  - Be present to share findings

In addition, the Core Team tasked subteams with charters and deliverables to progress particular elements of the project independently. Subteams were established for:

- Definitions
- Communications
- Influencing and industry engagement
- Quality culture
- Process capability

The importance of defining metrics carefully was identified early in the project. The Definitions Subteam developed the work described in Section 3.3. It was especially key to:

- Developing responses to frequently asked questions (FAQs)
- Producing the definitions given in Appendix 1
- Leading the process to develop the surveys given in Appendix 2.

The Influencing and Industry Engagement Subteam's role was to encourage companies to enroll for the Wave 1 Pilot and to arrange teleconferences with pilot lead individuals and senior leaders in participant companies. Additionally, this subteam took the lead in preparing material for presentation at the FDA meetings.

The Quality Culture Subteam explored new ideas and potential leading quality metrics. Given the findings from the Core Team and the level of public interest expressed at many meetings, this subteam focused on sharing current quality culture best practices. The team initially planned to explore whether a quantitative Quality Culture Index could be established. Subsequent work, however, including discussion and engagement across industry and academia, has suggested that quality culture evaluation requires a holistic approach, and centralized reporting of a standardized assessment is not desirable. The Quality Culture Subteam is developing a cultural excellence framework entitled *The Six Dimensions of Quality Culture*; future publications are also planned.

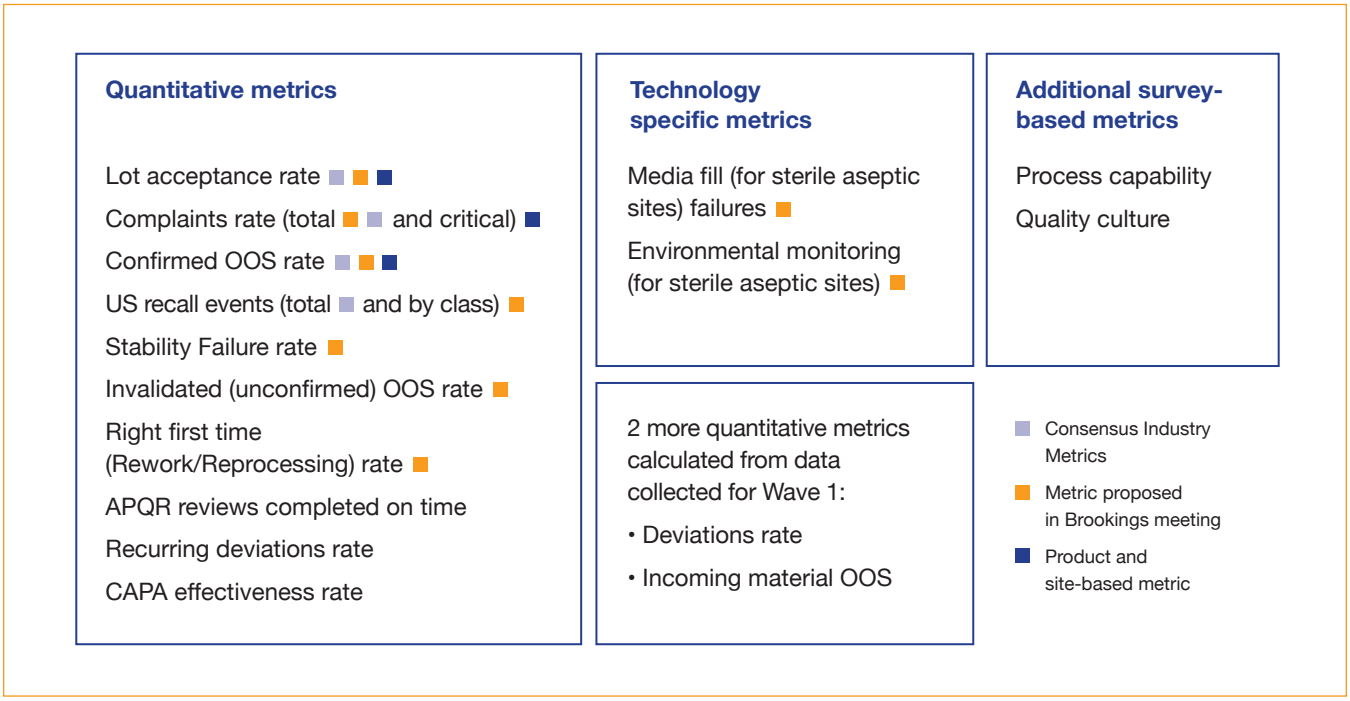
The Process Capability Subteam, working under the wider PQLI umbrella, was established based on a recommendation from the Quality Metrics Core Team. Its objectives are to produce a series of articles and/or white papers, as well as case studies, a potential baseline guide and industry sessions at ISPE meetings to examine the use of process capability measurements by the pharmaceutical industry globally. This team also contributed to the process capability survey questions conducted as part of the Pilot Wave 1.

3.2 Choice of Metrics for the Wave 1 Pilot

This section outlines the list of metrics used in the ISPE Quality Metrics Pilot Program, *Wave 1* and their associated definitions.

The choice of metrics for the Wave 1 Pilot evolved over time by taking account of many influences and input. Output from the Brookings meeting was considered by the ISPE project team, as well as FDA's request for product-based metrics. There was also a strong desire by most parties to start to understand the impact of quality culture. Additionally two technology-specific metrics for sterile product manufacture were included. A summary of the list of metrics used in the ISPE Quality Metrics Pilot Program, *Wave 1* and their origins is shown in Figure 3.

Figure 3: Summary of Final Metrics Collected During ISPE Industry Wave 1 Pilot



3.2.1 A Note on Product-Based Metrics

The following metrics were collected on both site and product bases to help gain an understanding of the differences between these approaches:

- Lot acceptance rate
- Total and critical complaints rate
- Confirmed out-of-specification (OOS) rate

Product-based metrics in the Wave 1 Pilot were collected on relevant unit operations performed at a specific site. For this pilot, metrics were not aggregated across multiple sites to the final packaged and labeled dosage form level or to a new drug application (NDA) level.

When reviewing product-level metrics reporting, it's important to ensure that definitions and expectations are clearly defined and understood. US NDAs, for example, can include multiple dosage form strengths, and each strength may be assembled into a series of packs. This means that one NDA may include many "products" – if a "product" is defined as one dosage form strength assembled into one pack.

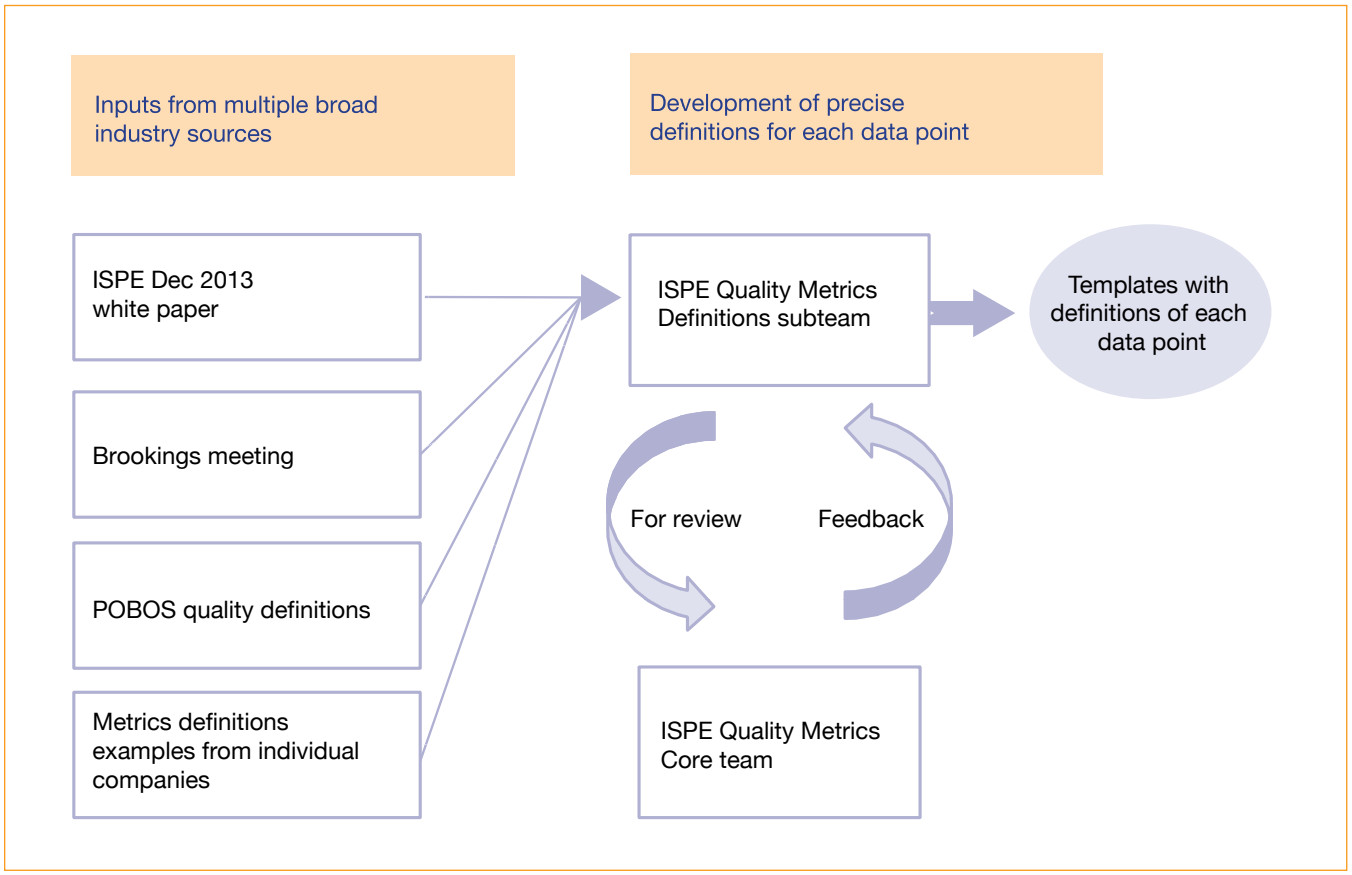
3.3 Achieving a Standardized Definitions Set

Experience from the project team, feedback at ISPE public meetings and preparation of the white paper all identified the importance of defining metrics that:

- Are clear to the project team.
- Are understood by participants.
- Match those currently used by companies as closely as possible.
- Measure quality performance accurately.
- Reduce the opportunity for "gaming".
- Minimize unintended consequences.

The Definitions Subteam established a thorough and robust approach to derive the definitions used in the Wave 1 Pilot, using an iterative process to reach consensus. This process is depicted in Figure 4.

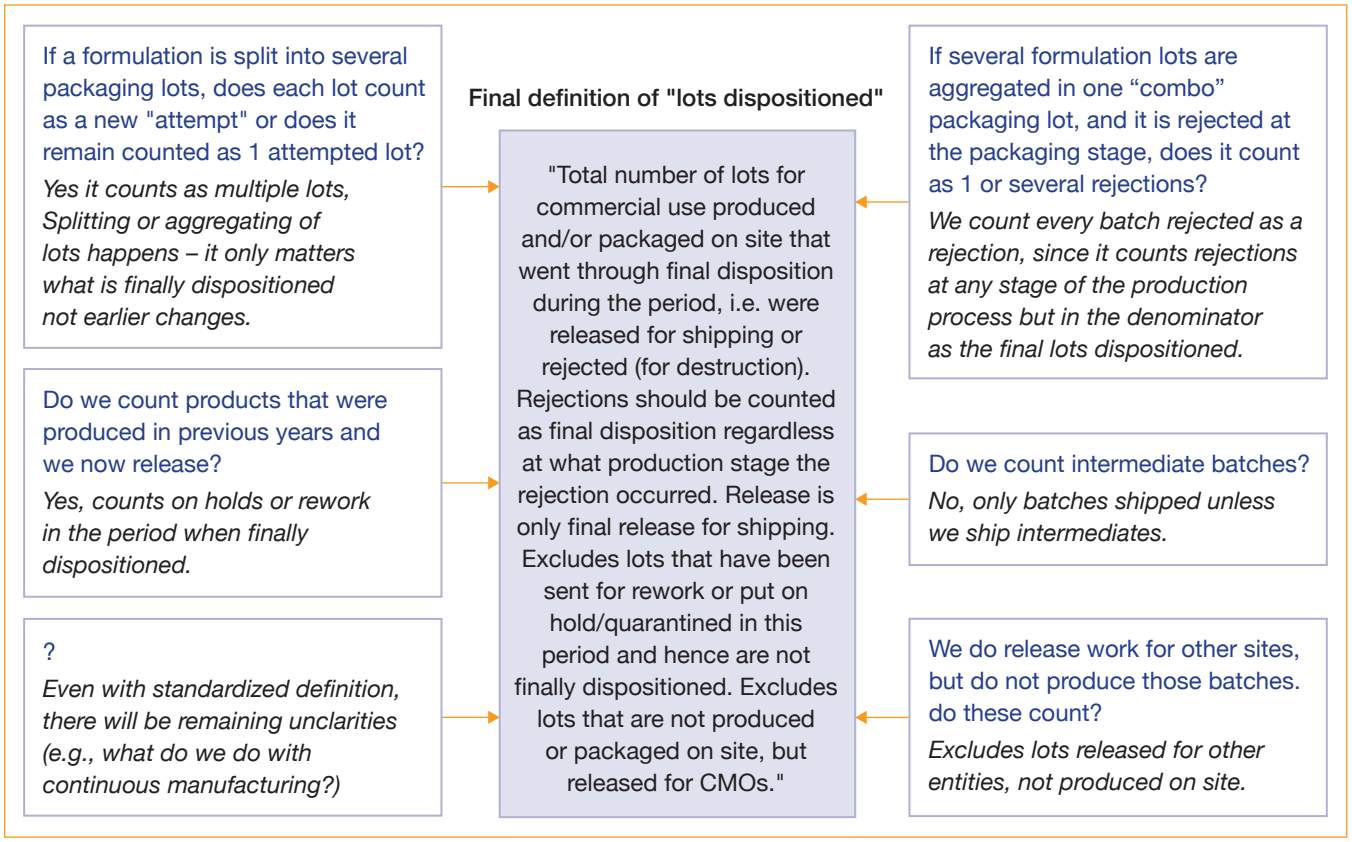
Figure 4: Process to Derive Definitions for Wave 1 Pilot



The Definitions Subteam received inputs from multiple sources. Outputs were collated into a set of agreed-upon definitions in the Excel data-collection templates produced by the McKinsey team, then given to participating companies for completion.

One example of this complex series of interactions can be seen in the definition of “lot dispositioned.” This is a critical term, which required consensus, as it is the denominator for several metrics collected in the pilot. A representation of this consensus building process is given in Figure 5.

Figure 5: An Example of Challenges Defining “Lot Dispositioned”



In addition to developing and designing definitions, the Definitions Subteam also assessed any questions raised by participants during the pilot kickoff or in the early phases of completing the data-collection templates. The subteam issued a weekly list of clarifications in the format of a Frequently Asked Questions document (FAQs) to all participants.

While analyzing data in the Wave 1 Pilot, it became apparent that two additional metrics – deviations rate and incoming material OOS rate – could also be calculated automatically, since the data required for these metrics was already being collected. This brought the number of metrics analyzed to 14.

Definitions for all quantitative metrics used in the Wave 1 Pilot are given in Appendix 1 of this report.

3.4 Additional Surveys Included in the Wave 1 Pilot

Two qualitative surveys were also conducted as part of the Wave 1 Pilot. These explored the prevailing quality culture at the site and examined the use of process capability monitoring and trending on the site.

3.4.1 Quality Culture Survey Development

Using McKinsey's previous experience in conducting quality culture–type surveys, the Definitions and the Quality Culture Subteams reviewed an existing POBOS Quality Culture Shop Floor Survey for inclusion in the Wave 1 Pilot. Using this survey, the subteams developed a 15-question assessment tool that measured five cultural elements: Leadership, Governance, Integrity, Mindset and Capabilities.

Although completing this quality culture survey could be a substantial amount of work for participants, it was considered a necessary tool to test the hypothesis of how quality culture may impact the quality performance outcomes at a given site.

3.4.2 Process Capability Survey Development

Previous work undertaken by the ISPE Quality Metrics Core Team and additional review with industry colleagues indicated two things:

- Application of process capability measures was not widespread in industry
- Sites that did measure process capability used a wide variety of approaches

It was decided, therefore, to develop a survey to assess the tools and processes used to monitor process capability by the participating sites.

Both the Quality Culture Survey and the Process Capability Survey are included in Appendix 2 of this report.

3.5 Estimates of Data Collection and Submission Effort

Companies were asked to estimate the person-hours of effort that they used to set up and carry out their part of the Wave 1 Pilot. Each site completed a template with estimates of the time and effort spent collecting each individual metric:

- Time (hours) spent to collect individual metric data at both site and product levels, for both the retrospective and the prospective periods.
- Degree of difficulty on a scale of 1 to 4 (easiest to most difficult) for collecting each metric at both site and at product levels.
- Site ratings of whether the data was available in the requested form or required recalculation/aggregation, or was collected from fragmented sources.

Companies did not indicate how much effort was required to complete survey questions. The Quality Culture Survey, estimated to take approximately five minutes per respondent, was completed by more than 10,000 staff in Wave 1. European surveys required approval by union representatives; this was also not included in the estimation of effort.

In addition, Wave 1 Pilot data collection and submission sites were allowed to provide "good enough" data. This means that they may not have conducted all the checking and approval steps that would otherwise be required for formal submission to FDA.

McKinsey estimated the operational effort required to set up and support the Wave 1 Pilot, including:

- Preparing submission templates
- Establishing the database
- Defining the collection process for data submission
- Supporting companies in submitting their data
- Analyzing the data

These estimates did not include the time spent by McKinsey personnel working as part of ISPE's project team, contributing to the pilot design and definitions development, producing the report for ISPE and individual companies' benchmarking reports.

3.6 Data Collection Period

Part of the announcement at the launch of the Wave 1 Pilot included details of the data collection period:

*Companies **to provide data** [emphasis added] for approximately one year (historic) and 3-months "real-time," but individual flexibility possible to accommodate data availability.*

A primary goal of the Wave 1 Pilot was to have findings accrued by the end of 2014 so that they could be available either before FDA issued its Federal Register quality metric notification or for consideration during the public comment period. Given this tight timeline, and to ensure meaningful data was collected in short order, it was decided to collect data using two data periods – retrospectively for 12 months for certain metrics where company data already existed, and prospectively for 3 months for metrics that may not have been previously collected or measured at a site. Templates were designed to collect data. An example of site data frequency and collection period is given in Table 1, showing a nonsterile finished dosage form as an example.

Table 1: An Example of Frequency and Period of Collecting Retrospective and Current (Prospective) Data in the Pilot

Data Points/Metric		Retrospective (Nominal)	Current (Nominal)
Baseline Data	Production volume in units	Monthly for 12 months	Monthly for 3 months
	Packs released		
	Lots dispositioned		
	Lots tested—total		
	Lots tested—stability only		
	Site head count	Annual	3 Months
	Site quality head count		
	Number of products		

Data Points/Metric		Retrospective (Nominal)	Current (Nominal)
Site Data	Rejected lots	Monthly for 12 months	Monthly for 3 months
	Reworked/Reprocessed lots		
	Confirmed OOS—total		
	Confirmed OOS—stability failures only		
	Unconfirmed OOS		
	Total recall events		
	Recall Events—Class I and II		
	Rejected lots		
	Total recalled lots		
	Total complaints		
	Critical complaints		
	Products subject to APQR	Annual	No data collected
	APQR on time		
	Number of CAPAs with effectiveness checks	3 monthly in 4 periods, April 2013 to March 2014	One 3 month period
	Number of effective CAPAs		
	Number of deviations		
	Number of recurring deviations		
Product Data	Total complaints for the product for the reporting year	Annual for individual products	Current period, typically 3 months for individual products
	Total critical complaints for the product for the reporting year		
	Total packs released for the product for the reporting year		
	Total lots dispositioned for reporting year		
	Total lots tested for reporting year		
	Rejected lots		
	Confirmed OOS		

4 Operational Processes for the Wave 1 Pilot

This section discusses some of the key operational aspects of the Wave 1 Pilot for both the McKinsey support team and the participant companies.

4.1 McKinsey Operational Process

The McKinsey support processes for performing the work associated with the Wave 1 Pilot were based on their experience gained from their POBOS benchmarking programs.

An overview of the process is as follows:

- Preparing templates for data submission:
  - Different templates were required for drug products, sterile and nonsterile drug products, and for labs.
  - Templates for each metric included detailed definitions, fields for each data point time period – monthly, quarterly or annual – and a commentary field.
  - Templates were validated using built-in checks for data consistency (e.g., total OOS by product = total OOS for site) and locked before they were sent to sites.
- Setting up databases for input and analysis.
- Defining the data-collection process.
- Translating the quality culture survey into the appropriate language for each participating site.
- Answering exploratory questions from interested companies, such as:
  - How much time and effort and resources will we need for the pilot?
  - How much does it cost?
  - What is involved?
- Enrolling companies into the Wave 1 Pilot by arranging:
  - Confidentiality agreements
  - Purchasing orders
  - Explaining data submission requirements
- Answering questions during the data-collection phase and updating the FAQs document.
- Reviewing and clarifying data (i.e., potential outliers).
- Processing the Quality Culture and Process Capability Surveys.
- Analyzing data, running correlations, profiling metrics.
- Reporting results of the Wave 1 Pilot data analysis.

Using agreed-upon definitions and survey questions, project timelines and data-collection frequencies, McKinsey prepared the templates in Excel. An example is provided in Appendix 3.

For the Wave 1 Pilot, companies completed the Excel spreadsheets manually and sent them to the McKinsey support team. An automated data-entry process may be considered for the future.

A data lock was applied at the end of November 2014, and all participating companies complied.

4.2 Experiences from Participating Companies

Companies participating in the Wave 1 Pilot provided feedback throughout the data collection and analysis phases; Pilot Leads meetings held by the Industry Engagement Subteam also provided feedback.

The majority of companies that participated in the Wave 1 Pilot reported benefits arising from their involvement. These included:

- The opportunity to trial a centralized submissions process gave them a deeper understanding of the impact of standardized metrics definitions and design.
- Participation enhanced the maturity of their internal metrics programs.
- Each participating site received a confidential benchmarking report that outlined their performance with respect to their peer group(s).

With respect to reporting "good enough" data, some participants noted that some data they collected were derived from non-GMP systems (e.g., product portfolios). Discussions have indicated the need for a validated/cGMP-based metrics collection, storage and reporting system that could be reviewed by inspection teams to confirm the veracity of any metrics reported to FDA. Concerns raised about the potential burden associated with this will require further consideration.

A detailed example of some of the key aspects of one company's experience of participating in the Wave 1 Pilot is provided in a case study in Appendix 4. A summary of the case study's main points are:

- Company A has a large product range and very complex supply chains, which make assigning product-level metrics extremely difficult and time-consuming.
- Changing their current IT systems to a standardized set of metrics that could produce product-level data would require significant investment.
- Data reported into the Wave 1 Pilot were "good enough" to examine the data-collection and -submission systems', mechanics, but they were not subjected to the review and checking that would be required for official submission to FDA.

5 Findings from the ISPE Quality Metrics Initiative Wave 1 Pilot

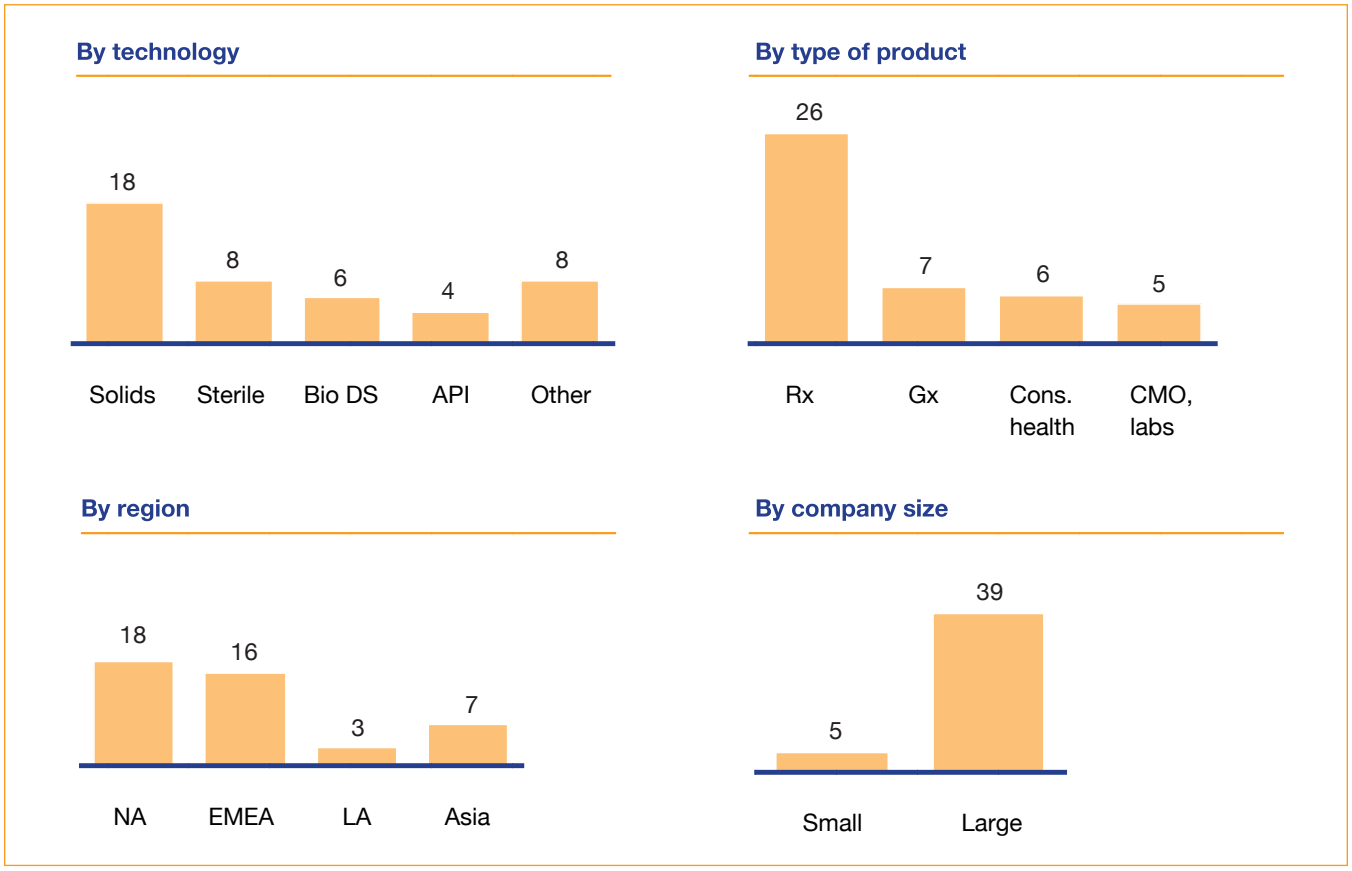
This section discusses the main findings from the ISPE Quality Metrics Wave 1 Pilot and includes considerations of the sample size, metric and survey data analysis, collection and submission effort data and the key relationships observed.

5.1 Sample Size

The Wave 1 Pilot collected data from 18 participating companies at 44 individual sites. Distribution of the sites by technology, type of product/business, region and company size is given in Figure 6 and Table 2.

Figure 6: Sample Distribution of Participating Sites

Diverse Sample: 18 Participating Companies with 44 Sites/Technologies



Note: If a site has more than one technology we count the number of separate templates they will fill, usually one per technology

Table 2: Figure 6 Abbreviations

Technology	
Bio DS	Biopharmaceutical or biological drug substance site
API	Small-molecule drug substance (active pharmaceutical ingredient)
Type of product	
Rx	Originator company
Gx	Generic company
Cons. health	Consumer health or OTC
CMO	Contract manufacturing organization
Labs	Contract research and testing laboratories
Region	
NA	North America
EMEA	Europe, Middle East and Africa
LA	Latin America
Company size	
Small	< $1 billion in revenues
Large	> $1 billion in revenues

The Quality Culture Survey sample size comprised 10,300 respondents from 37 participating sites. This differs from the total pilot sample size of 44 sites because some sites had two different product technologies located on the same physical site. These sites completed two Wave 1 Pilot metrics templates as two separate sites, yet submitted their quality culture survey assessments as one site.

The Wave 1 Pilot sample was considered suitably diverse, both by region and technology, to facilitate a representative analysis. Other aspects of the sample, however, had similarities that should be acknowledged:

- Most participant companies originated from developed countries, and therefore did not have any significant language or interpretation issues with the standardized definitions or the pilot template submission instructions.
- All enrolled sites may be considered in good standing with respect to quality; they did not have any quality or compliance (e.g., consent) issues.
- The majority of companies were classified as large.
- All companies and their sites consented to enroll in the Wave 1 Pilot and, therefore, had an open and positive disposition to the use of quality metrics to monitor and drive enhanced quality performance.

5.2 Metric and Survey Data Analysis and Discussion

To assist with data analysis and presentation, metrics collected in the pilot were grouped into different categories shown in Table 3. These include:

- **External Quality Outcomes:** An outcome that may affect the patient directly (e.g., the patient makes the complaint) or indirectly (e.g., a recall leads to product unavailability).
- **Internal Quality Outcomes:** An outcome observed by a company that could affect business output (e.g., a rejected product), product does not leave the company control, however, and is not available to the patient.
- **Supplier Quality:** Confirmed OOS of an incoming raw material is a measure of supplier quality.
- **Laboratory Quality:** An unconfirmed OOS is a measure of laboratory quality, whether or not the cause is identified.
- **Site Maturity:** Metrics that could be considered measures of site maturity, such as annual product quality reviews (APQRs) completed on time, high values of corrective and preventive action (CAPA) effectiveness rate and low values of recurring deviations rate.

Table 3: Categories of Metrics

Categories	Metrics
External Quality Outcomes (market)	Total recall events—US
	Recall events Class I and II—US
	Recalled lots—US
	Total complaints rate
	Critical complaints rate
Internal Quality Outcomes	Lot acceptance rate
	Confirmed OOS rate—release
	Confirmed OOS rate—stability
	Deviations rate
	Right first time (rework/reprocessing) rate
	Media fills successful
	Environmental monitoring (EM) action limit investigations rate
	EM action limit rejects rate
Supplier Quality	Confirmed OOS rate—incoming materials
Laboratory Quality	Unconfirmed OOS rate
Site Maturity	APQR completed on time
	CAPAs effective rate
	Recurring deviations rate

5.3 Wave 1 Pilot Quality Metric Data Analysis

Individual company data was compared with the total sample to develop a benchmarking report for each company. These data were confidential to the company and McKinsey. ISPE did not have access to individual company data.

Total industry-level data for all metrics collected, relationship determinations and comments are provided in Appendix 5 of this report.

Total industry-level data for each metric were tested and evaluated as appropriate, using either scatter plots (correlation of a leading indicator vs. an outcome) or quartiles analysis (range of outcome values against leading indicator values split into quartiles) or profiling (for metrics with discrete values).

Many figures presented in Appendix 5 of this report also contain a summary of the statistical tools applied and include explanation, where necessary.

As per standard statistical practices, incomplete data and extreme outliers were excluded from analyses. The resulting sample sizes allowed statistical analysis with some limitations, as follows:

- To allow for sufficient sample size, most analyses were done for finished dosage sites overall, not by technology.
- Product data was collected on annual bases, and did not allow time-lag analysis to see how product metrics correlate over time.
- Any relationships identified between metrics were statistically significant (less than 5% likelihood of a coincidence), however:
  - The strength of the relationships vary and may be relatively low (e.g., some may correlate with R^2 of 30% or 40%), since these metrics are influenced by multiple factors.
  - Correlation doesn't imply causation. Understanding the underlying factors and direction of a relationship will require further work, ideally on a larger data set from a larger sample size.

From the total industry database, median ranges of individual metrics split by technology are given in Figure 7 and Figure 8 below.

Figure 7: Metric Ranges by Technology: Solid Dosage and Sterile

		Solids			Steriles		
		TQ	Median	BQ	TQ	Median	BQ
External (market) quality outcomes	• Recall events – US	0.0	0.0	1.0	0.0	0.0	0.0
	• Recall events class I and II – US	0.0	0.0	0.5	0.0	0.0	0.0
	• Recalled lots – US	0.0	0.0	1.5	0.0	0.0	0.0
	• Complaints rate (per million packs)	6	16	26	24	55	76
	• Critical complaints rate (per million packs)	0.2	0.4	1.2	0	0.1	1.7
Internal quality outcomes	• Lot acceptance rate (% released out of all finally dispositioned lots)	99.7%	99.4%	98.8%	99.6%	98.8%	96.1%
	• Confirmed OOS – release (per 000' lots release-tested)	0.8	1.5	2.9	0.5	1.0	8.3
	• Confirmed OOS – stability (per 000' stability lots tested)	0.0	2.5	6.3	0.0	2.3	6.2
	• Deviations rate (per 000' lots dispositioned)	94	134	230	257	472	744
	• Rework rate (% lots dispositioned)	0%	0.1%	1.0%	0.0%	0.3%	0.7%
	• Media fills successful (%)				100%	100%	100%
	• EM action limit investigations rate (%)				0.6%	2.12%	8.8%
	• EM action limit rejects rate (%)				0	0	0.02%
Supplier quality	• Confirmed OOS – incoming materials (per 000' RM/PM lots tested)	0.3	1.1	4.9	0.9	2.7	5.4
Lab quality	• Unconfirmed OOS (per 000' lots tested)	1.6	3.2	4.3	0.6	0.3	4.1
Site maturity	• APQR on time (%)	100%	100%	96.2%	100%	100%	90.9%
	• CAPAs effective (%)	98.5%	97.4%	33.4%	96.7%	94.9%	92.6%
	• Recurring deviations (%)	5.2%	12.9%	27.5%	5.0%	9.5%	23.6%

Figure 8: Metric Ranges by Technology: API, Bio and Other

Median

		API	Bio	Other FP
External (market) quality outcomes	• Recall events – US	0.0	0.0	0.0
	• Recall events class I and II – US	0.0	0.0	0.0
	• Recalled lots – US	0.0	0.0	0.0
	• Complaints rate (per million packs)	N/A	N/A	111
	• Critical complaints rate (per million packs)	N/A	N/A	0.9
Internal quality outcomes	• Lot acceptance rate (% released out of all finally dispositioned lots)	100%	95.3%	98.1%
	• Confirmed OOS – release (per 000' lots release-tested)	2.9	25.1	6.1
	• Confirmed OOS – stability (per 000' stability lots tested)	0.9	15.4	0.0
	• Deviations rate (per 000' lots dispositioned)	464	9678	154
	• Rework rate (% lots dispositioned)	1.6%	0.2%	0.6%
	• Media fills successful (%)			
	• EM action limit investigations rate (%)			
	• EM action limit rejects rate (%)			
Supplier quality	• Confirmed OOS – incoming materials (per 000' RM/PM lots tested)	3.6	2.2	3.7
Lab quality	• Unconfirmed OOS (per 000' lots tested)	4.2	1.7	1.4
Site maturity	• APQR on time (%)	100%	87.5%	100%
	• CAPAs effective (%)	95.8%	33.2%	64.9%
	• Recurring deviations (%)	7.5%	13.6%	20.7%

Although a few sites experienced a recall during the Wave 1 Pilot period, the median for recalls in both Figure 7 and Figure 8 was zero. (For further details see "Recall Events" in Appendix 5.)

In addition to the data collected, participating sites provided very useful feedback on the definitions used in the pilot. An example of this includes the extensive feedback received on the definition for "recurring deviations rate," shown in Figure 9.

Figure 9: Feedback on Definition of Recurring Deviations Rate

Pilot definition

Number of deviations (out of the total reported in line 49) for which during the 12 month period preceding each deviation, at least one other deviation has occurred with the same root cause within the same process and/or work area.

Feedback/ alternative definitions

- Period considered for the recurrence may be based on the type of issue or work area (e.g. depends on the occurrence of the specific process with deviation), or left to the quality personnel judgment;
- Some sites use 6 months or 2 years as look back period;
- Deviation may be considered recurring if reoccurred anywhere in the plant, not just in the same work area
- At least one site considers recurring only deviations that have had a CAPA (as recurrence indicates ineffective CAPA implementation (other deviations with same root cause are considered "repeat"))
- Many sites feel that final assessment depends how deep you go into the root cause – from a more general "operator error" to a very specific error description (has to be the same product, nature of incident, root cause category) – which would influence how recurrence is identified;

Recurring deviations rate was not the only metric that produced lots of questions regarding definition and how the data should be calculated and submitted. Other definitions in need of refinement and further alignment are critical complaints, rework rate and CAPA effectiveness rate.

5.4 Wave 1 Pilot Quality Culture Survey Data Analysis

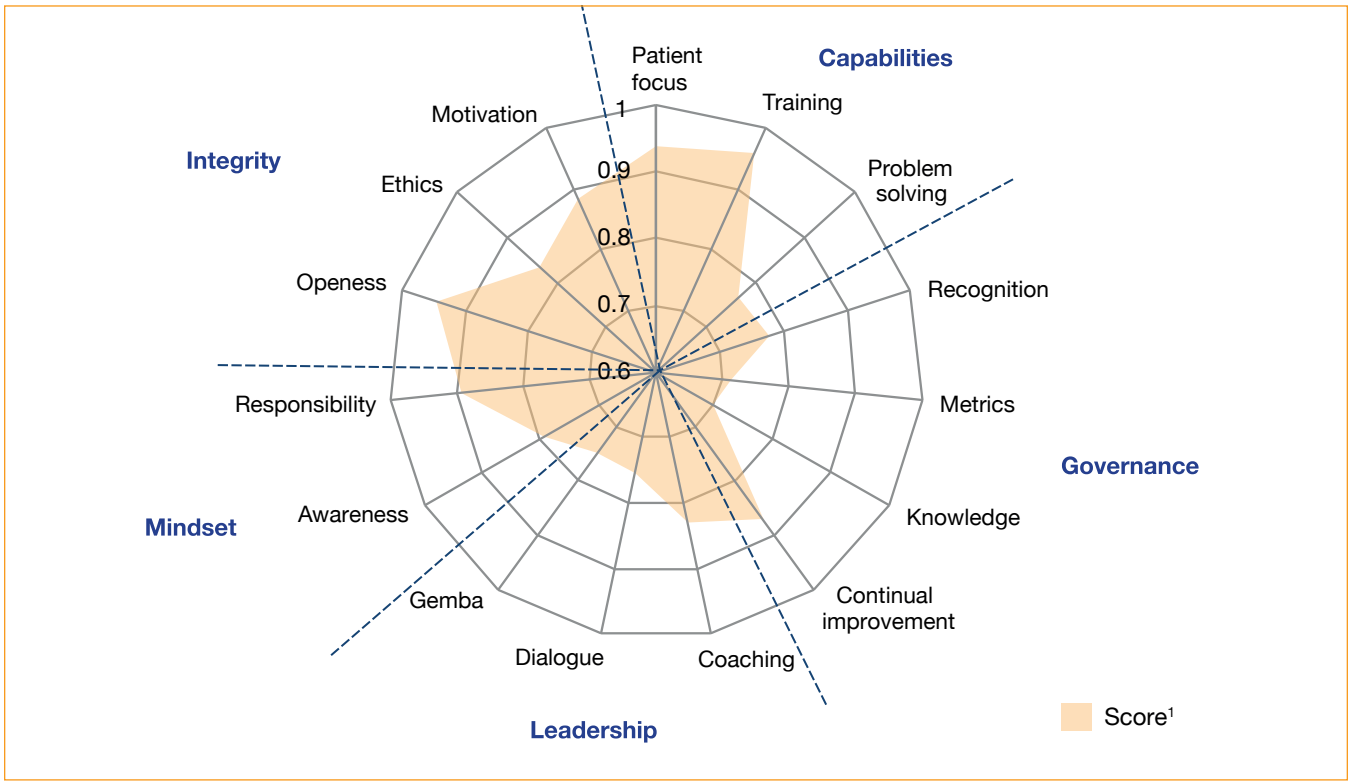
For the quality culture survey, each of the 15 questions could be answered using one of five response options:

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

To facilitate data analysis and relationship mapping with the quality metrics data set, scoring mechanism was established 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." These proportional values were then plotted on a radar diagram for each question, as shown in Figure 10.

Figure 10: Quantitative Quality Culture Scores Plotted on a Radar Diagram

Total of 10,300 respondents from 37 sites

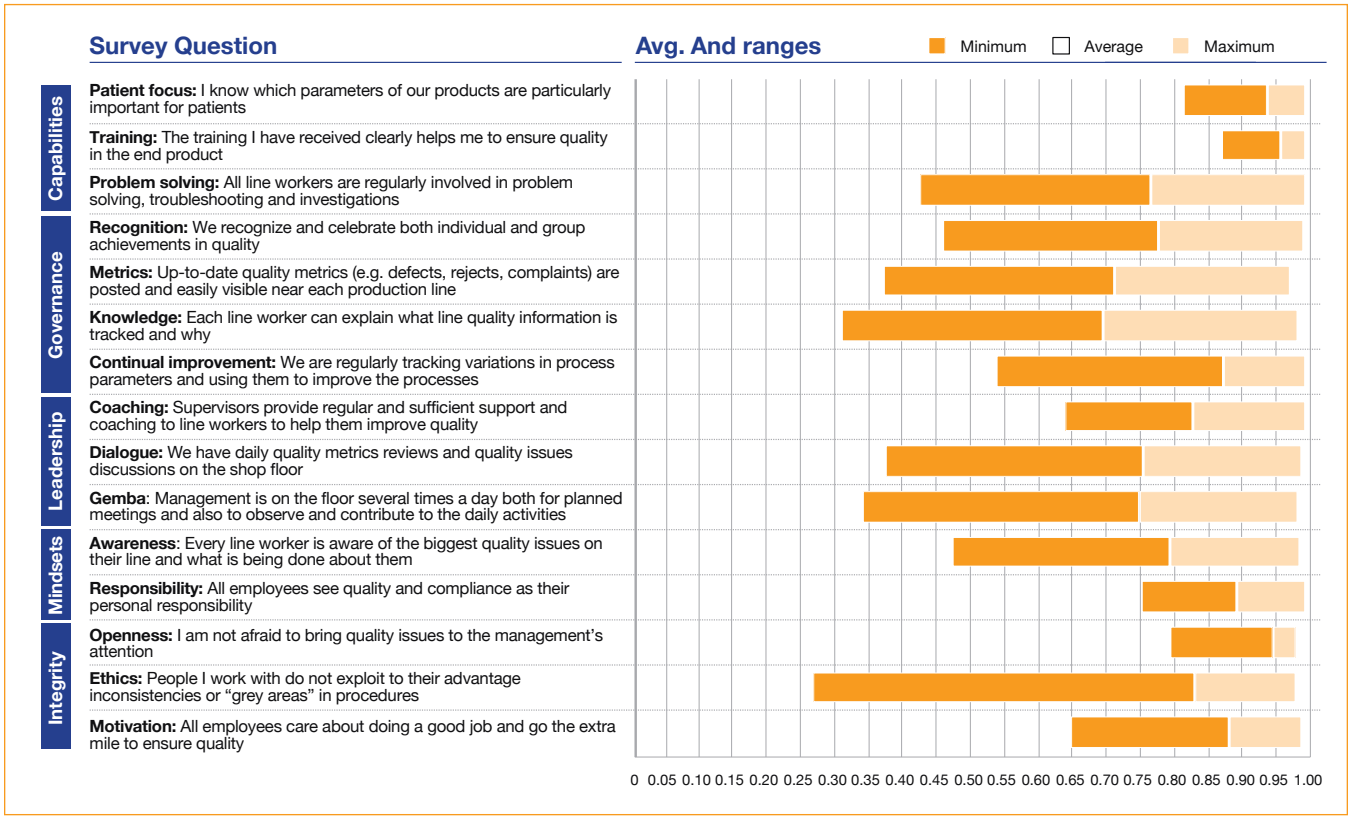


[1] Total score calculated as "top boxes" (share of "agree" and "strongly agree" responses) ratio. 100% = all respondents agree or strongly agree, 0% = nobody agrees or strongly agrees.

The results showed that at the site level, questions pertaining to the cultural elements of Capabilities and Integrity received the highest ratings, while those associated with Governance and Leadership scored relatively weaker. The top-boxes approach facilitated a quantitative analysis of the quality culture findings and provided some interesting insights, but it is broadly agreed that further work is required, potentially in a Wave 2 Pilot, to understand these findings better.

The range of values assigned to the responses for each quality culture question are given in Figure 11. Orange shading represents the range of site responses that fall below the industry average, while light orange shading indicates the range of site responses exceeding the industry average.

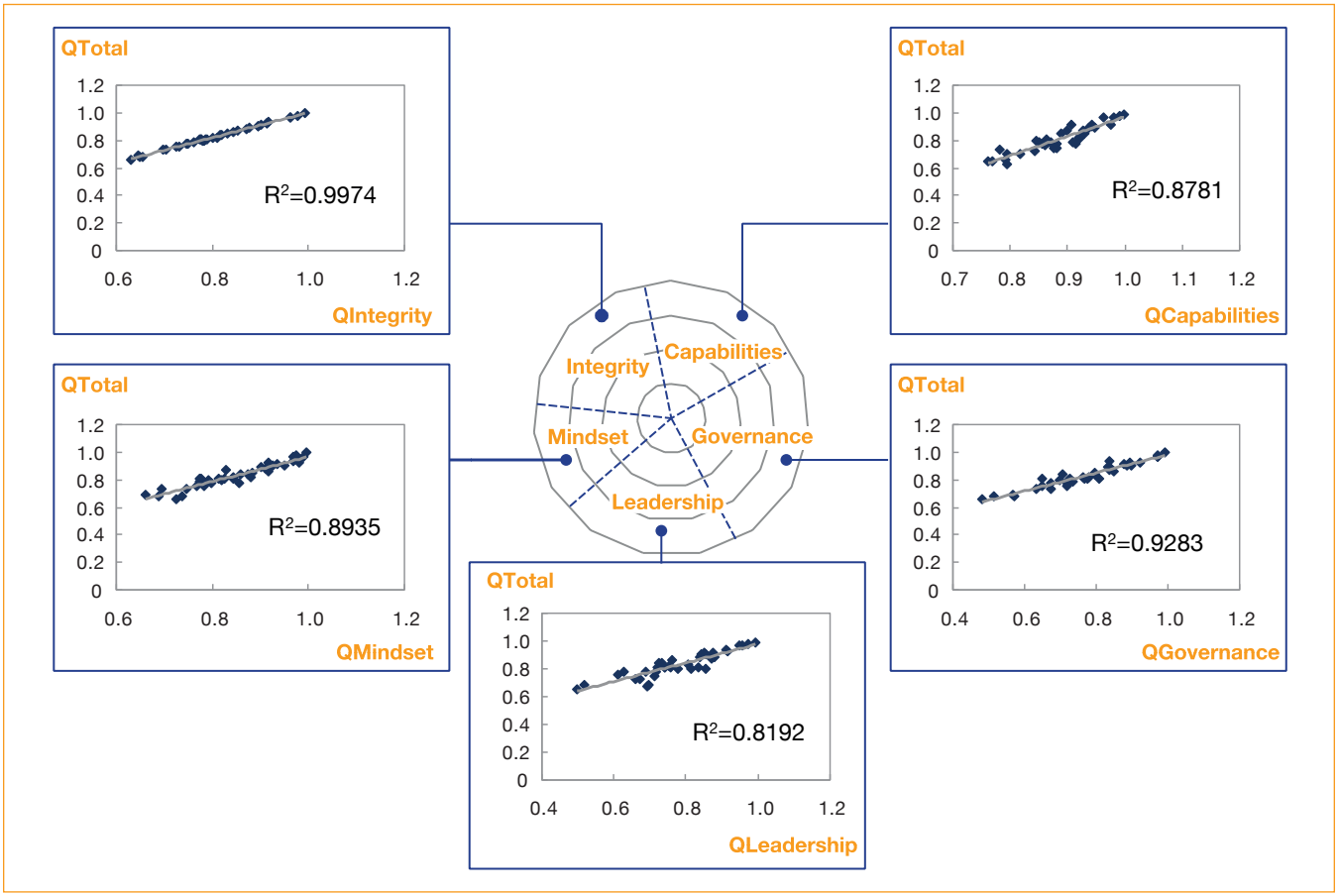
Figure 11: Range of Values for Quality Culture Responses



Responses to questions on ethics, knowledge (e.g., metrics tracking), and Gemba (Japanese for "at the site," it refers observations of a process in action, or management presence on the shop floor) had the highest variation in responses.

To explore whether there are differences in responses to questions within each of the cultural dimension examined, the values for each dimension were plotted against values of total quality culture responses. These plots are given in Figure 12.

Figure 12: Plot of Total Quality Culture Values for Each Quality Culture Dimension



[1] Total score calculated as "top boxes" (share of "agree" and "strongly agree" responses) ratio

This analysis shows that the cultural dimensions included in the quality culture survey are highly consistent between each other—i.e., the line slopes are similar and each dimension is highly correlated with the overall value.

This analysis confirmed that attempts to determine relationships between quality culture and other quality metrics values can use total quality culture values.

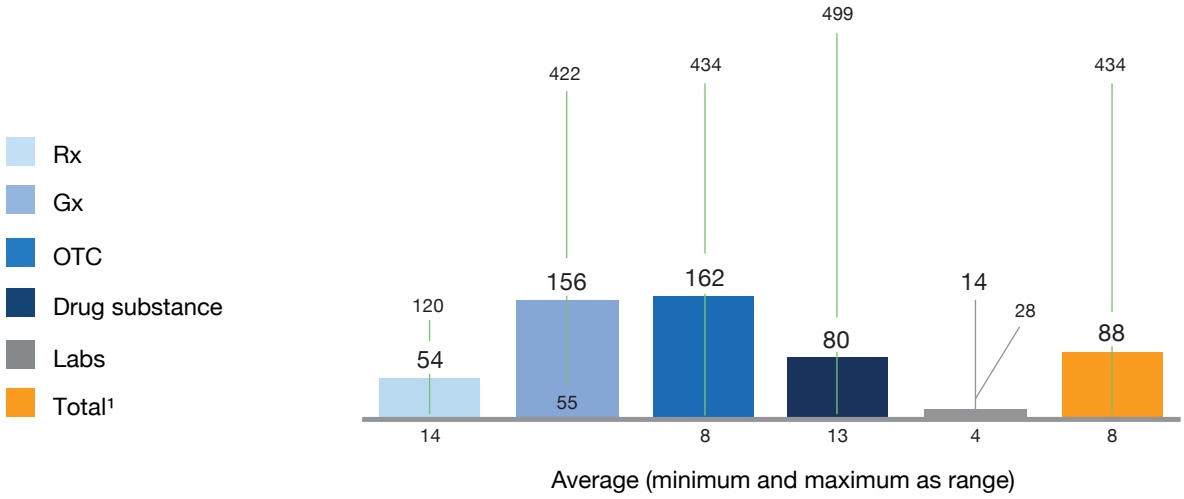
5.5 Data Collection and Submission Effort Data

Estimates of effort expended by companies as an industry total were provided and the results of this analysis are provided in Appendix 5 for each individual metrics.

The average time on each site to collect the annual data was 88 hours. Effort estimates (hours) split between the various types of business are also given in Figure 13 below.

Figure 13: Average Time to Collect Annual Quality Metric Data

Total time spent on collecting data (12 months), [Hours]



[1] Excluding DS and Labs

The results show that OTC and generic sites took three times longer than originator companies to collect similar data. Participants seemed to indicate this difference could be due to issues such as increased number of products and (potentially) the increased complexity of supply chain for OTC and generic sites. Site volumes (packs or lots dispositioned) and site complexity (number of products) did not appear to influence the collection and submission effort required.

Drug substance and laboratory sites required significantly less time than finished product sites because less data is collected. Even though the Wave 1 Pilot had limited visibility in such sites, their workload was calculated at 45 hours collection and reporting per site, the average workload for both lab and drug substance sites.

Approximately 12,000 sites globally have Federal Establishment Identifier (FEI) numbers; [17] close to 6,000 are registered as Finished Dosage and API sites, and the remaining 6,000 sites include medical gases, medical feed, labs, Center for Biologics Evaluation and Research (CBER) establishments, and others. Wave 1 Pilot data estimated the average workload for the API and Finished Dosage sites at 90 hours, while the average workload for the other sites is estimated at 45 hours. Using a typical labor cost (including overhead) of $75,000 per year, collecting this amount of data would cost the industry approximately an **additional $35 million** annually.

This estimate is considered conservative, however, because it does not include several factors, such as:

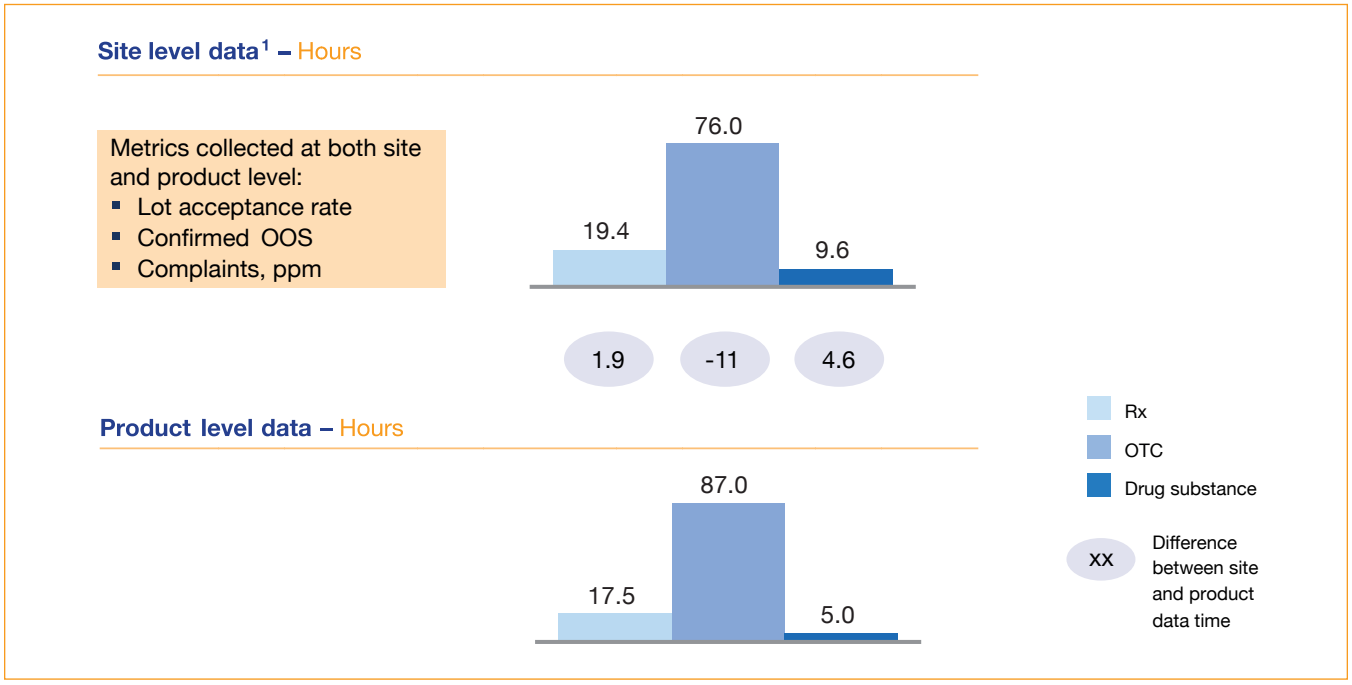
- Wave 1 Pilot sites were allowed to provide "good enough" data. Submission to FDA could require more thorough and complete data collection, additional review and data verification steps, potentially at different levels and disciplines and would have to be accompanied by considered comments.
- Time for internal discussions, management review and above-site guidance was not included.
- Effort to develop and validate new/modified IT systems was not included. (This was not required for the Wave 1 Pilot.)
- Participants had flexibility to provide most pragmatic data set (e.g., for all products at site or only those for the US market).
- Data were provided within each site, not through full product supply chain.
- Participants had mature systems and capabilities.
- Majority of sites were from developed countries.

The additional cost to produce official submissions could bring the annual cost of such a program to **$100+ million**.

The time spent to collect the 3 months of data for the three metrics collected at both product and site bases (lot acceptance rate, confirmed OOS and complaints) are given in Figure 14.

Figure 14: Time Spent To Collect Product Level Data and Site Level Data for 3 Months

Time spent on collecting site[1] and product level data, [Hours, (15 months)]



Note: Sites that didn't submit full product data were excluded as effort likely to be understated. Gx sample was too small to report results

[1] Only for metrics with product level granularity

No data is included in the analysis shown in Figure 14 for Generic companies due to insufficient sample size. Also sites that did not submit full product data were excluded since they were likely to underestimate the effort required.

The findings from the analysis shown in Figure 14 are:

- Rx sites were able to collect these metrics at product level as easily as at site level (there was an approximately 10% difference in time for the three metrics) due to several factors.
- Some OTC sites found product-level data for these metrics more difficult to collect than site data (15% more time for the three metrics).

Potential reasons why Rx sites were able to collect product level data as easily as site level data include:

- These three metrics were chosen intentionally to be collected easily by product level.
- Some sites already had systems set up to collect product-level data.
- For the Wave 1 Pilot, product data was collected within each individual site rather than across a full supply chain.
- Most sites selected for the pilot had good systems and proficient personnel; some even had systems set up to allocate data by product and then aggregate it at site level.
- Definitions allowed accurate allocation by product (e.g., using lots dispositioned rather than lots attempted).

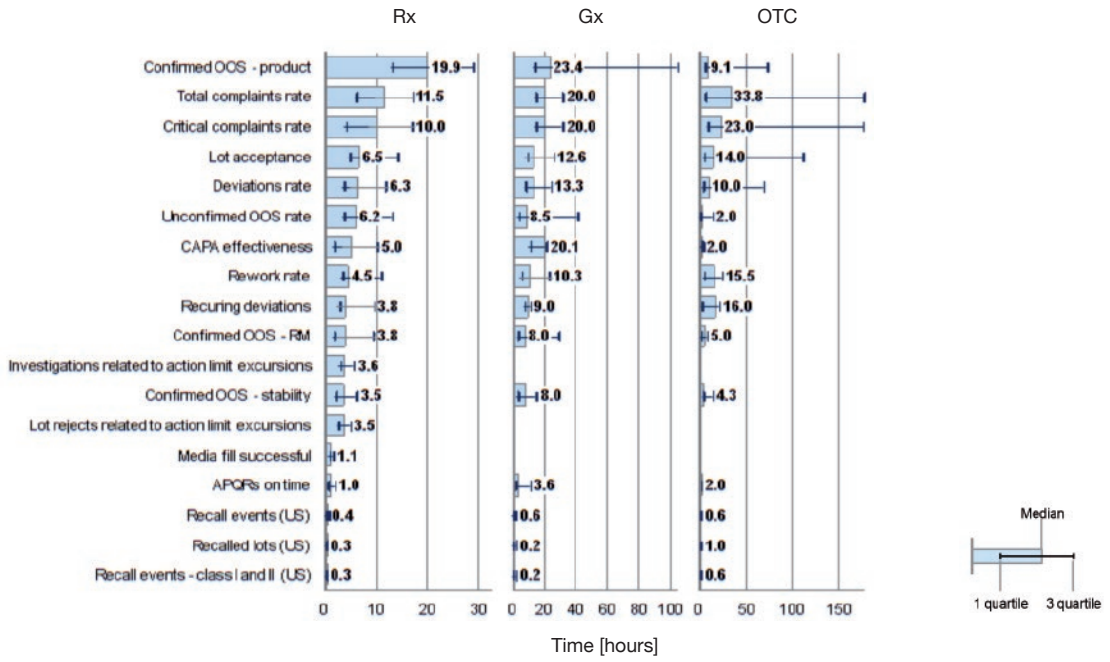
OTC sites spent more effort collecting site level metrics from Figure 13 and Figure 14, and even more effort collecting product level metrics from Figure 14, potentially due to:

- Separating complaints down to individual formulations rather than at a product family (e.g., several flavors or colors) level.
- Counting pack or unit data (vs. cases), and disaggregating data from APR's to match the time frame of the data collection period.

Using the full 15 months of data, the time to collect individual metrics is listed in Figure 15, with columns given for Rx, Gx and OTC companies.

Figure 15: Effort per Metric for Rx, Gx and OTC Companies

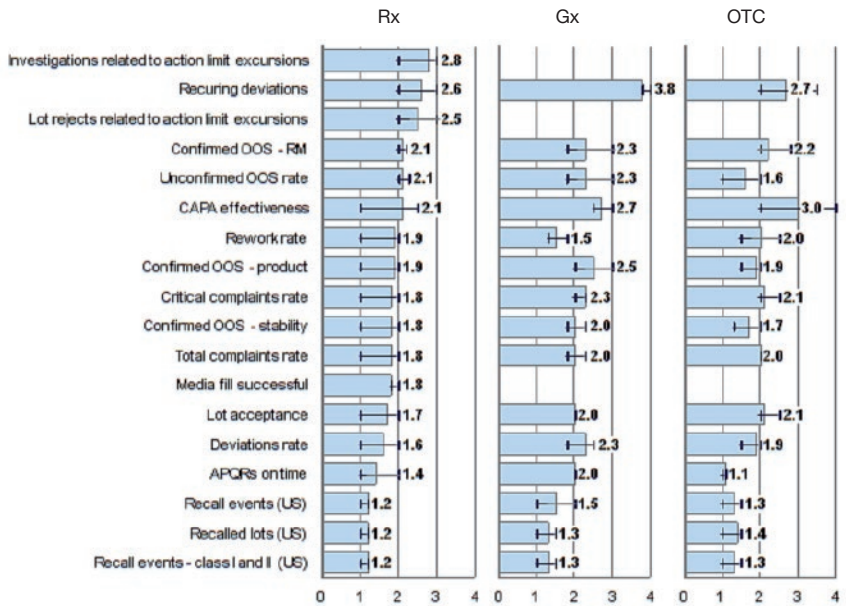
15 months of data



This analysis indicates that the most time-consuming metrics to collect were; OOS for final product and complaints, both total and critical. The least time-consuming metrics to collect were recalls and APQRs on time.

In addition to capturing the number of hours required to collect each metric, Wave 1 Pilot companies were also asked to report the degree of difficulty of collecting a particular metric using a scale of 1 to 4, (1 being easiest and 4 most difficult). This analysis is shown in Figure 16, as estimated by the participating companies themselves.

Figure 16: Degree of Difficulty of Collecting Each Metric

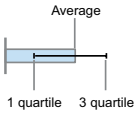


1 - Very easy, all the data we have in an electronic system/paper based documents in the form required here

2 - Fairly easy, most of the data we had ready, some we had to recalculate/aggregate basing on the provided definitions and our raw numbers

3 - Difficult - we had to recalculate/aggregate most of the data, sometimes from very fragmented sources

4 - Very difficult, we did not have the data at all for retrospective period, we only started collecting it for current period

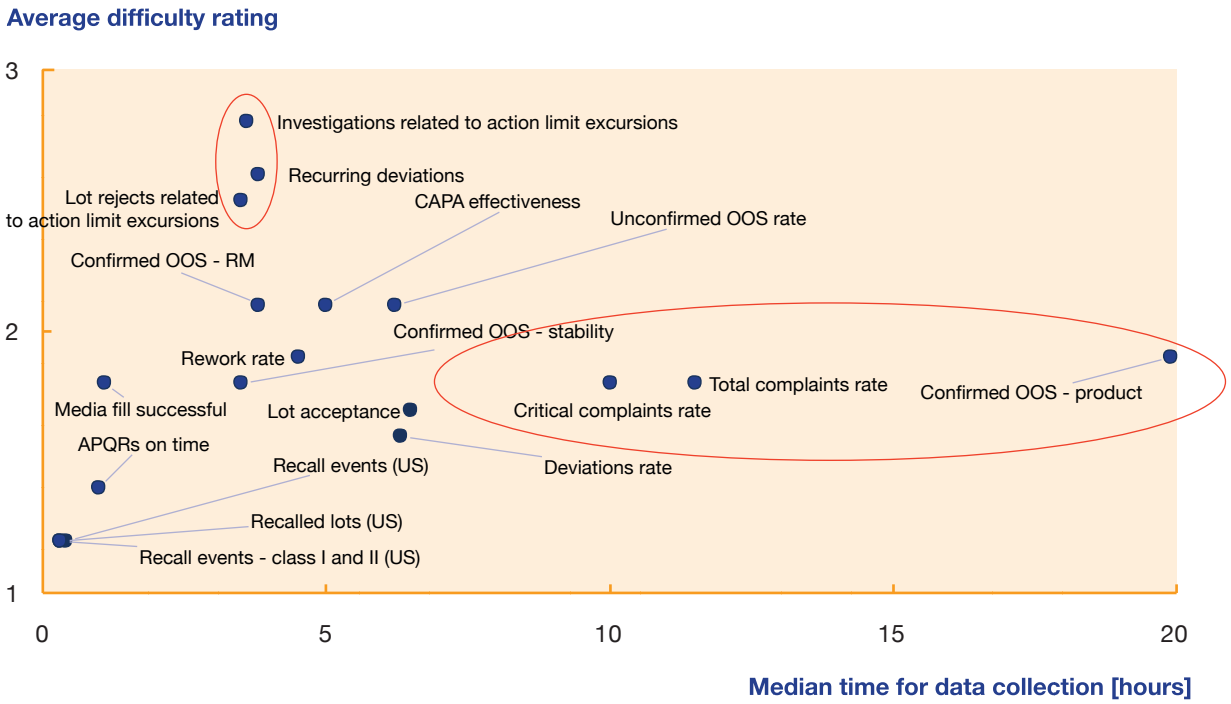


This analysis indicates that the most difficult metrics to collect were the recurring deviations rate and the two sterile specific metrics. The least difficult to collect were once again US recalls and APQRs on time.

Figure 17 shows a further analysis relating to ease of data access and burden to collect, plotting the average difficulty rating against the median time for data collection for Rx companies. This group was chosen as they presented the largest sample size in the Wave 1 Pilot.

Figure 17: Average Time for Collection of a Metric and Median Time for Collections

Finished dosage Rx sample,15 months of data



The results from Figure 17 show that the most time-consuming metrics are not necessarily rated as the most difficult.

The red circles highlight the most difficult to collect metrics:

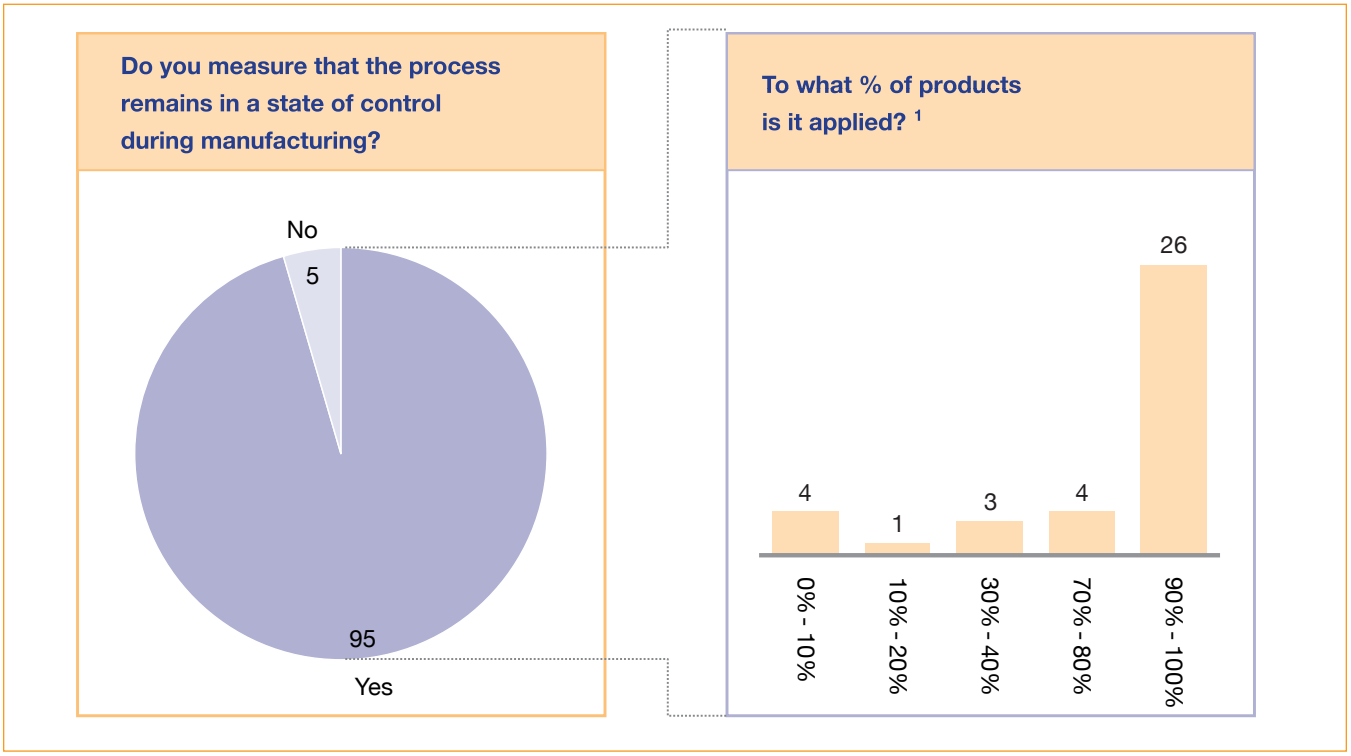
- Investigations and lots rejected related to environmental monitoring for sterile products.
- Recurring deviations: The majority of sites reported that they do not currently have processes in place to collect this metric accurately. Where it is collected there is broad variation in the definition of "recurrence."

As shown in Figure 15, OOS for product, and complaints, total and critical were the most time-consuming metrics to collect.

5.6 Process Capability Survey

The results of the Process Capability Survey (see Appendix 2) are given in Figures 18, 19 and 20. Responses to the first two high-level questions in are given in Figure 18.

Figure 18: Response to Process Capability Questions



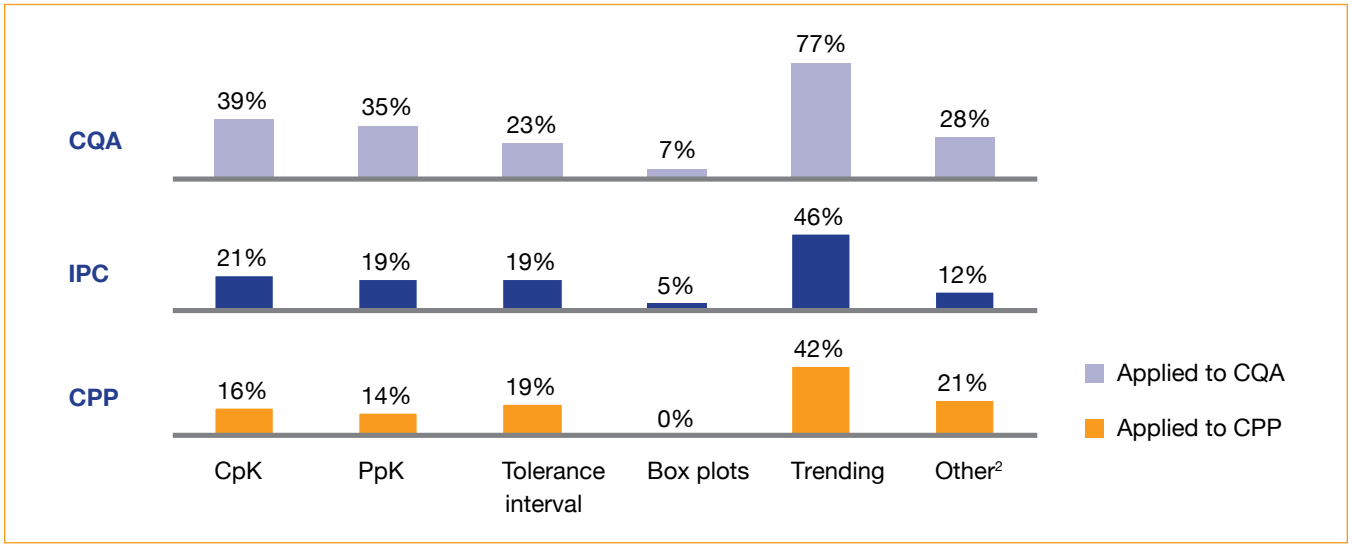
1 When only some products were chosen, choice was based on risk approach to customer and importance for business

From Figure 18 we observe that 95% of sites apply ongoing monitoring during production processes to an average 74% of their products. Where process monitoring was applied to a selected range of products, a risk-based approach was typically used to determine which products required monitoring.

More detailed questions asked which process capability statistical tool was most used and to what attributes [e.g., critical quality attributes (CQAs)] or parameters [e.g., critical process parameters (CPPs) or in-process controls (IPCs)]. Responses to these questions are given in Figure 19 and Figure 20.

Figure 19: Process Capability Survey Findings

Percentage of sites using each type of metric[1]

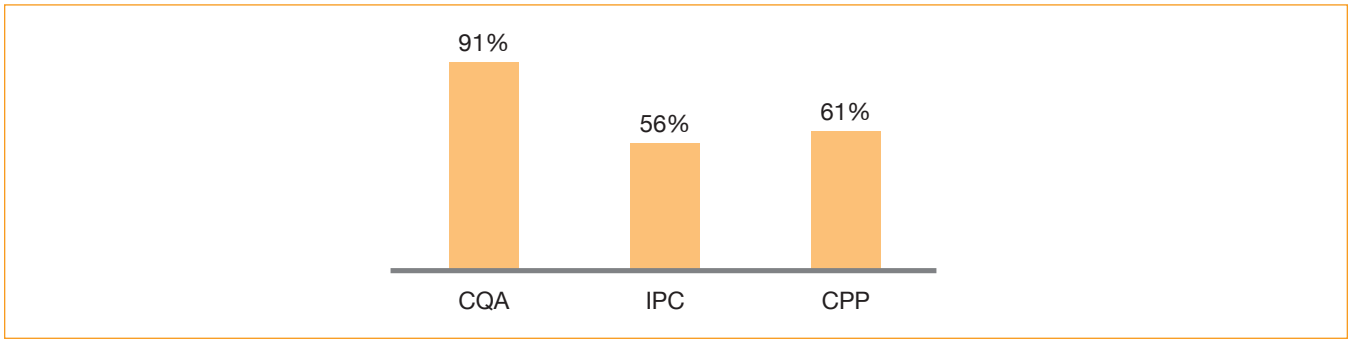


[1] Out of sites monitoring capability in any way
[2] Other mentioned metrics were pareto charts, monitoring via excursions trending, I-charts regression, 3 sigma, and Run/control charts

The results indicate that trending is most the widely used tool. Process capability index (CpK), process performance index (PpK) and tolerance intervals are used less often—by 39% and 22% of sites respectively. While, as seen in Figure 20, 91% of sites measure their current state of control through CQAs, while only 56% on IPCs and 61% on CPPs.

Figure 20: Process Capability Tools in Use

Percentage of sites monitoring each parameter type

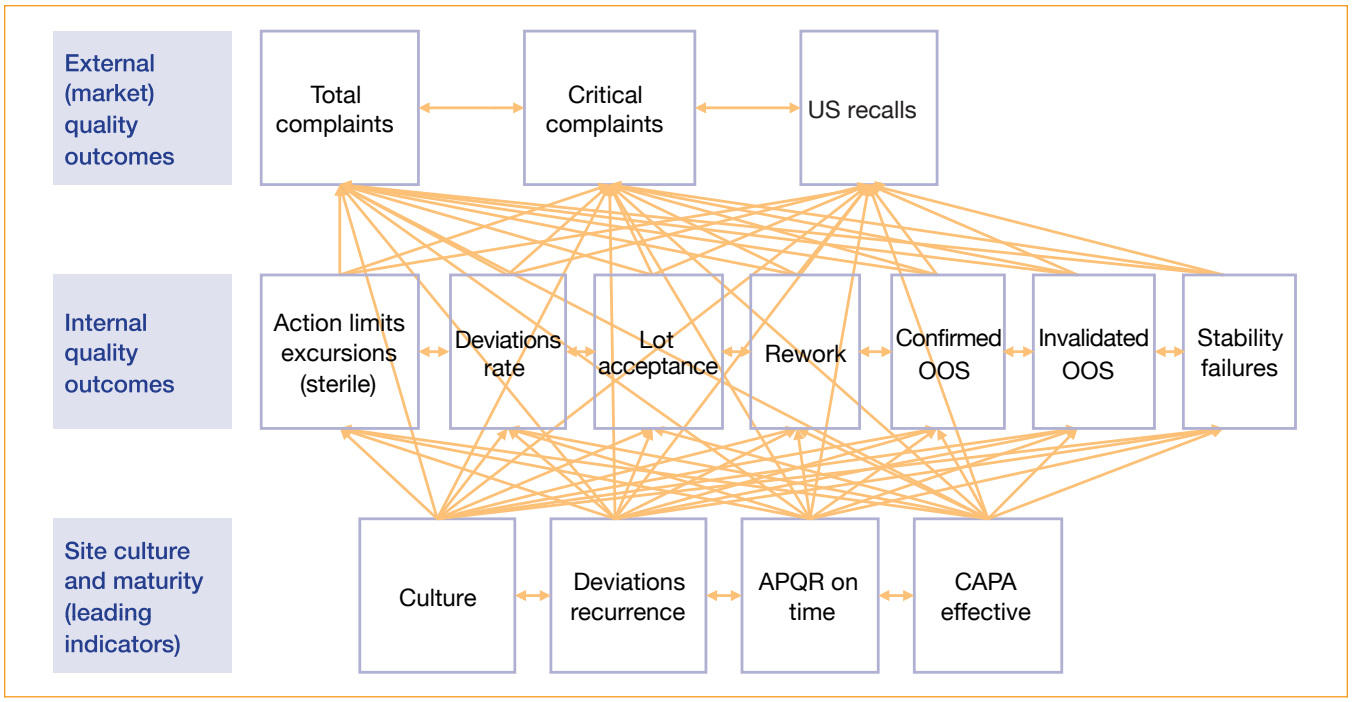


As anticipated the capability approach is variable across companies and in terms of use and applicability. The tool employed e.g. Ppk, Cpk, control charts etc. is contextual and there is no one tool that can be applied to all situations. These tools should be used internally for troubleshooting and identifying continual improvement opportunities rather than for monitoring compliance.

5.7 Establishing Statistically Significant Relationships

Relationships between the collected metrics were tested and assessed using the grouping of quantitative metrics shown in Table 3 (external quality outcomes, internal quality outcomes and site culture and maturity). Connections between each of the standardized quality metrics and Quality Culture Survey values are depicted by the orange lines in Figure 21.

Figure 21: Relationship Testing



As before, incomplete data and extreme outliers were excluded. The resulting sample sizes allowed statistical analysis with some limitations:

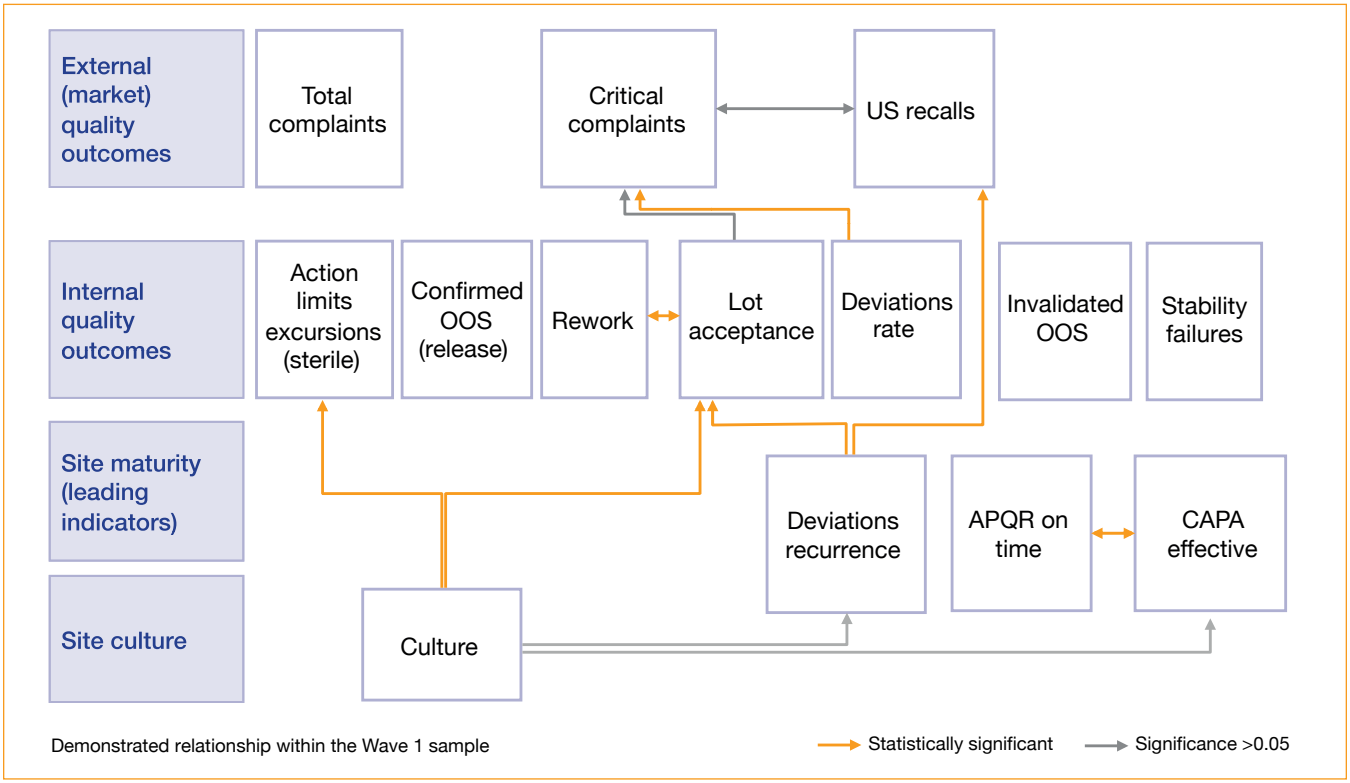
- To allow sufficient sample size most analyses were performed for finished dosage sites overall, not by technology
- Product data were collected on annual basis, not allowing time lag analysis to see how product metrics correlate over time

Identified relationships between metrics were deemed statistically significant when there was a less than 5% likelihood of a coincidence. The strength of the relationships may vary, however, and in some cases are relatively low (e.g., some may correlate with R^2 of 0.30 (30%) or 0.40 (40%). The size of the Wave 1 Pilot data set is acknowledged in these findings and it is also noted that these metrics under examination are influenced by multiple factors not currently included in this analysis.

R^2 measures how well variability of given metric X explains variability of metric Y. It ranges from 0 (no relationship) to 1 (perfect linear relationship). Pearson coefficient (R) measures the extent to which two variables move in the same direction. It varies from 0 (random relationship) to 1 (perfect linear relationship).

The relationships determined following this analysis are summarized in Figure 22.

Figure 22: Significant Relationships



The orange lines show relationships that are statistically significant (< 0.05%), while the grey line shows relationships that have a significance level exceeding 0.05% but are still considered worth examining further given the level of variability and relatively low sample size.

Multivariate correlation analysis was not performed since the sample size was insufficient for this.

A statistically significant relationship does not imply causation. Causation can only be proposed after studying underlying factors, which requires further work, as does understanding the degree of correlation and direction of a relationship.

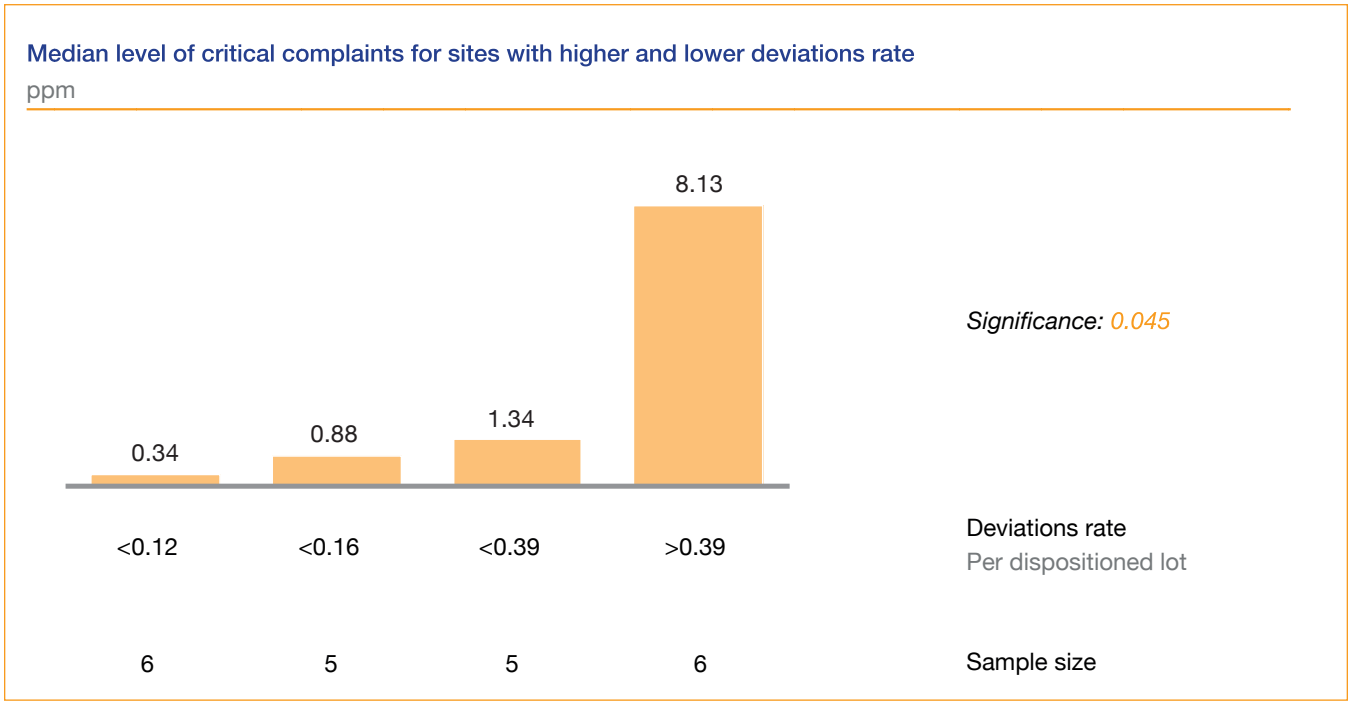
5.8 Statistically Significant Relationships in Wave 1 Pilot Data

Statistically significant relationships were found between the following metrics, or quality culture values, and those metrics as shown in the appropriate analysis figure.

- Critical complaints and deviations rate (Figure 23).
- US recalls and deviations recurrence (Figure 24).
- Lot acceptance rate and rework (Figure 25).
- Lot acceptance rate and quality culture values (Figure 26).
- Lot acceptance rate and deviations recurrence (Figure 27).
- Action limit excursions (sterile products) and quality culture values (Figure 28).
- APQR on time and CAPA effectiveness rate (Figure 29).

Figure 23: Critical Complaints and Deviations Rate

Critical complaints in selected intervals of deviations rate
Sample of 21 plants, finished dosage, average annual values

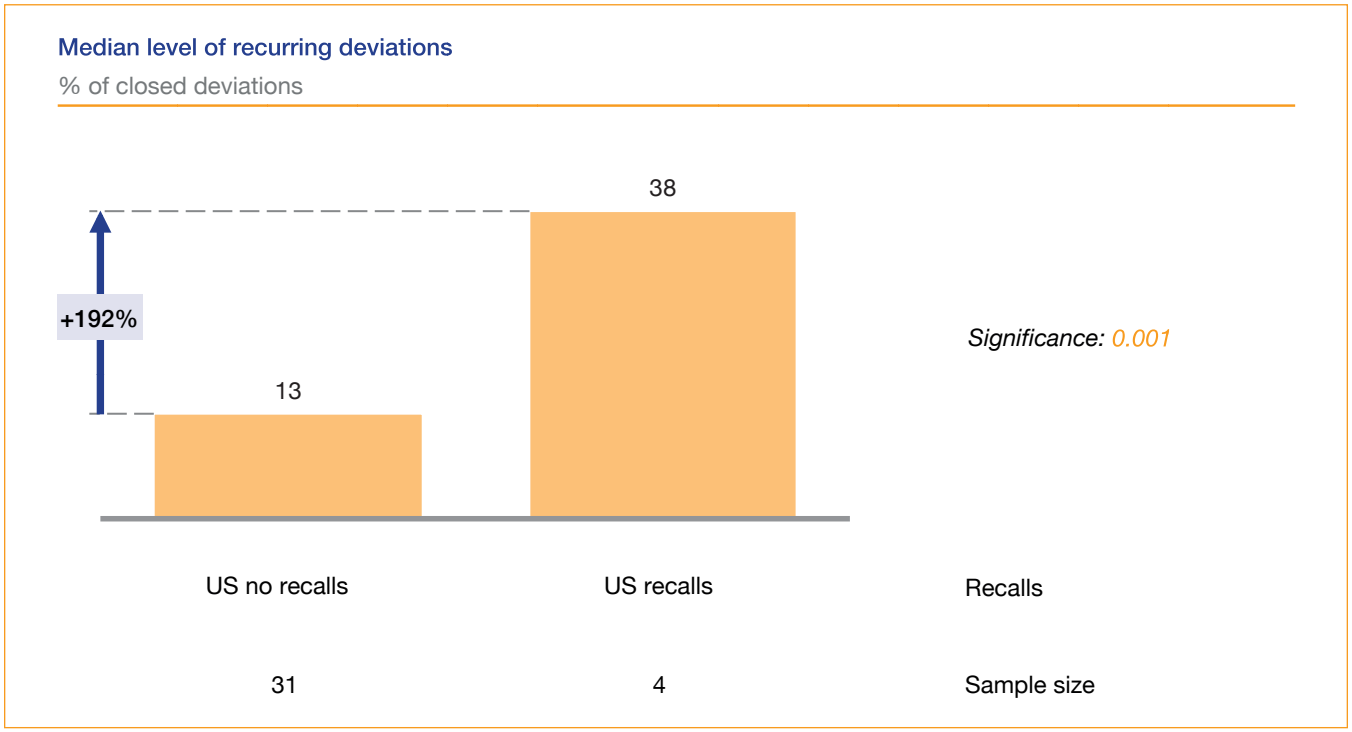


Significance level is a result of independent samples median (chi-square) test.
Value below 0.05 means that the medians of Variable are significantly different between categories.

Critical complaints rate increases with increase of deviations rate. Values for deviations rate are quartile boundaries. The critical complaints rate values are statistically different (0.045%) using the chi-squared test.

Figure 24: US Recalls and Deviations Recurrence

Recurring deviations for plants with and without recalls
Sample of 35 plants, all technologies, average annual values

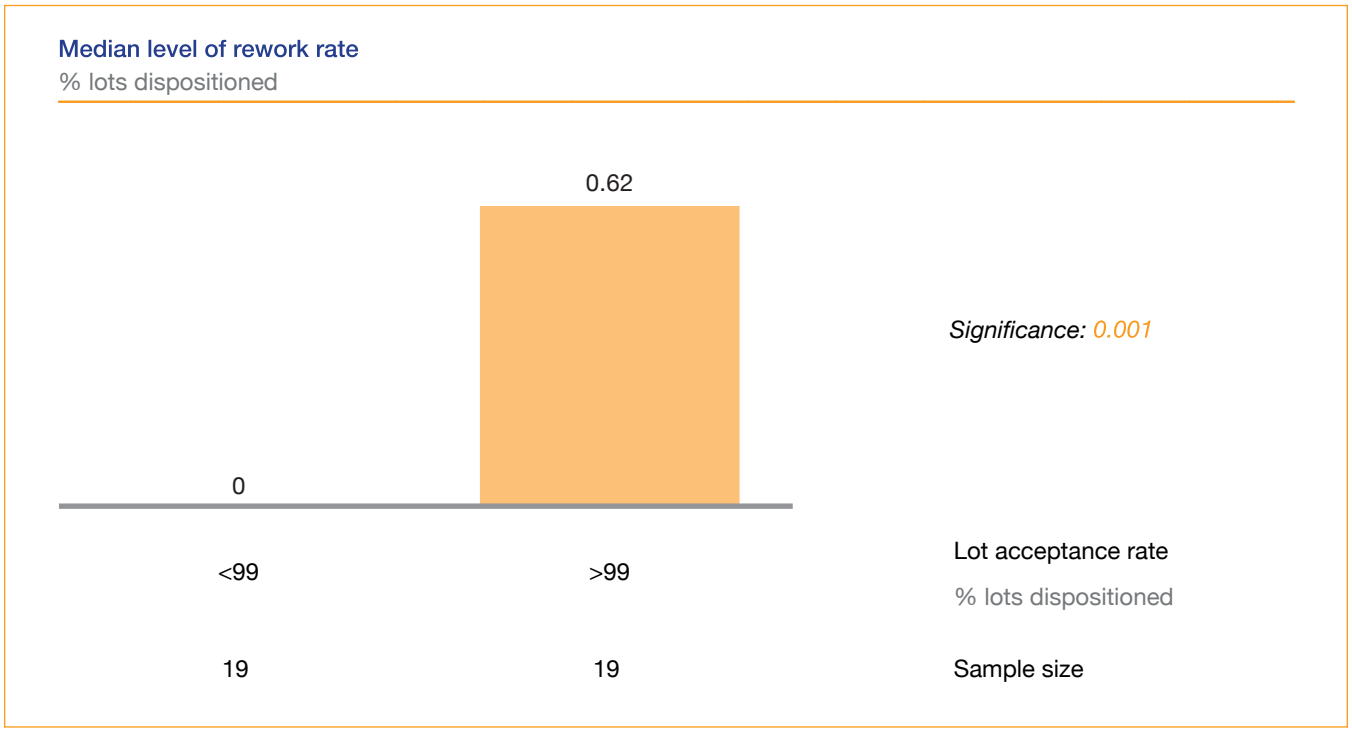


Significance level is a result of independent samples median (chi-squared) test.
Value below 0.05 means that the medians of variable are significantly different between categories.

There is a relationship (significance 0.001%) between US recalls and recurring deviations. Because only four of the 35 sites in this analysis had a recall, causal relationship between the recall and recurring deviations rate has not been established.

Figure 25: Lot Acceptance Rate and Rework

Rework rate in selected intervals of lot acceptance
Sample of 38 plants, all technologies, average annual values

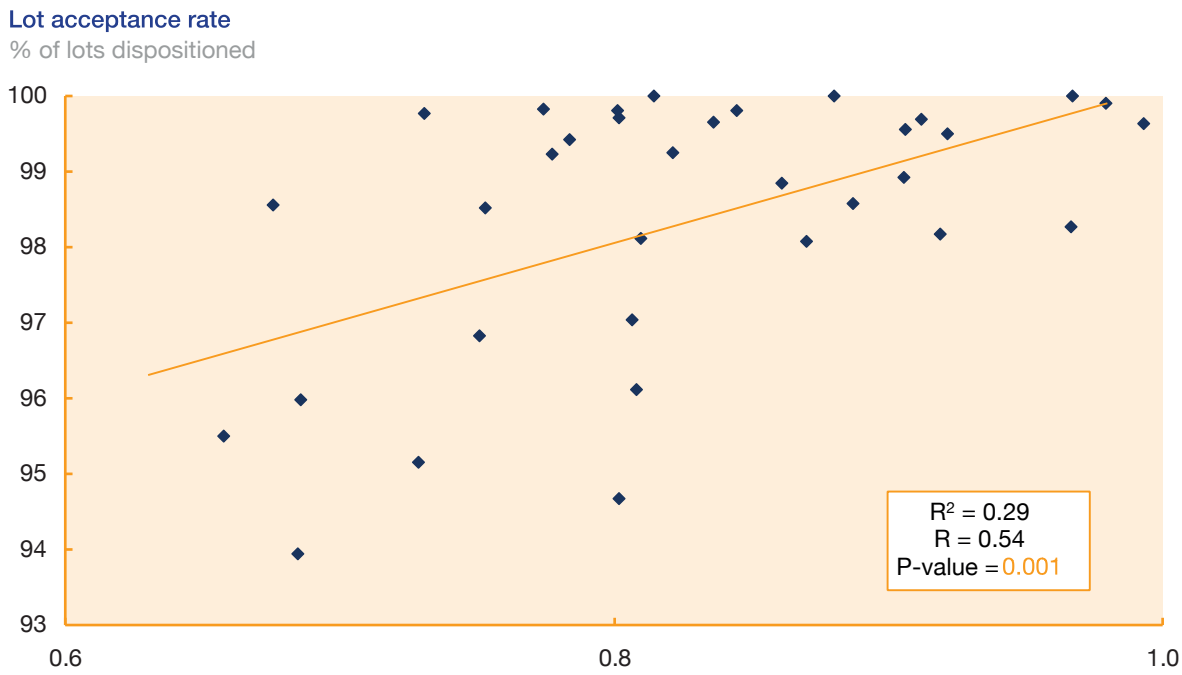


Significance level is a result of independent samples median (chi-square) test. Value below 0.05 means that the medians of variable are significantly different between categories. Significance represents non-linear dependency.

Higher rework rate is associated with higher lot acceptance rate (fewer rejects) at a significance level of 0.001%. Sample sizes are split equally between two levels of rework rate. Higher rework rates are also associated with fewer rejects (higher lot acceptance rate), therefore rework rate may have additional value as balancing metric, but the site practices driving this relationship still require better understanding.

Figure 26: Lot Acceptance Rate and Quality Culture Values

Sample of 34 plants, all technologies, average annual values

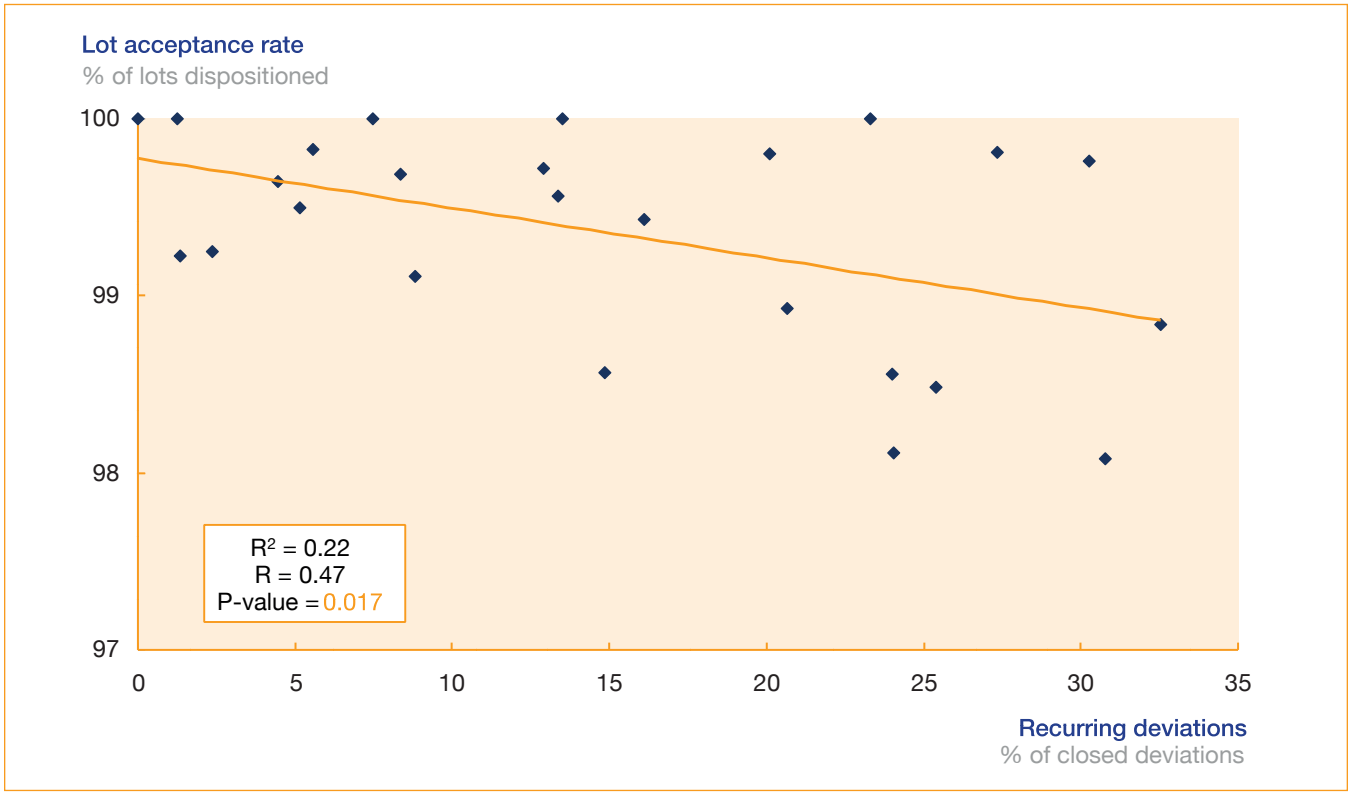


R^2 measures how well variability of given metric X explains variability of metric Y. It ranges from 0 (no relationship between X and Y) to 1 (perfect linear relationship). Pearson coefficient (R) is a measure to what extent two variables move in the same direction. It varies from 0 (random relationship) to 1 (perfect linear relationship) or -1 (perfect negative linear relationship). P-value is probability that correlation is zero (in this case this means there is no linear correlation between X and Y variables), value below 0.05 indicates significant results.

Stronger quality culture values (total) are associated with higher lot acceptance rate. The significance level of 0.001% is strong, however, the level of correlation is weak (R^2 = 0.29) since lot acceptance rate is also influenced by other factors besides quality culture.

Figure 27: Lot Acceptance Rate and Deviations Recurrence

Sample of 25 plants, all technologies, average annual values

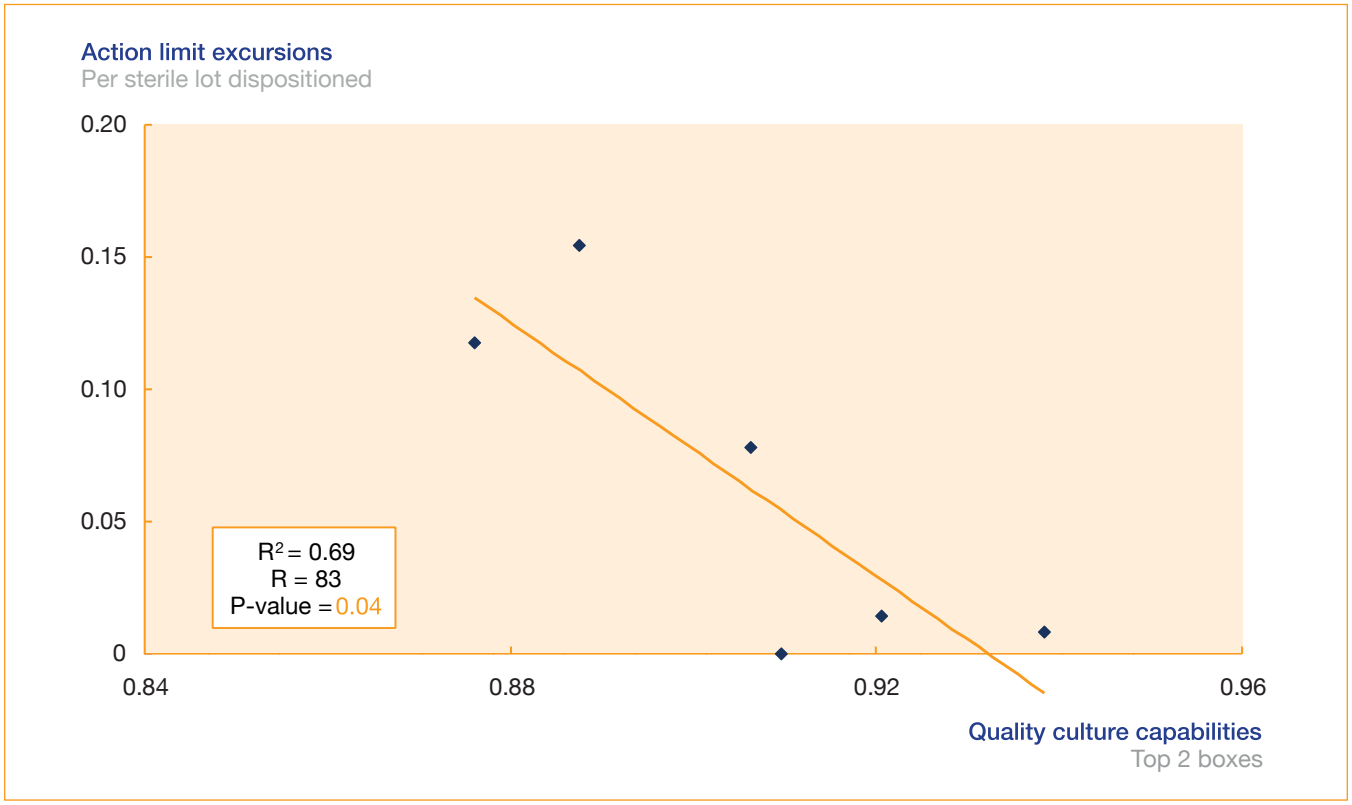


R^2 measures how well variability of given metric X explains variability of metric Y. It ranges from 0 (no relationship between X and Y) to 1 (perfect linear relationship). Pearson coefficient (R) is a measure to what extent two variables move in the same direction. It varies from 0 (random relationship) to 1 (perfect linear relationship) or -1 (perfect negative linear relationship). P-value is probability that correlation is zero (in this case this means there is no linear correlation between X and Y variables), value below 0.05 indicates significant results.

A higher deviations recurrence rate is correlated to lower lot acceptance rate at a significance level of 0.017%, with higher quality culture values also associated with lower recurring deviations rate. The level of correlation (R^2 = 0.22), however, is weak, since lot acceptance rate is influenced by factors other than deviations recurrence.

Figure 28: Action Limit Excursions (Sterile Products) and Quality Culture Values

Sample of 6 plants, steriles, average annual values

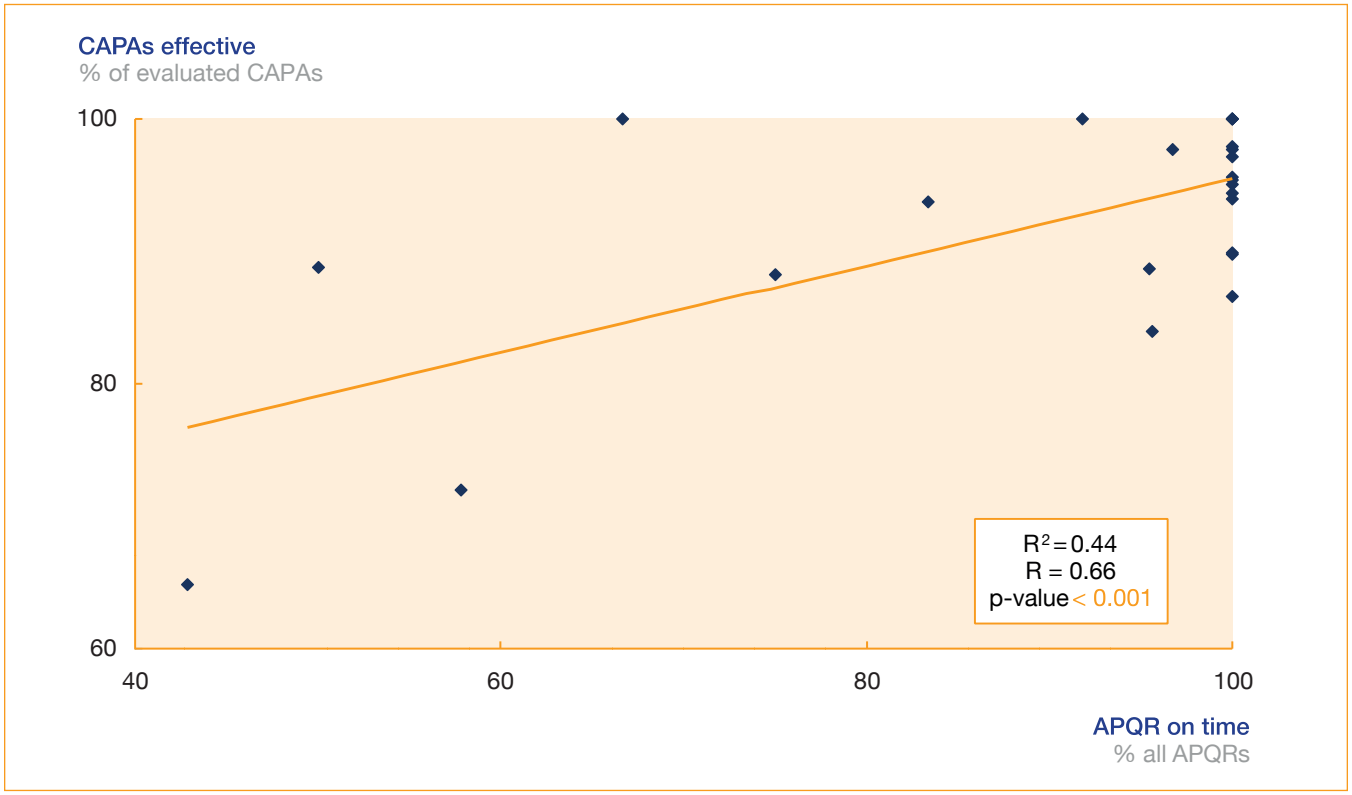


R^2 measures how well variability of given metric X explains variability of metric Y. It ranges from 0 (no relationship between X and Y) to 1 (perfect linear relationship). Pearson coefficient (R) is a measure to what extent two variables move in the same direction. It varies from 0 (random relationship) to 1 (perfect linear relationship) or -1 (perfect negative linear relationship). P-value is probability that correlation is zero (in this case this means there is no linear correlation between X and Y variables), value below 0.05 indicates significant results.

Stronger quality culture values (capability dimension) also link to fewer action limit excursions for sterile product manufacture at a significance level of 0.04%.

Figure 29: APQR on Time and CAPA Effectiveness Rate

Sample of 24 plants, all technologies, average annual values



R^2 measures how well variability of given metric X explains variability of metric Y. It ranges from 0 (no relationship between X and Y) to 1 (perfect linear relationship). Pearson coefficient (R) is a measure to what extent two variables move in the same direction. It varies from 0 (random relationship) to 1 (perfect linear relationship) or -1 (perfect negative linear relationship). P-value is probability that correlation is zero (in this case this means there is no linear correlation between X and Y variables), value below 0.05 indicates significant results.

CAPA effectiveness rate is strongly related (significance < 0.001%) to APQRs on time. This relationship may possibly be driven by similar culture and capabilities. This has not been demonstrated, however.

In conclusion, the following metrics and survey values have statistically significant relationships either directly to the main external quality outcome (US recalls) or via an internal quality outcome as shown in Figure 22:

- Critical complaints
- Lot acceptance rate
- Deviations rate
- Recurring deviations rate
- Quality culture values

Rework rate is strongly linked to lot acceptance rate but is not included in the above list.

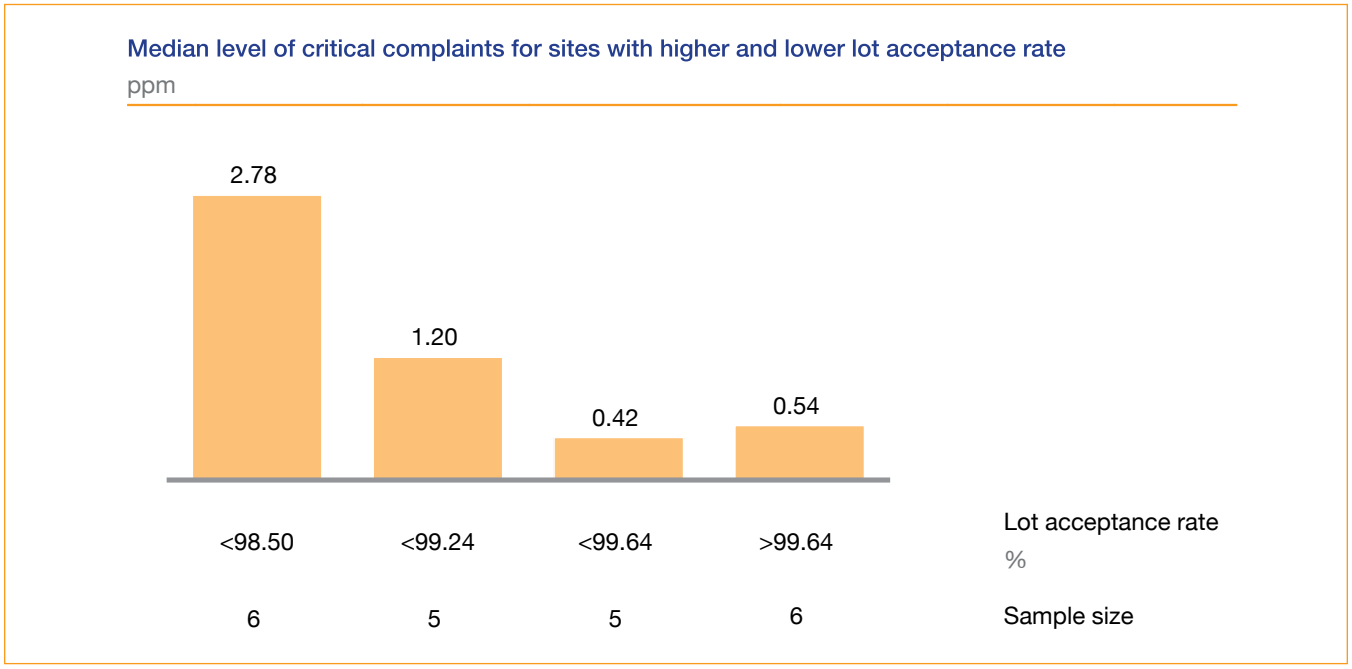
5.9 Relationships at Lower Levels of Significance

In addition to the analysis outlined above, relationships at a weaker significance level (more than 5% chance that the relationship is coincidental) were also seen between:

- Critical complaints and lot acceptance rate (Figure 30)
- Critical complaints and US recalls (Figure 31)
- Deviations recurrence rate and quality culture values (Figure 32)
- CAPA effectiveness rate and quality culture values (Figure 33)

Figure 30: Critical Complaints and Lot Acceptance Rate

Critical complaints in selected intervals of lot acceptance
Sample of 22 plants, finished dosage, average annual values



Lot acceptance rate may have a relationship to critical complaints. Values for lot acceptance rate in Figure 30 are quartile boundaries.

Figure 31: Critical Complaints and US Recalls

Complaints for plants with and without recalls
Finished dosage, average annual values

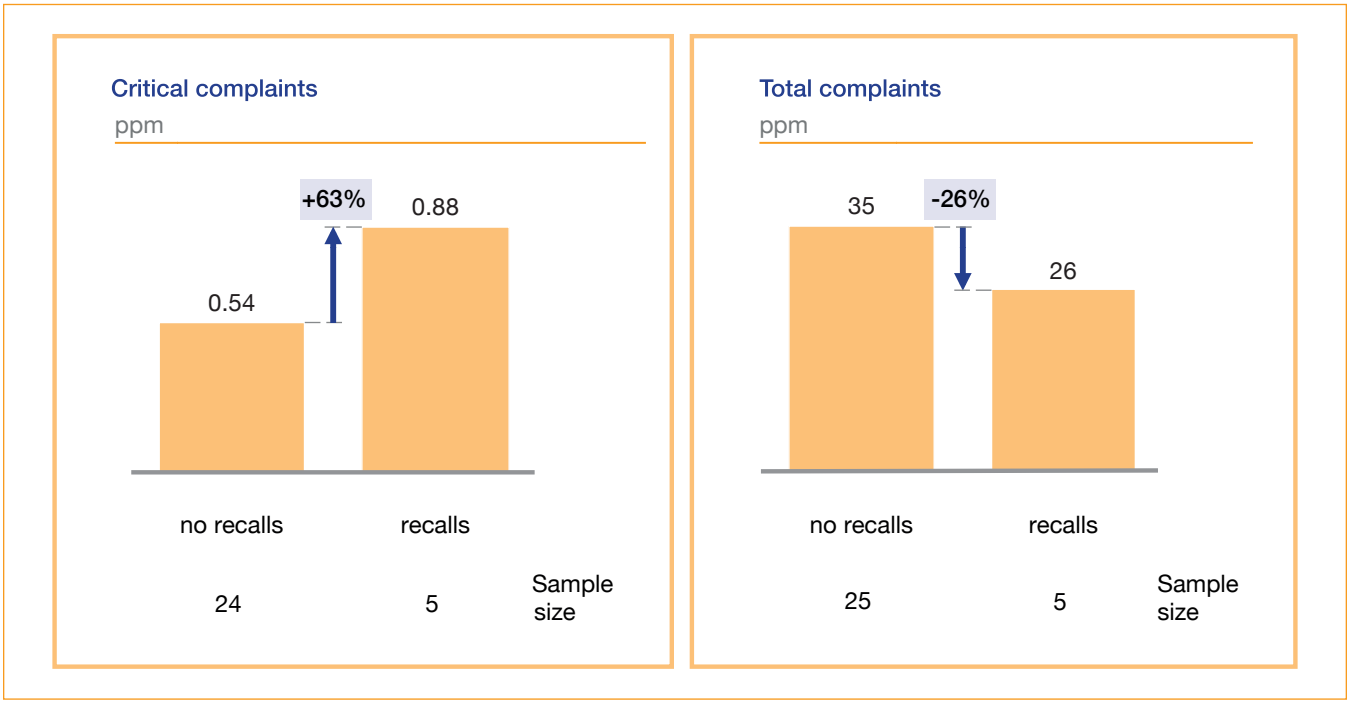


Figure 31 shows that critical complaints trend is consistent with number of US recalls from the left-hand graph, while total complaints goes in the opposite direction. A possible reason for this difference is that critical complaints, although difficult to define precisely, could be much less variable than total complaints, which includes many categories, such as subjective defects.

Figure 32: Deviations Recurrence Rate and Quality Culture Values

Recurring deviations in selected intervals of Quality culture overall
Sample of 9 plants, solids, average annual values

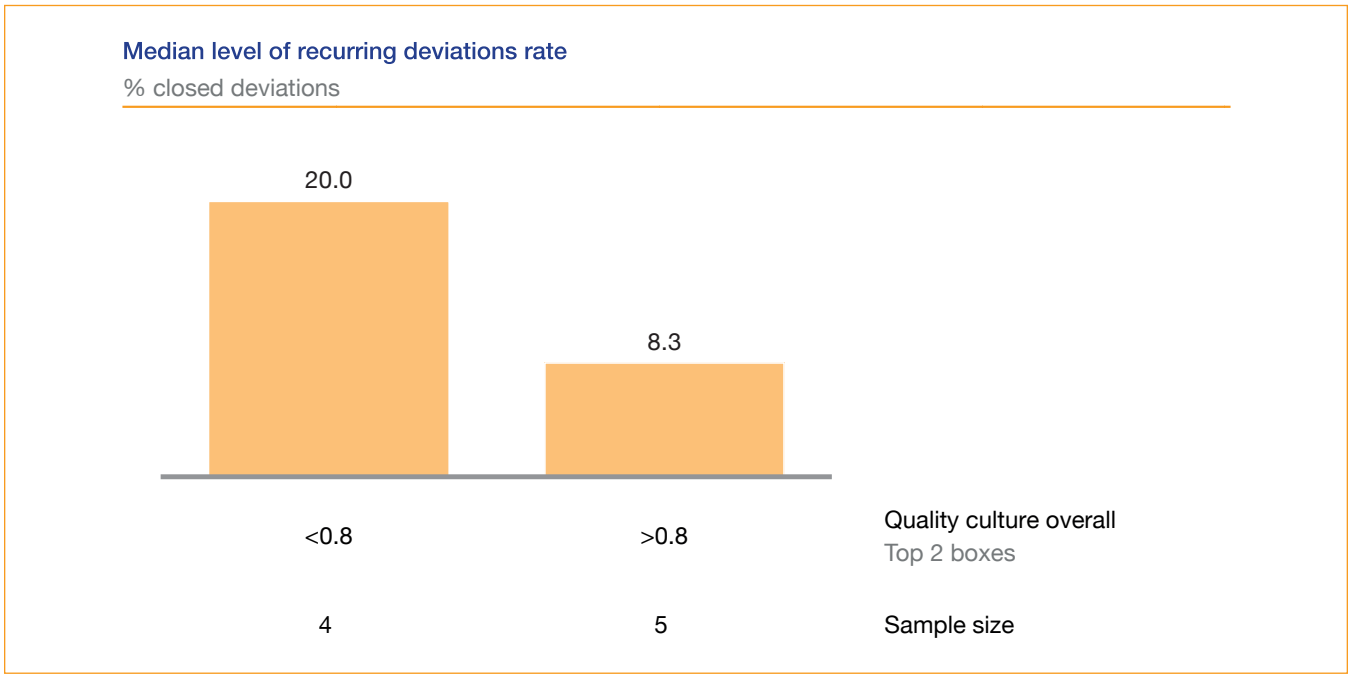
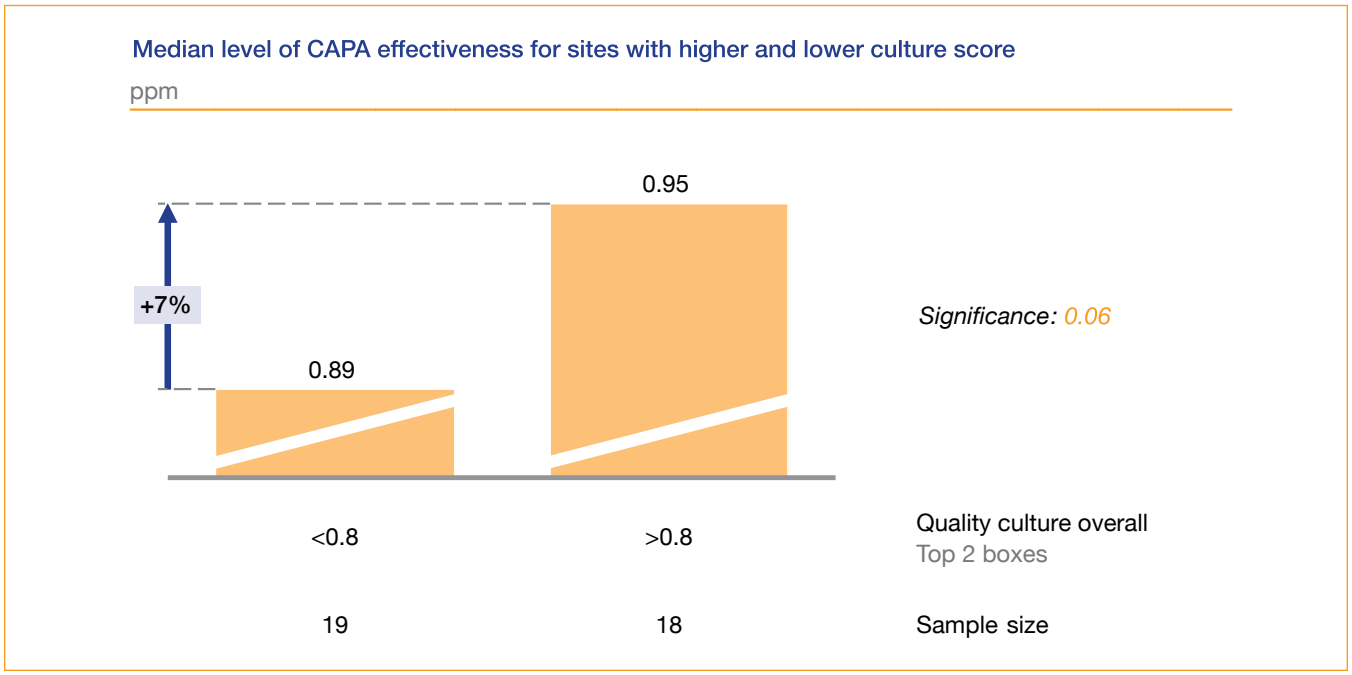


Figure 32 indicates that there is some evidence that quality culture values are related to deviations recurrence rate. When the sample is split into two parts, those sites having a quality culture values greater than 0.8 are found to also have a lower recurring deviations rate (8.3 versus 20.0)

Figure 33: CAPA Effectiveness Rate and Quality Culture Values

CAPA effectiveness in selected intervals of quality culture
Sample of 74 average quarterly values, all technologies



Using the same quality culture value as given in Figure 32 of 0.8, there may also be a difference (7% observed) in CAPA effectiveness rate. The significance level for this relationship is close to, but above, 0.05% (0.06%) and this is shown in Figure 33.

Note: The quality culture value of 0.8 also splits the sample into two almost equal parts.

5.10 Comparisons Where Metrics Are Not Differentiated or Are Inconclusive

Some metrics were not differentiated or had no conclusive results and these included:

- APQR on time, which is also not highly differentiating, reported as 100% by the majority of sites.
- For stability failures, confirmed OOS (product release and incoming materials) and Unconfirmed OOS, the pilot sample did not yield clear conclusions.
- The technology-specific metrics were not sufficiently tested due to smaller sample size, and some of them (media fills, environmental rejects) were not differentiated between sites.

5.11 Discussion of Relationships

Combined findings on metrics, quality culture survey values in terms of relationships, time spent to collect metrics, difficulty of collection and difficulty of definition are summarized below in Table 4.

Table 4: Overview of Findings from Pilot

Site level data

	Value	Time	Difficulty[1]	Definitions discussion[2]
US recalls	Relationship	Low	1	
Total complaints rate	Inconclusive yet	High	2	
Critical complaints rate	Relationship	High	2	**Yes**
Lot acceptance rate	Relationship	Moderate	2	
Rework rate	Inconclusive yet	Moderate	2	**Yes**
Confirmed OOS rate	Inconclusive yet	High	2	
Stability failure rate	Inconclusive yet	Moderate	2	
Deviations rate	Relationship	Moderate	2	
Invalidated OOS rate	Inconclusive yet	Moderate	2	
APQR on time	Inconclusive yet	Low	1	
Recurring deviations rate	Relationship	Moderate	3	**Yes**
CAPA effectiveness rate	Inconclusive yet	Moderate	2	**Yes**
Successful media fills	Not differentiating	Low	2	
Action limits excursions	Inconclusive yet	Moderate	3	
Environmental rejects	Not differentiating	Moderate	3	
Culture	Relationship			

[1] 1 = very easy to 4 = very difficult; - site ratings related to whether the data are available in the requested form, or requires recalculation/aggregation, or collection from fragmented sources;

[2] Metrics where definitions vary significantly across companies, and feedback was provided through pilot

Further discussion of each column in Table 4 is given below.

5.11.1 Column 1: Value

As illustrated in Figure 22 the following metrics/values have statistically significant relationships (as discussed in Section 5.8) directly or indirectly with an external quality outcome (US recalls or critical complaints):

- US recalls (outcome)
- Critical complaints (outcome)
- Lot acceptance rate
- Deviations rate
- Recurring deviations rate
- Quality culture

In column 1, the above metrics/values are shown with grey shading.

Critical complaints is related to US recalls and may be a relevant measure of external quality, however, there was much feedback from site leaders from participant companies that the definition of critical complaints led to difficulty in interpretation of values to submit and as noted earlier there was a only a very small number of recalls from the pilot study.

Lot acceptance rate, deviations rate, deviations recurrence rate and quality culture values all have relationships (some statistically significant) to quality outcomes (critical complaints and US recalls).

As stated at the beginning of this section, it cannot be stressed too strongly that the **statistically significant relationships observed in the Wave 1 Pilot sample do not indicate cause**. Further testing and studies are required to better understand these relationships.

Some metrics had inconclusive comparisons with other metrics and further testing is required to indicate the presence or absence of relationships as shown in Table 4.

Some metric comparisons for technology-specific metrics (successful media fills and environmental rejects) were not differentiated; these are highlighted with orange shading in column 1.

5.11.2 Column 2: Time

Time spent collecting each individual metric is tabulated in column 2. Metrics with lowest amount of time are shaded grey and those with the highest time are shaded orange. Some possible reasons for differences are:

- Complaints are usually received centrally, and effort is required to understand the complaint and distribute to a particular site in a supply chain, and then for that site to categorize the complaint and submit a value. Also some complaints may need to be forwarded to a number of different sites (e.g., to a packaging site if the complaint is related to the packaging operations).
- A confirmed OOS metric may be hard to collect on a site basis since it is usually collected on a product basis and effort is required to aggregate to a site level.
- APQRs on time metrics are probably either already collected or easy to collect.
- US recalls are a low number and of such high importance that the metric is easy to collect.
- For sterile product manufacturing sites, the successful media fills metric is a relatively low number and easy to collect.

Estimates of time to collect Quality Culture Survey data were not collected in the overall effort estimates, since even the approximately 5 minutes required for each individual response would still present a large effort given the overall number (10,300) of survey respondents.

5.11.3 Column 3: Difficulty

These rankings are the participating sites' estimates of the difficulty of collecting a metric as shown in Figure 16.

Most difficult was the recurring deviations rate, since this metric does not appear to be collected routinely by most sites. As shown in Figure 17, there is not necessarily a correlation between degree of difficulty of collecting a metric and median time for collection.

5.11.4 Column 4: Definitions

In the last column, metrics are given where the McKinsey support team spent most time explaining a metric definition and where, from feedback, definitions varied most between companies and between a company definition and the definition used in the pilot. The metrics most difficult to define were:

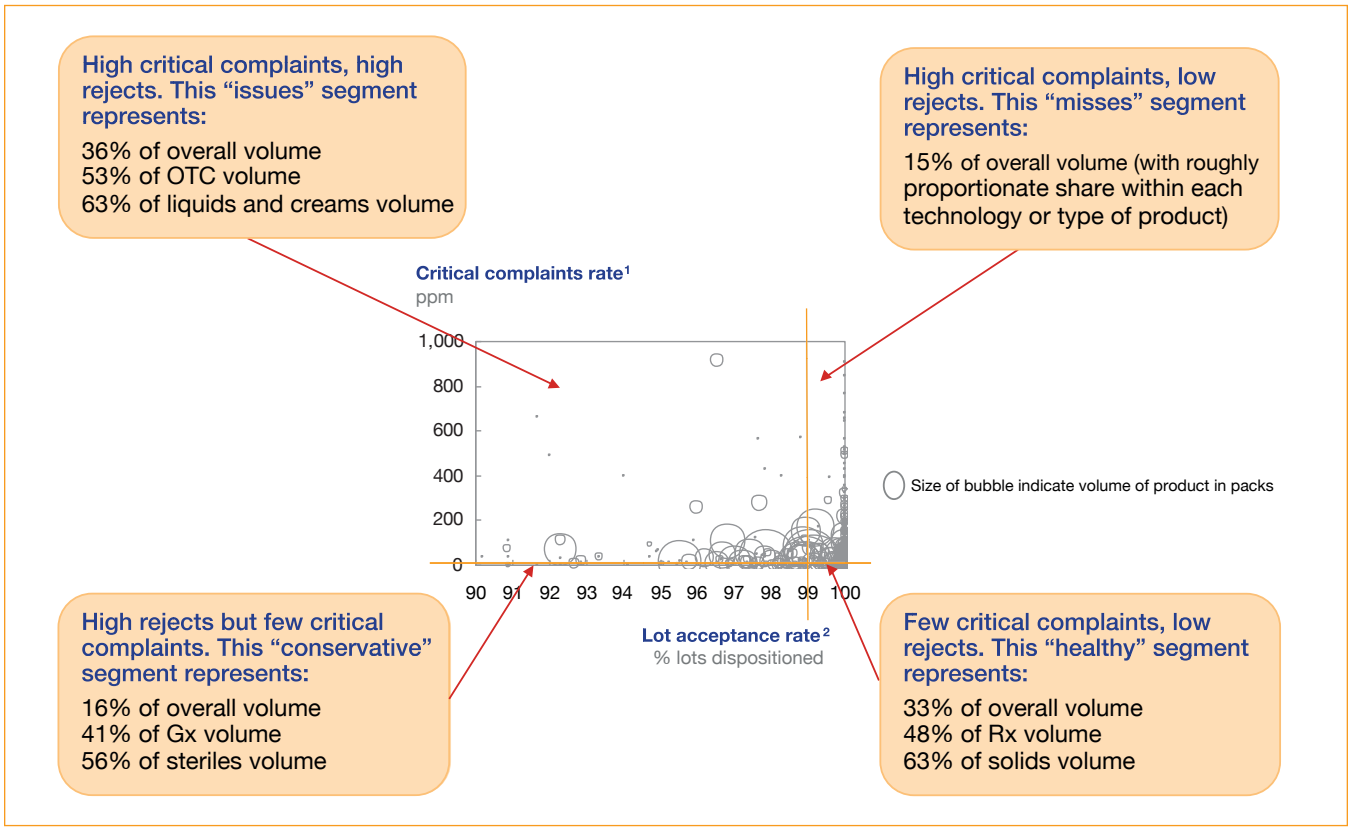
- Critical complaints rate
- Rework rate
- Recurring deviations rate
- CAPA effectiveness rate

5.12 Complaints Analysis

The database was examined to evaluate relationships between critical complaints, lot acceptance rate and type of product and technology. Data were plotted in a scatter plot as given in Figure 34.

Figure 34: Scatter Plot of Critical Complaints, Lot Acceptance Rate, Technology and Type of Product

Product level data, retrospective period (12 m.), finished dosage plants



[1] Products with over 1000 critical complaints per milion packs have been omitted on the chart
[2] Products with less than 90% lot acceptance have been omitted on the chart

The red lines represent the 80th percentile boundaries for lot acceptance rate (99.24, lower to the left) and critical complaints (0.12, lower below line).

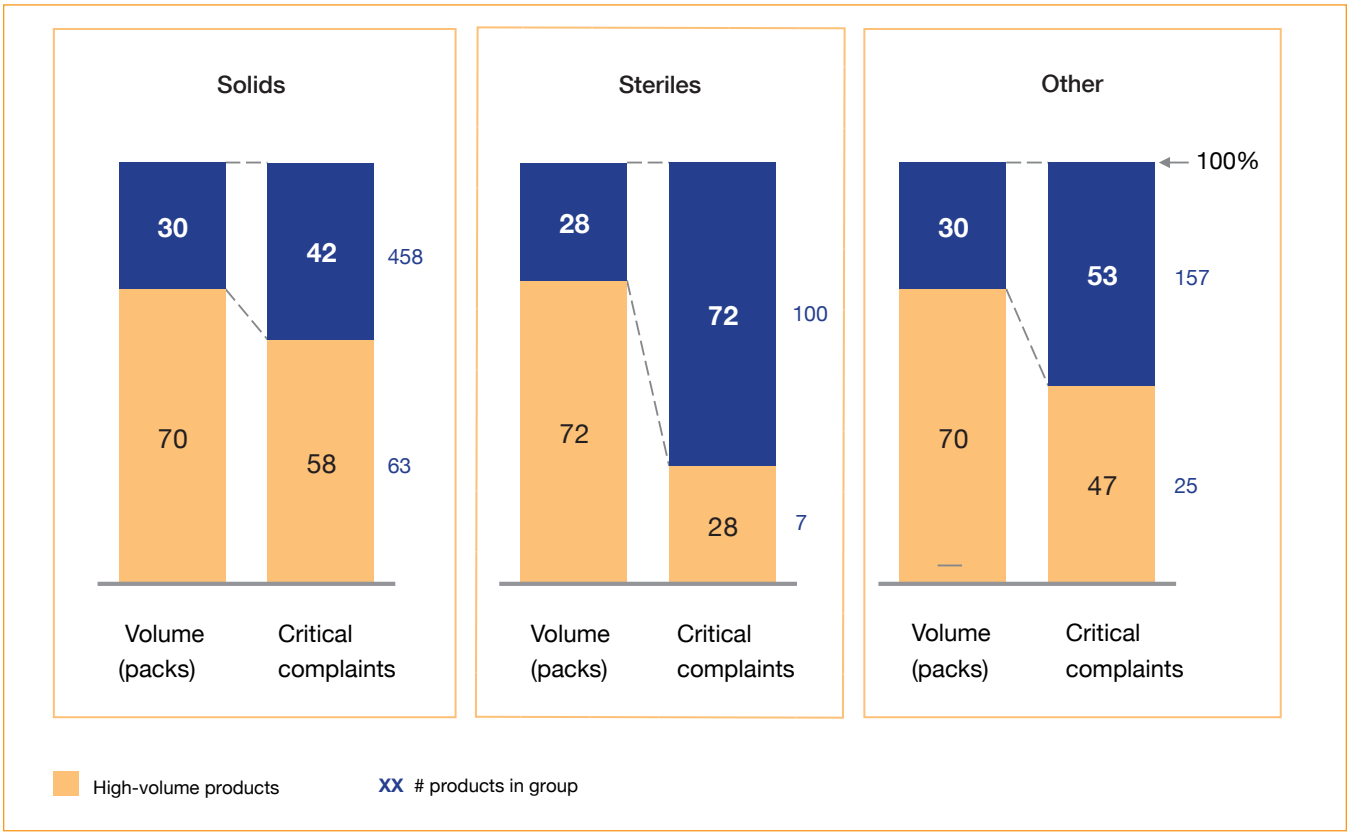
This shows that products vary in their rejects and complaints profile depending on both the technology and the type of product. It is therefore hard to draw any specific conclusions at this time from this analysis and available data set.

5.13 Analysis of Product-Based Metrics

Product-based metrics were analyzed for the complaints and lot acceptance rate metrics. The analysis indicated that "tail" products (90% of products representing 30% of volume) represent a different level of critical complaints and rejects depending on technology as shown in Figure 35 and Figure 36.

Figure 35: Critical Complaints and Volumes for Product-Based Data

Product level data[1]

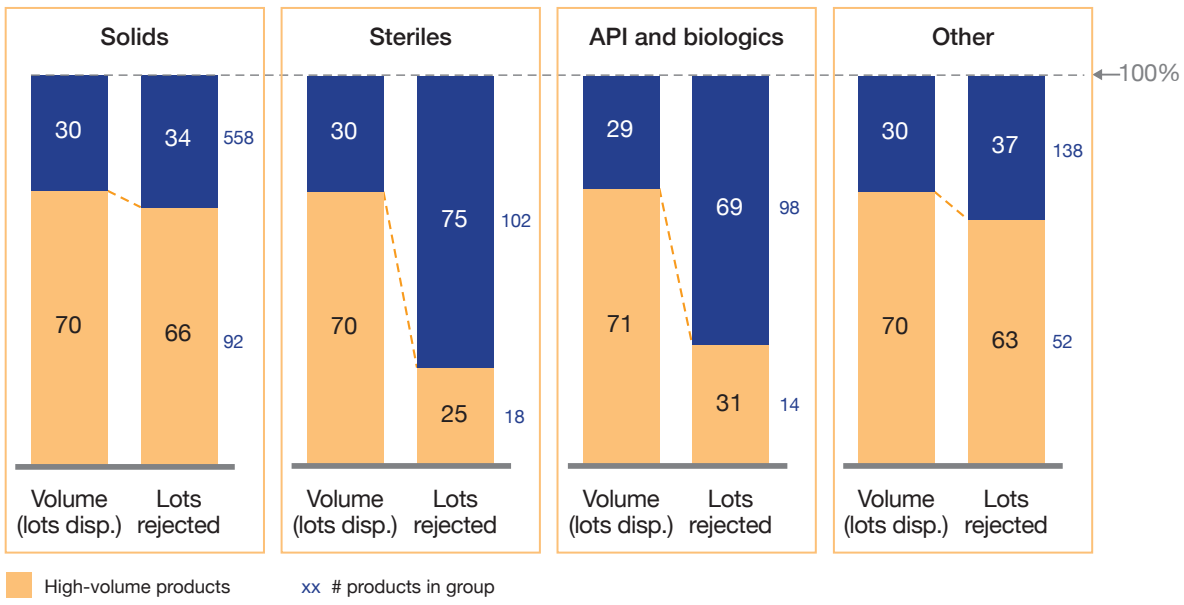


[1] Excludes inactive products (with no packs released in reporting period)

The multiple tail products, representing 30% of volume, result in 42% of critical complaints for solids, 70% of critical complaints for sterile products and 53% of critical complaints for other technologies.

Figure 36: Rejects (from Lot Acceptance Rate) and Volumes for Product-Based Data

Product level data



For rejects observed as part of lot acceptance rate, 30% of the volume gave 34% rejects for solids, 70–75% of rejects for both sterile product and drug substance sites and 37% rejects for other technologies.

5.14 McKinsey Analytical Effort and Observations

Analysis by McKinsey regarding the amount of effort that their support personnel had performed was estimated as:

- Each site required on average 22 hours of dedicated support—explaining the data requirements, answering questions, reviewing and verifying the submitted data through discussions with the site.
- The more structured and supported the data submission process is, the more accurate the received data.
- Checking submitted data for coherence/accuracy was required even after several submission cycles.
- Preparation of submission templates, databases and process, and analyzing the gathered data required approximately 400 additional hours

The above may indicate the level of support that FDA will be requested to provide, at least for early rounds of data collection.

During the course of the Pilot there were many interactions between participating companies and McKinsey personnel and the following success factors were observed:

- **Good definitions:** To make data submission as easy as possible and minimize the need for follow-up discussion, metrics must be defined to a very high standard. Industry-consensus definitions almost always differ from those being used by companies; hence it is necessary to allow companies to adjust.
- **Strong support:** Experienced, dedicated support was required throughout the collection period to answer questions quickly and make necessary clarifications.
- **Engaged and committed participants:** Given that participating in the pilot was voluntary, it was clear that participant company staff were committed and knowledgeable; most companies also had mature systems. Using McKinsey personnel and teleconferences of participant company site leaders permitted much sharing and learning between companies.
- **Built-in checks:** Submitted data required careful checking to ensure consistency and accuracy.
- **Feedback from participating companies:** McKinsey collected nonconfidential and cross-company comments about pilot design and operation. Definition differences between ISPE and participating companies for “recurring deviations rate” generated considerable feedback, for example.

This type of feedback reinforces the need to have precise, agreed definitions. There is the possibility that even with precise definitions, differences in product and process flow, could lead to undesirable variation in companies’ submissions.

6 Output and Lessons Learned from ISPE Quality Metric Wave 1 Pilot

This section discusses success factors for the Wave 1 Pilot and an overview of findings such as a summary of effort required and comparison of site- and product-based data.

The ISPE Quality Metrics Project Team have produced the following lessons learned and outputs from the Wave 1 Pilot based on findings and data generated during the Pilot, comments made by participants as well as comments by McKinsey personnel comparing their experiences with this ISPE Pilot with other benchmarking exercises.

6.1 Success Factors

The following are listed as success factors for the pilot:

- Metrics lists prealigned during industry meetings.
- Precise definitions, developed with input from multiple companies.
- Engaged and committed participants with mature systems.
- Acceptable to submit “good enough” data.
- Strong collaboration across companies and between experts to support a process that allows learning and continuously improving data accuracy.
- Strong McKinsey Support, providing:
  - Structured data submission with detailed guidance on how to report the data.
  - Ability to comment on data points to enable interpretation.
  - Experienced, dedicated support for questions and clarifications during and throughout the data collection effort.
  - Built-in checks and joint review between McKinsey and the participating company to ensure consistency and accuracy of data.
  - Transparency throughout the process, engaging participants in frequent debriefs and discussions.

These factors support starting with a quality metrics program containing a relatively small number of well-defined metrics with a learning period to refine scope, definitions and process.

6.2 Definitions

Definitions are extremely important:

- Definitions must be exact: Denominators in particular are highly sensitive to issues around lot aggregation and final disposition.
- Even common terms like “lot,” “deviations,” “complaints” and “reviews” must be specified in great detail to minimize multiple interpretations.
- Even with detailed definitions, support and answering questions throughout the process is necessary to ensure more accurate data submission.
- Standardized definitions will differ from current company definitions, thus requiring additional work.
- Product and process differences will generate differences and variations in metric ranges.
- Commentary on data points is essential to interpretation and analysis.
- Some variation must be expected due to differences in product, process flows and product/process complexity.

The pilot showed that standardizing metrics definitions across companies is feasible.

6.3 Metrics Collection by Site and by Product

Attempts were made to generate product-level data, however, there are challenges:

- The pilot collected metrics only within a given site, not across multiple sites or across a full supply chain.
- A company’s ability to collect data by site and/or by product differs by metric and how their systems are set up.
- Most companies collect data at site level and may have to manually disaggregate data in order to report at a product level.
- Some metrics data collected at sites cannot be easily allocated to a product (e.g., deviation rate).
- Some companies report some metrics (e.g., OOS) at a product level and have to aggregate data to report at a site level.
- Some data collected at product level may be derived from APRs.
- Depending on the size and complexity of the site, APR preparation may be spread by product on different cycle times during the year.
- Complaints data are generally gathered centrally (corporate level) and distributed to sites for investigation.

In summary, few companies aggregate metrics across the supply chain to be able to report at product/application level.

6.4 Industry Effort

Industry effort is summarized as:

- On average the participating sites in the Wave 1 Pilot spent ~90 hours in pure data collection time. For 12,000 sites with FEI number (6,000 API/ Finished Dosage and 6,000 Labs/Drug Substance/'other') at typical quality labor cost, collecting this amount of data would **cost the industry an additional ~$35 million annually**.
  - This conservative estimate **does not include several factors** that could bring the cost of such a program to $100+ million, such as:
- The 90 hours observed might be underestimated for sites that did not report all products at that site.
- Workload estimates for "good enough" data submission vs. an official submission.
- Time for internal discussions, management review and above-site guidance not included.
- The need for new/modified IT systems was not included in Wave 1.
- Participants had flexibility to provide the most pragmatic data set (e.g., for all products at site or only those for the US market).
- Data was provided within each site and **not** through the full product supply chain.
- Most participants had mature systems and capabilities regarding performance measurement.
- Majority of sites were from developed countries.

The program scope and design can influence these industry costs significantly.

6.5 McKinsey Analytical Effort

There was substantial effort from McKinsey:

- Each site required on average 22 hours of dedicated support—explaining the data requirements, answering questions, reviewing and verifying the submitted data through discussions with the site.
- The more structured and supported the data submission process is, the more accurate the received data.
- Checking incoming data for coherence/accuracy was required even after several data submission cycles.
- Preparation of submission templates, databases and process, and the analysis of the gathered data required approximately 400 additional analytical hours.

As already mentioned in Section 5.14, the above points may indicate the level of support that FDA will be requested to provide, at least for early rounds of data collection.

7 Proposals

Based on findings from the Wave 1 Pilot, in terms of ease and effort of collection and submission of metrics, it is suggested that future quality metrics programs start with a relatively small number of standardized metrics.

From the relationship findings discussed in Section 5 a "starting set" of five metrics are proposed for inclusion in future industry quality metrics programs:

1. Lot acceptance rate (normalized by lots dispositioned) at site level.
2. Lot acceptance rate (normalized by lots dispositioned) at product level within a site.
3. Critical complaints (normalized by number of packs released) at product level by application, not broken down by site.
4. Critical complaints (normalized by packs released) at site level, undifferentiated by product.
5. Deviations rate at site level.

Rationale for including these metrics in a Wave 2 Pilot are given below.

7.1 Rationale for Metrics Proposed as a Starting Set for Wave 2 Pilot

- In most cases the metric was already captured in the Wave 1 Pilot, and continued monitoring is desired over a longer timeframe and broader set of companies, technologies, regions.
- The metric selected demonstrated a statistically significant relationship to one of the following:
  - Deviations recurrence
  - Quality culture values
  - Critical complaints
  - Lot acceptance rate
- It has been demonstrated to be relatively easy to collect and submit.
- It was deemed an important metric for determining site quality performance.
- It will assist the company to identify continual improvement opportunities.
- While the critical complaints metric (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.

The starting set of metrics are now being considered for further analysis in an ISPE Quality Metrics Wave 2 Pilot that will include an extended time period of prospective data collection and increased sample size of participating companies. The momentum and knowledge gained from the Wave 1 Pilot will be leveraged to facilitate an early transition to Wave 2. At the Quality Metrics Summit in Baltimore, April 2015, notification was given for both new and continued participation in a Wave 2 Pilot with a target enrollment commencement date of June 2015. Further design of the Wave 2 Pilot will then get underway.

In addition to the starting set of five metrics listed above, other metrics of interest under consideration for a Wave 2 Pilot include:

- **Deviations recurrence rate:** How to achieve a consistent and accepted industry definition and practice.
- **Unconfirmed OOS:** Test on a bigger sample its relation to culture, particularly useful for laboratories, and assess against outcomes.
- **Quality culture:** How best to assess the influence of quality culture on quality performance outcomes at industry scale.

Briefing sessions with existing participating companies are underway to discuss the implications for ongoing participation. Many other companies in attendance at the Quality Metrics Summit who expressed a willingness to get involved in future pilot studies will be contacted in the near term regarding potential enrollment.

The Wave 2 Pilot will give participating companies a deeper understanding of metrics definitions, the opportunity to contribute to the final pilot study design, and the chance to experience the challenges of a centralized metrics submissions process. In addition, participants will have access to an industry benchmarking report that will allow them to examine their progress with respect to their peers. Finally, participation also provides opportunities to enhance the maturity of their internal metrics programs.

Those interested in receiving further details should contact the ISPE Quality Metrics Initiative at PQLI@ispe.org.

8 Conclusions

The objectives of the Wave 1 Pilot were achieved and consequently the pilot is considered a success.

Evidence of several statistically significant and interesting relationships was found between company quality performance, prevailing quality culture and key patient-related quality outcomes. The insights and knowledge gained from this Wave 1 Pilot confirm the benefits of utilizing quality metrics to monitor, create transparency and drive enhancement in pharmaceutical manufacturing.

Industry engagement throughout this process has confirmed that many sites and companies, including those participating in the Wave 1 Pilot, are already using metrics to drive and monitor internal continual improvement programs. This pilot was an attempt to develop, collect and analyze a standardized set of metrics and has indicated that this is achievable. While there are inherent challenges and costs associated with this exercise, the pilot was able to identify several key benefits also. The pilot therefore accomplished its primary objectives and has set the stage for additional work in Wave 2.

These recommendations were discussed with industry representatives and FDA at the ISPE Quality Metrics Summit and broad support was given to the outcomes and future plans. A significant opportunity for trust building between the industry and the agency has also been realized though this Wave 1 Pilot.

Communications will continue with all parties as the ISPE Quality Metrics Initiative further develops its plans for the Wave 2 Pilot.

Ten years ago Dr. Janet Woodcock outlined the ultimate goal of the desired state [18] for pharmaceutical manufacturing. This Quality Metrics Initiative Pilot has provided early support of this journey toward pharmaceutical excellence.

9 References

1. US Department of Health and Human Services: Food and Drug Administration, et al. “Guidance for Industry: Quality Systems Approach to Pharmaceutical CGMP Regulations.” September 2006. http://www.fda.gov/downloads/Drugs/.../Guidances/UCM070337.pdf

2. ———. “Pharmaceutical GMPs for the 21st Century—A Risk-Based Approach.” Final Report. September 2004. http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/Manufacturing/QuestionsandAnswersonCurrentGoodManufacturingPracticescGMPforDrugs/UCM176374.pdf

3. ———. “Guidance for Industry: PAT—A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance.” September 2004. http://www.fda.gov/downloads/Drugs/Guidances/ucm070305.pdf

4. ———. “Guidance for Industry: Process Validation: General Principles and Practices.” Revision 1. January 2011. http://www.fda.gov/downloads/Drugs/Guidances/UCM070336.pdf

5. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonized Tripartite Guideline. “Pharmaceutical Development: Q8 (R2).” Step 4 version, August 2009. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_R1/Step4/Q8_R2_Guideline.pdf

6. ———. Quality Risk Management: Q9.” Step 4 version, 9 November 2005. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q9/Step4/Q9_Guideline.pdf

7. ———. “Pharmaceutical Quality System: Q10.” Step 4 version, 4 June 2008. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q10/Step4/Q10_Guideline.pdf

8. ———. “Development and Manufacture of Drug Substances (Chemical Entities and Biotechnological/Biological Entities): Q11.” Step 4 version, 1 May 2012. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q11/Q11_Step_4.pdf

9. Government Printing Office. Food and Drug Administration Safety and Innovation Act (FDASIA). Publication L. 112–144. July 9, 2012. http://www.gpo.gov/fdsys/pkg/PLAW-112publ144/pdf/PLAW-112publ144.pdf

10. Brookings Institution. “Measuring Pharmaceutical Quality through Manufacturing Metrics and Risked-Based Assessment,” Expert Workshop, 1–2 May 2014 http://www.brookings.edu/events/2014/05/01-measuring-pharmaceutical-quality

11. McKinsey & Company. “Pharma Operations Benchmarking of Solids (POBOS).” http://solutions.mckinsey.com/pobos/

12. ISPE. “ISPE Proposals for FDA Quality Metrics Program—Whitepaper.” 20 December 2013. www.ISPE.org

13. US Department of Health and Human Services: Food and Drug Administration. "Guidance for Industry: Self-Identification of Generic Drug Facilities, Sites, and Organizations." August 2012 http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm316721.htm

14. ISPE. "ISPE Drug Shortages Prevention Plan: A Holistic View from Root Cause to Prevention." October 2014. www.ISPE.org/DrugShortagesPreventionPlan

15. US Department of Health and Human Services: Food and Drug Administration. "Food and Drug Administration Drug Shortages Task Force and Strategic Plan; Request for Comments." *Federal Register*, Docket No. FDA-2013-N-0124. 12 February 2013. http://www.gpo.gov/fdsys/pkg/FR-2013-02-12/html/2013-03198.htm

16. Woodcock, Janet. "The Quality Revolution: The Future of Pharmaceutical Manufacturing and Its Regulation. "Presented at the ISPE Quality Metrics Summit, Baltimore, MD, 22 April 2015.

17. http://rx-360.org/Portals/1/Guidance/FDA Annual Inspection Report on Establishments/2015 Annual Report on Insp of Establishments in FY 2014 Final.pdf

18. Woodcock, J. "The Concept of Pharmaceutical Quality." *American Pharmaceutical Review* 47(6): 1–3, 2004.

Appendix 1

Definitions of Quantitative Metrics Used in Pilot

Lot acceptance rate

Metric	Definition
Lot acceptance rate = Total lots released for shipping out of the total finally dispositioned lots for commercial use in the period	▶ Total lots 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. 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. ▶ Total lots rejected = total full lots 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. ▶ Total lots released (accepted) = total lots dispositioned less total lots rejected.

Complaints Rate (Total and Critical)

Metric	Definition
Total complaints rate = Total complaints received in the reporting period, related to the quality of products manufactured in the site, normalized by the number of packs released	▶ Packs released = 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. ▶ Total complaints = All complaints received in the reporting period, related to the quality of products manufactured in the site, regardless whether subsequently confirmed or not. All complaints received by the site should be counted, even if a complaint affects more than 1 site, or if eventually the root cause analysis attributes the issue to another site. Complaints related to lack of effect should be counted as well.
Critical complaints rate = All critical complaints, normalized by the number of packs released	▶ Critical complaints = Post-distribution product quality complaints which may indicate a potential failure to meet product specifications, 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.

Appendix 1

OOS Rate, Stability Failure, Invalidated (Unconfirmed) OOS Rate

Metric	Definition
Confirmed OOS rate = Total confirmed OOS (test results that fall outside the specifications or acceptance criteria), out of all lots dispositioned by the lab during the period	▶ Total lots tested/dispositioned by the lab = 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: – Lots for release testing (counted as 1 lot, even if sampled separately for chemical and microbiological testing, or for in-process analytical testing in lab or on shop floor). – Lots of incoming materials for analytical testing (count 1 per each analytically tested raw material and/or packaging material lot). Includes water used as raw material. – Lots for stability testing in that period (counted as 1 per each timepoint and condition sampled per the approved stability protocol). – Does not include environmental monitoring samples. ▶ Confirmed OOS = all test results that fall outside the specifications or acceptance criteria established in drug applications, drug master files (DMFs), official compendia, formulary or applied by the manufacturer when there is not an 'official' monograph.
Stability failure rate = Total confirmed OOS related to stability testing	▶ Subset of the Confirmed OOS rate – based on stability lots tested and confirmed OOS related to stability only.
Invalidated (unconfirmed) OOS rate = Total unconfirmed OOS, out of all lots tested during the period	▶ Unconfirmed OOS = all OOS minus confirmed OOS (see the definition of confirmed OOS).

US recall events (Total and by Class)

Metric	Definition
Recall rate	▶ Recall events = all US market recall events. ▶ By class = all US market recall events, class I and II. ▶ Recalled lots = Include lots recalled either voluntarily or by regulatory order (recall implies physical removal of product from field, not just a field action or correction). Include US market recalls only.

Appendix 1

Right First Time (Rework/Reprocessing)

Metric	Definition
RFT (rework/reprocessing rate) = Total lots released that have not been reworked or reprocessed out of the total finally released lots for commercial use in the period	▶ Total lots released (accepted) = total lots dispositioned less total lots rejected (see the definition of Lot acceptance rate). ▶ Total lots reworked or reprocessed = all lots that have gone through rework (using alternative process) or reprocessing (using again the original process) before that final disposition in order to meet requirements for release. Only count rework or reprocessing necessitated by quality issues (for example contract manufacturing sites should exclude rework due to customer order changes). If a lot was sent for rework and received a new lot number, it should still be counted as undergone rework when finally dispositioned.

APQRs on Time

Metric	Definition
APQR completed on time = Number of Annual Product Quality Reviews in the period that were completed by the original due date, normalized by all products subject to APQR	▶ Products subject to APQR = Total number of products subject to Annual Product Quality Reviews - annual evaluations of the quality standards of each drug product to verify the consistency of the process and to highlight any trends in order to determine the need for changes in drug product specifications or manufacturing or control procedures (as required by CFR Sec. 211.180, General requirements, section (e) and ICH Q7, GMPs for APIs, section 2.5 or EU Guidelines for Good Manufacturing Practice for Medicinal Products for Human and Veterinary Use, Chapter 1, Pharmaceutical Quality System, section 1.10). Does not include the data packages that a site prepares to its customers when acting as a CMO. ▶ Number of Annual Product Quality Reviews on time = completed by the original due date.

Appendix 1

Recurring Deviations Rate

Metric	Definition
Recurring deviations rate = Number of deviations that have re-occurred during the preceding 12 month period out of all closed deviations	▶ 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. ▶ Recurring deviations = Number of deviations for which during the 12 month period preceding each deviation, at least one other deviation has occurred with the same root cause within the same process and/or work area. If redundant/duplicative processes or equipment exist, please consider deviation events common to the grouping/work center as recurring (still within the 12 month timeframe). 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 (i.e. as 2 same deviations, rather than 1 different for each line).

CAPA Effectiveness Rate

Metric	Definition
CAPA effectiveness rate = Number of CAPAs effective out of all CAPAs with effectiveness check in the reporting period	▶ CAPAs with effectiveness check = Number of CAPAs evaluated for effectiveness in the reporting period. All CAPAs should be counted, including those related to inspection or audit observations. ▶ CAPAs effective = those evaluated CAPAs where the quality issue subject of the CAPA was resolved, and/or has not reoccurred, and there have been no unintended outcomes from the CAPA implementation.

Media Fill Failures (sterile/aseptic only)

Metric	Definition
Media fill rate = Number of media fills dispositioned as successful out of all media fills to support commercial products dispositioned during the period	▶ Media fills = Total number of media fills (regardless of number of runs in each) to support commercial products that were dispositioned (as successful or failed) during the period. If the media fill was dispositioned as failure and a rerun was needed, that repeat is counted as a separate media fill. Includes all media fills - both for initial and periodic qualifications. ▶ Successful media fills = All media fills that were not dispositioned as failures.

Appendix 1

Environmental Monitoring (sterile/aseptic only)

Metric	Definition
Environmental monitoring ▶ Lots with action limit excursions, normalized by all sterile dispositioned lots ▶ Lots rejected due to environmental monitoring reasons, normalized by all sterile dispositioned lots	▶ Sterile dispositioned lots during the period (see definition for Lot acceptance rate). ▶ Lots with limit excursions = All sterile dispositioned lots during the period that had associated investigations related to exceeding environmental monitoring action limits. If a lot had more than 1 such investigation please count only 1 per lot. If an investigation has affected multiple lots, please count each lot separately. Action limit is an established microbial or airborne particle level that, when exceeded, should trigger appropriate investigation and corrective action based on the investigation. ▶ Rejected lots due to environmental monitoring reasons = All sterile dispositioned lots during the period that were rejected for exceeding environmental monitoring action limits. 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.

Deviations Rate

Metric	Definition
Deviations rate = Number of deviations closed during the period out of the total finally dispositioned lots for commercial use in the period	▶ 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. 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. ▶ Deviations rate = Number of deviations closed during the period out of the total finally dispositioned lots for commercial use in the period. ▶ Total lots 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. 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.

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 Questions

- Do you measure that the process remains in a state of control (the validated state) during commercial manufacturing? (yes/no)
- 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? (open question)

Please indicate which metric or metrics do you use for ongoing monitoring and to what parameters do you apply them	**CpK**	**Ppk**	**Tolerance interval**	**Box Plots**	**Trending of CQAs**	**Other (specify which in the comments field)**
Applied to CQA (critical quality attributes tested in the lab)						
Applied to IPC (in-process control) checks						
Applied to CPP (critical process parameters)						

Appendix 3

Examples for Site and Product Data Collection Templates

Baseline data

			Retrospective					
			Apr 2013	May 2013	Jun 2013	Jul 2013	Aug 2013	Se
Production volume in units	[SDUs]	Number of single dosage units produced in the period. One unit is typically 1 pill, 1 capsule, 1 vial, 1 syringe, etc.						
Packs released	[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						
Lots 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. 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.						
Lots tested - total	[#]	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: - lots for release testing (counted as 1 lot, even if sampled separately for chemical and microbiological testing, or for in-process analytical testing in lab or on shop floor) - lots of incoming inspection materials for analytical testing (count 1 per each analytically tested raw material and/or packaging material lot). Includes water used as raw material. - lots undergoing stability testing in that period (counted as 1 per each timepoint and condition sampled per the approved stability protocol) Does not include environmental monitoring samples.						

Data by product

For each of your site products (as defined in line 25 in Data by site tab) please provide all metrics for the reporting year (1 total annual number)	Total complaints for the product for the reporting year *(as defined in line 38 in Data by site tab)*	Total critical complaints for the product for the reporting year *(as defined in line 39 in Data by site tab)*	Total packs released for the product for the reporting year *(as defined in line 17 in Data by site tab)*	Total lots dispositioned for reporting year *(as defined in line 18 in Data by site tab)*	
Total for all products	0	0	0	0	
There is no need to use actual product names, feel free to use e.g. internal codes or any other coded names for products	match	match	match	match	
Product names	Complaints	Critical complaints	Packs Released	Lots dispositioned	

Appendix 4

Case Study Company A

Introduction

Company A, a participant in the ISPE Quality Metrics Wave 1 Pilot, has a large portfolio of over-the-counter (OTC) products manufactured and distributed globally:

- 8,000+ raw material lots tested per year.
- 20,000+ intermediate bulk, packed, and labeled products tested annually.
- 750+ million packs produced and released.

Given the extent of site-level activity, one of Company A's motivations for participating in a quality metrics program is its potential to reduce inspection frequencies. Unlike other pilot participants, however, Company A will not benefit from expedited approvals or improved efficiency in post-approval changes due to the nature of its OTC- and monograph-based business.

Company A also understands that collecting quality metrics may help reduce drug shortages. Indeed, the ISPE Drug Shortages Prevention Plan [14] states "well-defined metrics tailored to proactively identify the potential risk of a shortage, will help mitigate looming shortages." The plan acknowledges that a prescriptively selected set of metrics may not always identify drug shortages, however. For this reason, it recommends that each metrics program should be tailored to the supply chain situation it addresses.

Finally, because developing a healthy corporate culture is a particular focus for Company A, it knows that a well-designed set of performance metrics helps identify improvement opportunities.

Pilot Experiences

Company A uses a detailed set of operational performance metrics. The definitions of some of these metrics, however, differ from those used in the Wave 1 Pilot. These differences required Company A to perform additional work so it could report its data in the Wave 1 Pilot standardized format.

Metrics reported by Company A (and all Wave 1 Pilot participants) were:

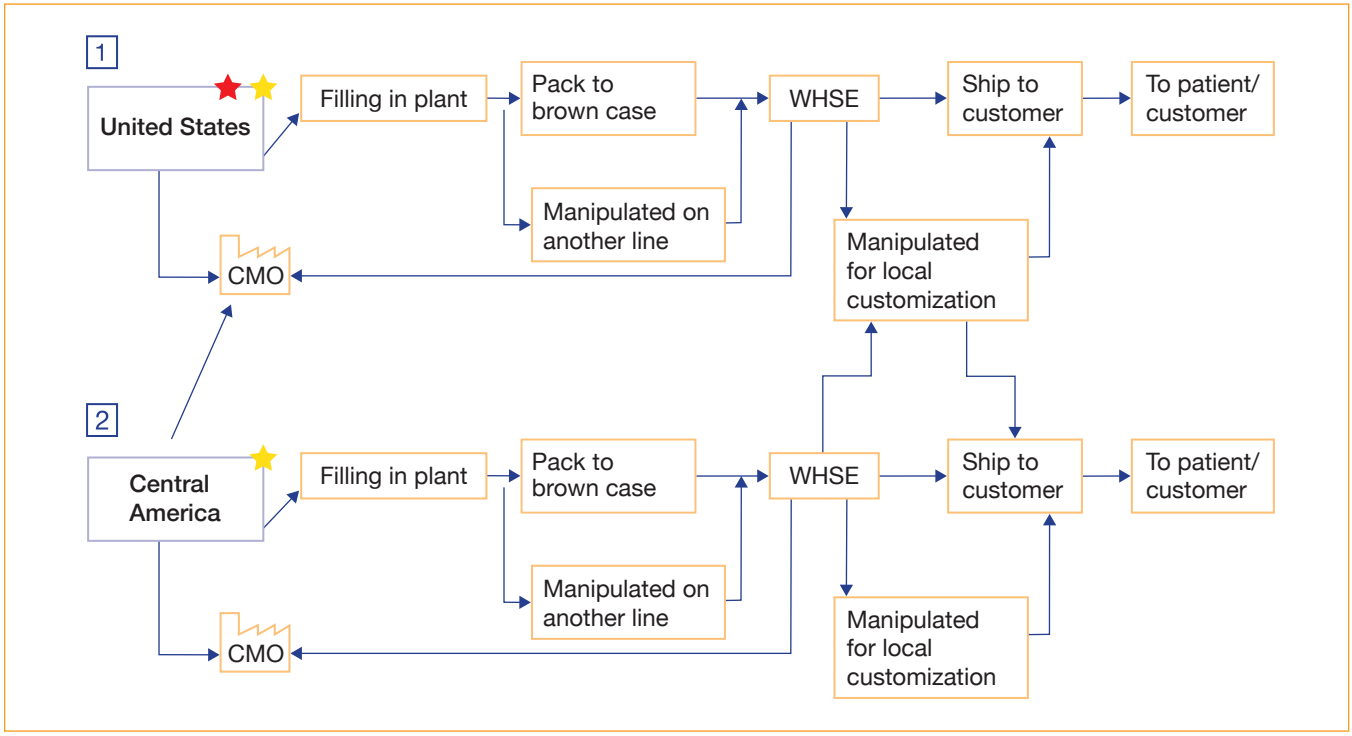
- Data mined retrospectively from the previous 12-month period.
- Measured prospectively for the current 3-month period.
- Evaluated at an operational/site level and (where required) aggregated by product/formula.
- Reported to McKinsey on a "good enough" basis; data was not subjected to the rigorous review and checking required for official submission to FDA.

Appendix 4

Understanding Supply Chain Complexity

Company A operations involve significant supply chain complexity, as shown in Figure A1.

Figure A1: Supply Chain Complexity



WHSE: Warehouse
Customer: Pharmacy, grocery chain or consumer store

Supply chains for two different formulas are shown in Figure A1:

- A red formula (red star) is a bulk drug product made in a manufacturing facility in Country 1 (United States).
- A yellow formula (yellow star) is a bulk drug product made in both Country 1 (United States) and Country 2 (Central America).

Appendix 4

As the figure shows, packaging, labelling, secondary packaging and customizing operations add significant complexity:

- The yellow formula can be transported in bulk to a contract manufacturing organization (CMO) for primary packing, secondary (outer) packaging, and in some cases customization.
- The yellow formula could also be sent to one of Company A's existing plants in either country to be filled, packaged (primary), dispositioned, and sent to the customer.
- The primary package could also be moved to another line for manipulation or customization.
- Adding secondary packaging into an outer carton and/or brown case could also include being shipped to a contract manufacturer to be customized.
- Customization can be very complex so that for example, a 24 pack case could consist of many product/pack variants – 24 different formulas, 2 packs of 12 formulas, 6 packs of 4 formulas, or 8 packs of 3 formulas
- A product could be shipped from a location in Country 2 to a location in Country 1 to be moved down the customer supply chain.
- Each primary package has a lot code and expiry date to allow full traceability.

This complexity relates directly to how data was reported by Company A in the Wave 1 Pilot: Because the back end of the supply chain includes variation-based choices of the final disposition point (i.e. data at the consumer pack level depends on where the final disposition point in the supply chain is allocated), plant-level and/or aggregated formula-level metrics that are normalized on a per-pack basis can be affected. Understanding this is critical to ensuring that metrics are captured and reported accurately.

Appendix 4

Supply Chain Complexity in Quality Metric Reporting: Challenges

Given the potential range of primary packs that a secondary (dispositioned) pack might contain for Company A, allocating a specific complaint to a particular product and/or supply chain is particularly difficult.

Assigning consumer complaints would depend on where the final disposition was designated and where the consumer purchased the 24-, 12-, or six-pack case.

For these and other reasons, complaint rate results in the Wave 1 Pilot are normalized for a product based on the dispositioned lot.

Other Product-Based Complexity Challenges

In addition to supply chain complexity, Company A also has product line complexity. While it is relatively easy to get information and quality metric data on product families prepared using a base formula differentiated by flavor, for example, this becomes much more complicated when there are multiple formulas within one family.

Company A's annual product quality reviews (APQRs) are bracketed by formula families that have a common base. One base formula with a particular flavor can be packed into many secondary-pack variations. This makes assigning product-based metrics extremely difficult.

Disaggregating quality metric data (complaints rate, for example) to products at the unique formula level or to a particular manufacturing site in a supply chain is extremely challenging; existing information technology (IT) systems at Company A do not currently facilitate this.

Conclusion

Company A has a large product range and very complex supply chains. This makes assigning metric data to a product level extremely difficult and time-consuming. Changing IT systems to a standardized set of metrics that could produce product-level data would require significant investment.

Appendix 5

Detailed Analysis of Data and Relationships for Each Individual Metric

Lot acceptance rate

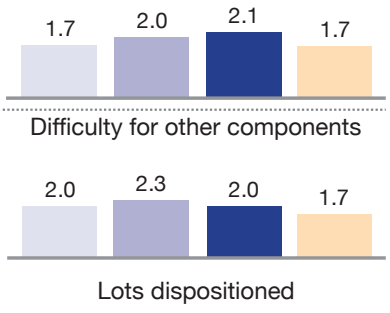
Lot acceptance rate defined as difference between 100% and ratio of rejects vs. all dispositioned lots

Lot acceptance rate correlates with:
Culture Deviations recurrence Critical complaints Rework

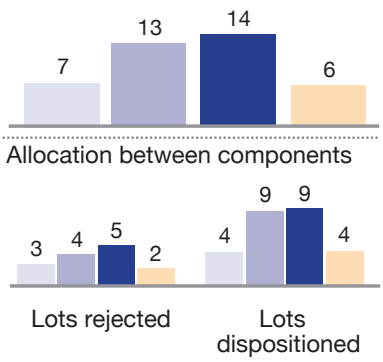
Preferred frequency of reporting

MONTHLY

Effort difficulty
Rejects difficulty[1]



Time consumed [h][2]
Overall



Comments

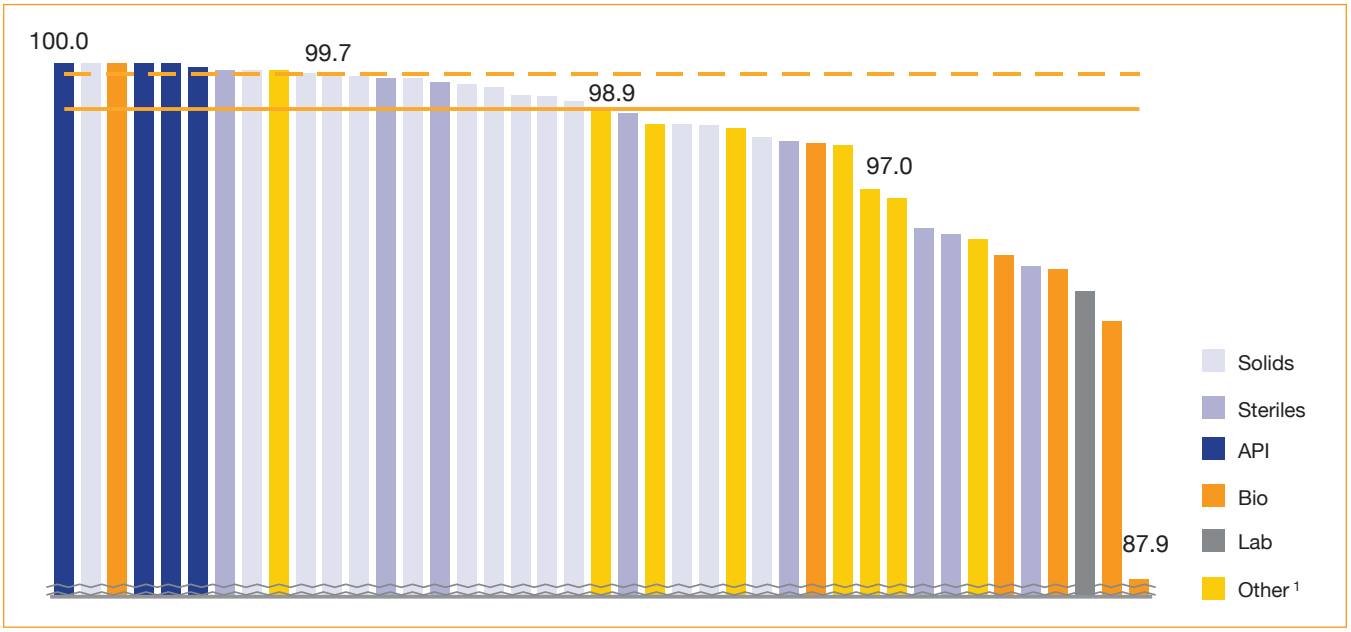
- Data pulling was not really hard but time consuming (manual pull) (few comments)
- We would like to have an option to nclude partial batch rejections

Rx
Gx
OTC
Drug subst.

[1] On a scale from 1 - easiest to 4 - most difficult
[2] Retrospective + current

Lot acceptance rate

% released out of all finally dispositioned lots



[1] Other includes Creams, Liquids and Other

Appendix 5

Confirmed OOS – product

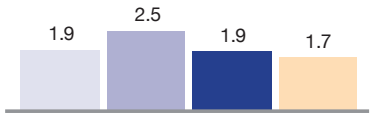
Confirmed OOS – product defined as (total confirmed OOS – confirmed stability OOS – cofirmed RM OOS)/(total lots tested – stability lots tested – RM lots tested)

Confirmed OOS – product correlates with:
Total complaints (lag) Critical complaints (lag)

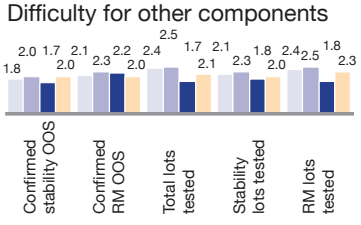
Preferred frequency of reporting

MONTHLY

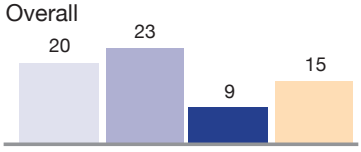
Effort difficulty
Total confirmed OOS difficulty[1]



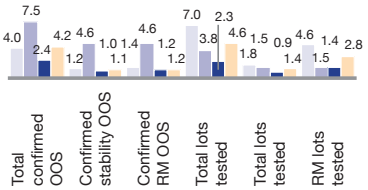
Difficulty for other components



Time consumed [h][2]
Overall



Allocation between components



Comments

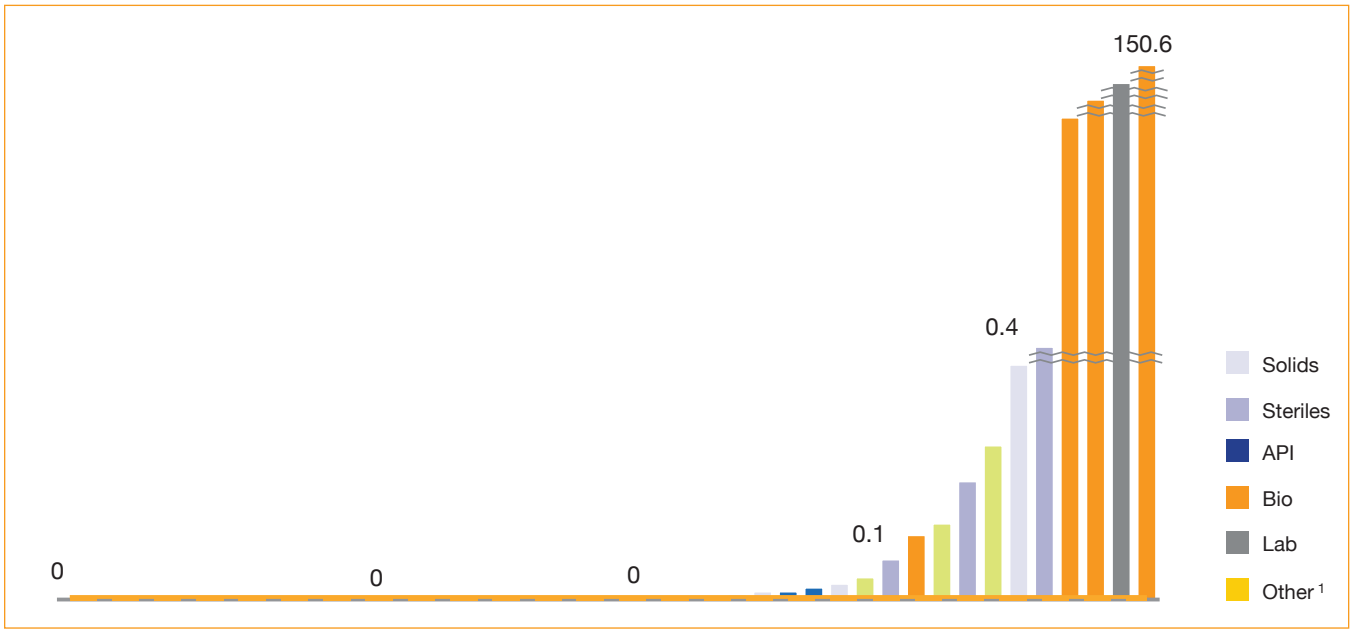
- Collecting separately release OOS and lots tested would simplify and speed up the process (2 components instead of 6 for KPI) (McK.)
- Numbers for lots tested get huge for multi-product plant or when lots of water is tested (few comments)
- Would like more clarity which lots to include in lots tested (e.g. placebo, micro samples)
- For OOS could be specifed, count when investigation closed' as only then it can be, confirmed'

Rx
Gx
OTC
Drug subst.

[1] On a scale from 1 - easiest to 4 - most difficult
[2] Retrospective + current

Confirmed RM OOS – product

No unit provided



[1] Other includes Creams, Liquids and Other

Appendix 5

Confirmed OOS – stability

Confirmed OOS – stability defined as ratio of confirmed OOS for stability to stability lots tested

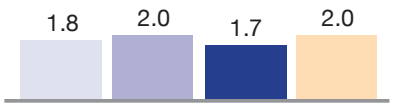
Confirmed OOS – stability correlates with:
-

Preferred frequency of reporting

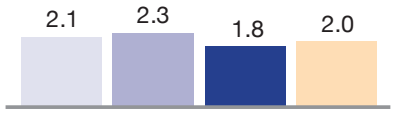
MONTHLY

Effort difficulty

Confirmed OOS stability difficulty[1]



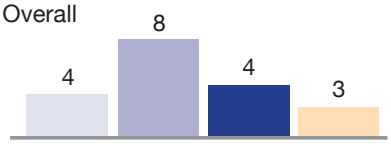
Difficulty for other components



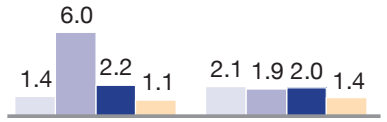
Lots dispositioned

Time consumed [h][2]

Overall



Allocation between components



Confirmed OOS stability

Lots tested stability

Comments

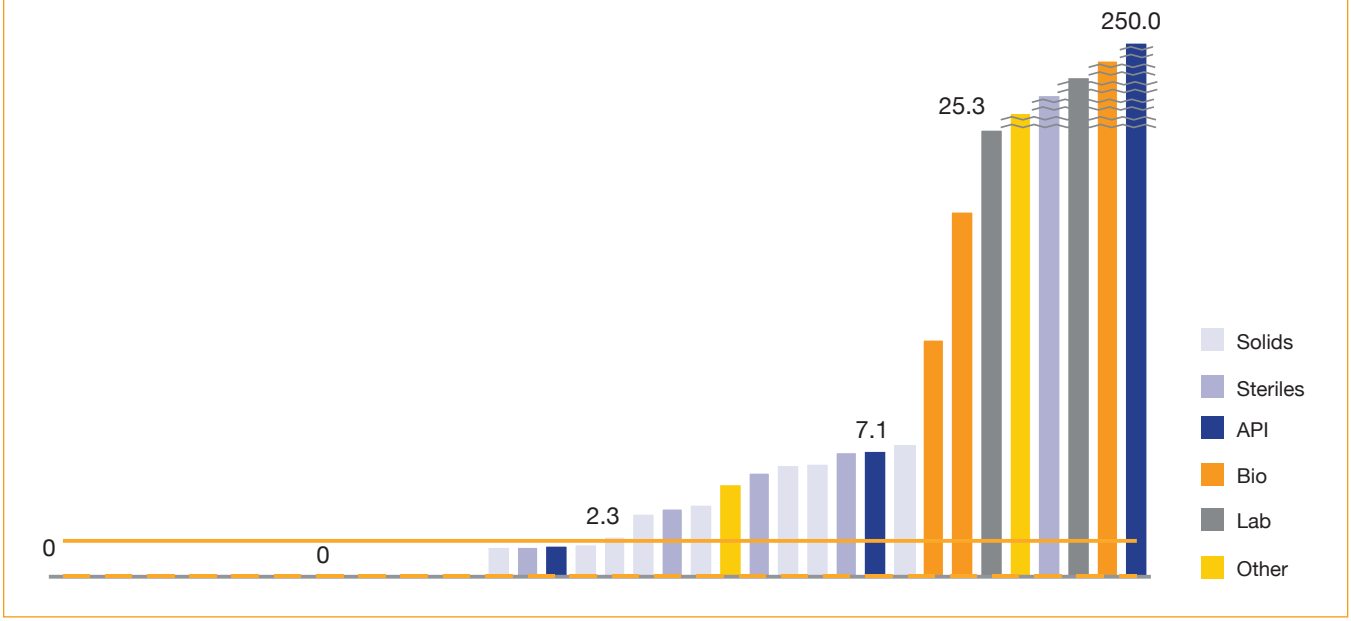
- Data for stability stored separately/ pulled manually (few comments)
- At least for sites with only few products it makes little sense to collect monthly (GMP requires 1 batch/ year/ product)

Rx
Gx
OTC
Drug subst.

[1] On a scale from 1 - easiest to 4 - most difficult
[2] Retrospective + current

Confirmed OOS – stability

Per 000' stability lots tested



[1] Other includes Creams, Liquids and Other

Appendix 5

Deviations rate

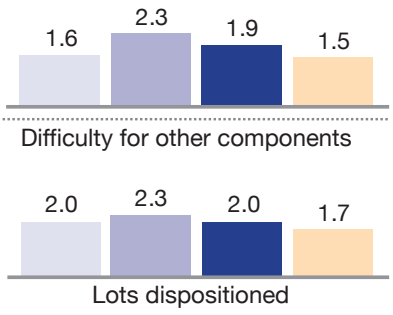
Deviations rate defined as ratio of deviations to lots dispositioned

Deviations rate correlates with:
Critical complaints

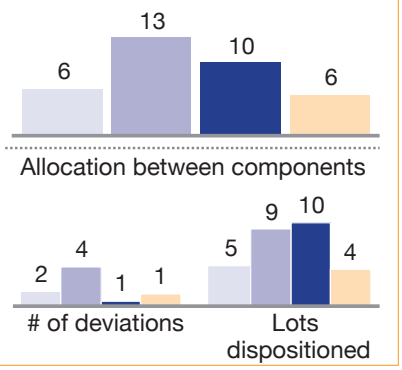
Preferred frequency of reporting

MONTHLY

Effort difficulty
# of deviations difficulty[1]



Time consumed [h][2]
Overall



Comments

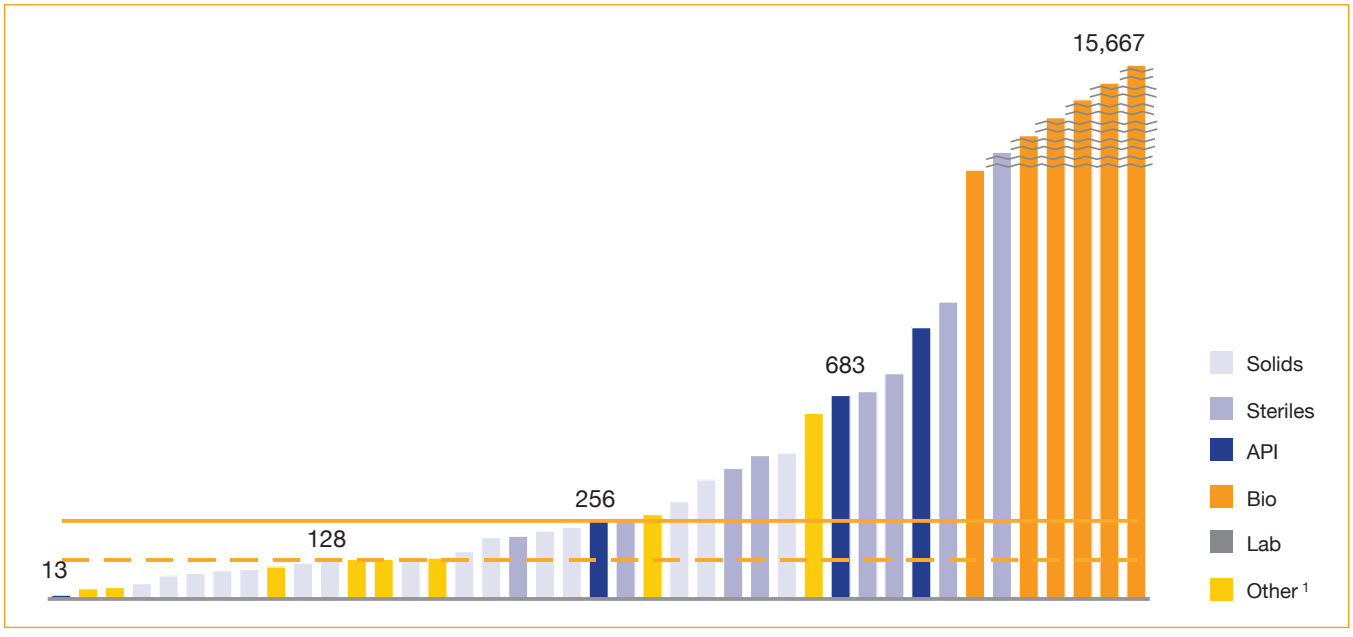
- Deviation def. is very broad and includes almost everything, should be narrowed, also adding OOS here confuses the picture
- Need more clarity if e.g. RM or buffer deviations are to be counted
- Issues with data pulling (e.g. separate systems for OOS and deviations and need for aggregation) (few comments)

Rx
Gx
OTC
Drug subst.

[1] On a scale from 1 - easiest to 4 - most difficult
[2] Retrospective + current

Deviations rate

Per 000' lots dispositioned



[1] Other includes Creams, Liquids and Other

Appendix 5

Rework rate

Rework rate defined as ratio of reworked and reprocessed lots vs. (lots dispositioned minus lots rejected)

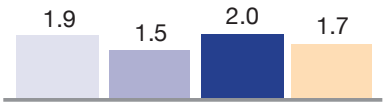
Rework rate correlates with:
Lot acceptance rate

Preferred frequency of reporting

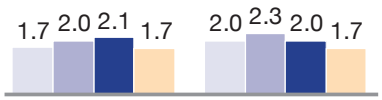
MONTHLY

Effort difficulty

Lots reworked difficulty[1]

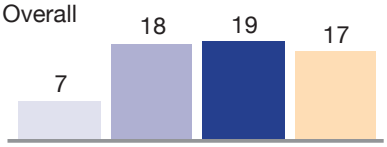


Difficulty for other components

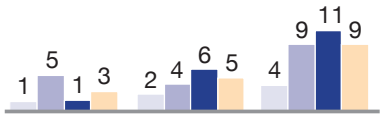


Time consumed [h][2]

Overall

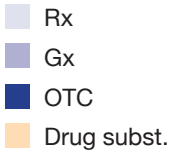


Allocation between components



Comments

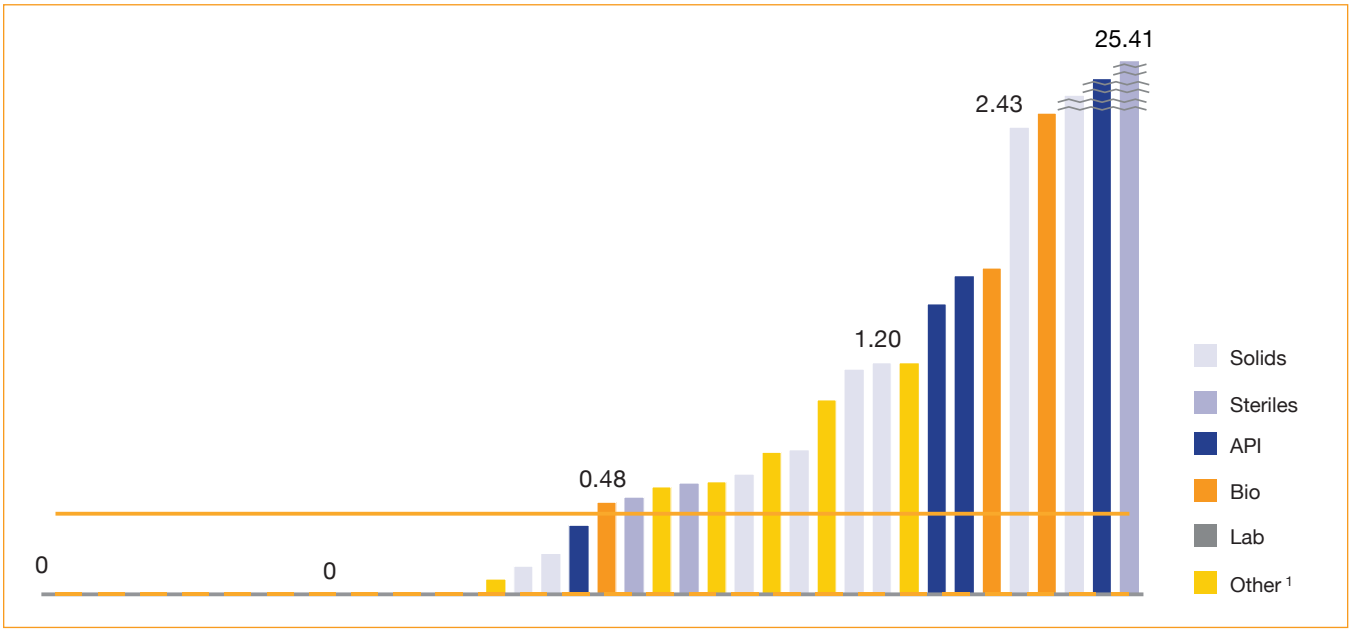
- We do not permit any rework at all (few comments)
- Need more clarity what rework or reprocess mean



[1] On a scale from 1 - easiest to 4 - most difficult
[2] Retrospective + current

Rework rate

% lots dispostioned



[1] Other includes Creams, Liquids and Other

Appendix 5

Confirmed OOS – raw materials

Confirmed OOS – raw materials defined as ratio of confirmed OOS for raw materials to raw materials lots tested

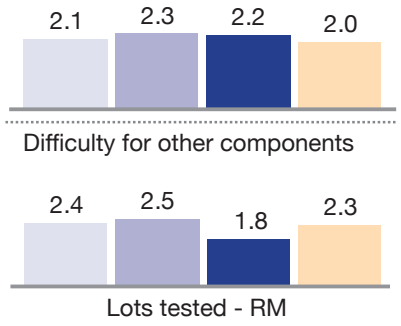
Confirmed OOS – raw materials correlates with:
-

Preferred frequency of reporting

MONTHLY

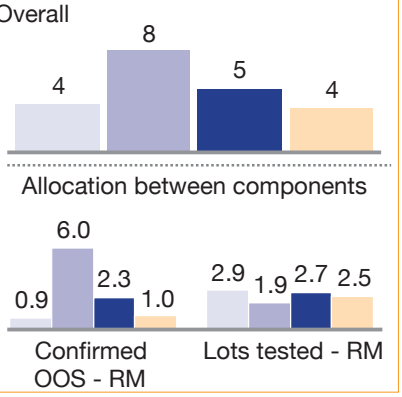
Effort difficulty

Confirmed OOS RM difficulty[1]



Time consumed [h][2]

Overall



Comments

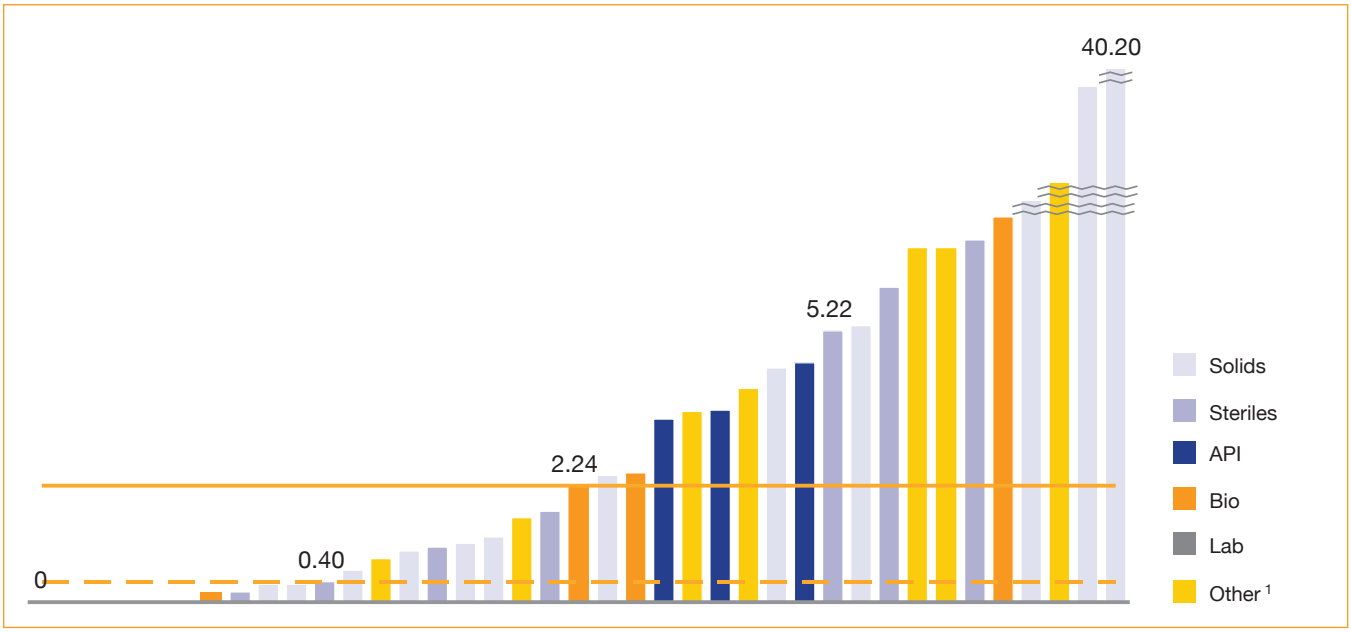
- 2 sites did not supply it
- Need more clarity what to include (e.g. which EM, why water is in and other EM not neccessarily) (few comments)
- Split between technologies and/or pharma/cosmetics businesses difficult (few comments)

Rx
Gx
OTC
Drug subst.

[1] On a scale from 1 - easiest to 4 - most difficult
[2] Retrospective + current

Confirmed OOS – raw materials

Per 000' RM/PM lots tested



[1] Other includes Creams, Liquids and Other

Appendix 5

Unconfirmed OOS

Unconfirmed OOS defined as ratio of unconfirmed OOS to total lots tested

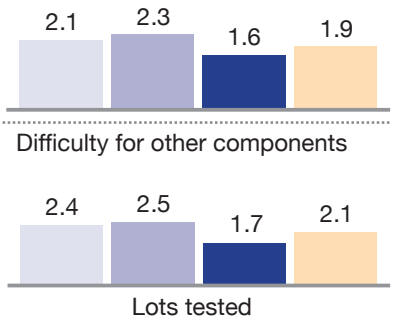
Unconfirmed OOS correlates with:
-

Preferred frequency of reporting

MONTHLY

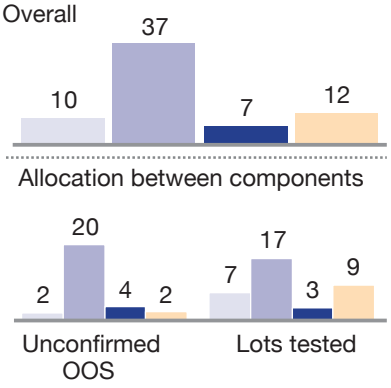
Effort difficulty

Unconfirmed OOS difficulty [1]



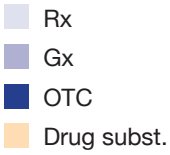
Time consumed [h][2]

Overall



Comments

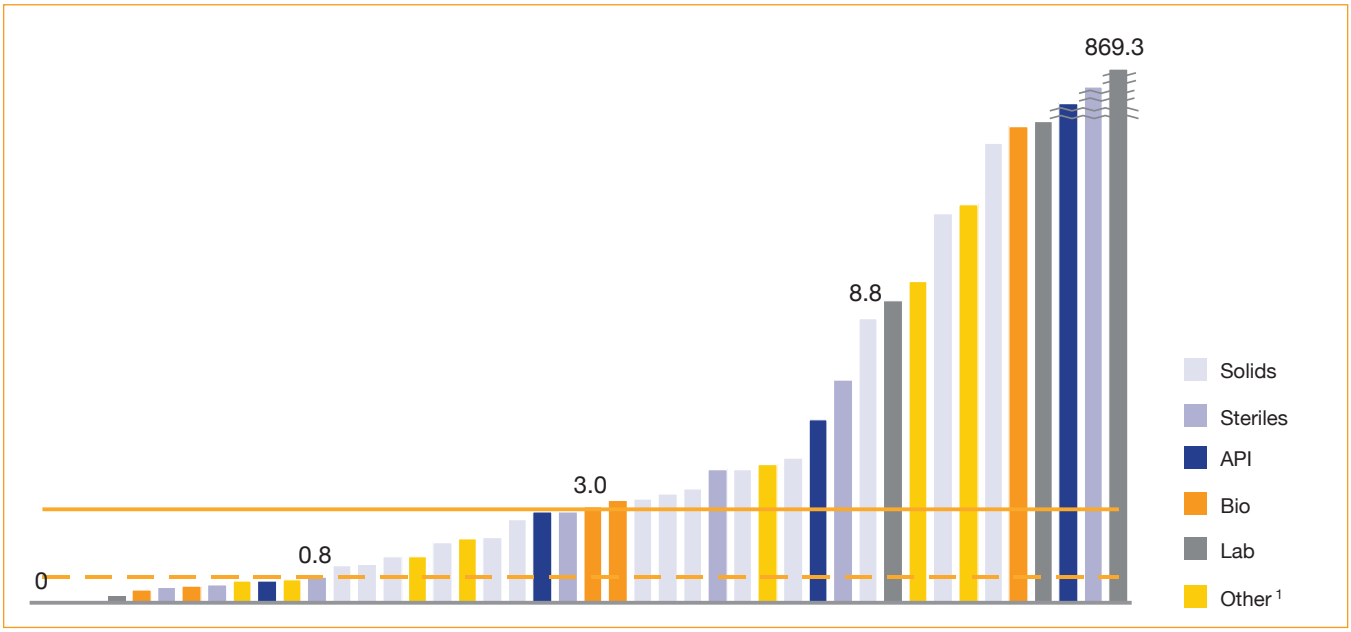
- Data is automatically available together with confirmed OOS (few comments)
- There was some manual aggregation needed due to scope (few comments)



[1] On a scale from 1 - easiest to 4 - most difficult
[2] Retrospective + current

Unconfirmed OOS

Per 000' lots tested



[1] Other includes Creams, Liquids and Other

Appendix 5

Recall events US

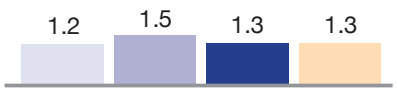
Recall events US

Recall events US correlates with:
Deviations rate Critical complaints

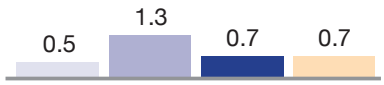
Preferred frequency of reporting

MONTHLY

Effort difficulty[1]



Time consumed [h][2]
Overall



Comments

- This is non-applicable for some drug substance sites
- Time was short as we had no recalls, not sure if this reflects situation with recalls
- We should not limit ourselves to US market

Rx
Gx
OTC
Drug subst.

[1] On a scale from 1 - easiest to 4 - most difficult
[2] Retrospective + current

Recall events US

of events

[1] Other includes Creams, Liquids and Other

Appendix 5

Recall events class I and II US

Recall events US class I and II US

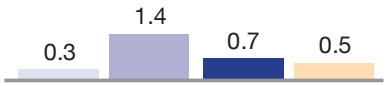
Recall events US class I and II US correlates with:
Not tested

Preferred frequency of reporting

MONTHLY

Effort difficulty[1]

Time consumed [h][2]
Overall



Comments

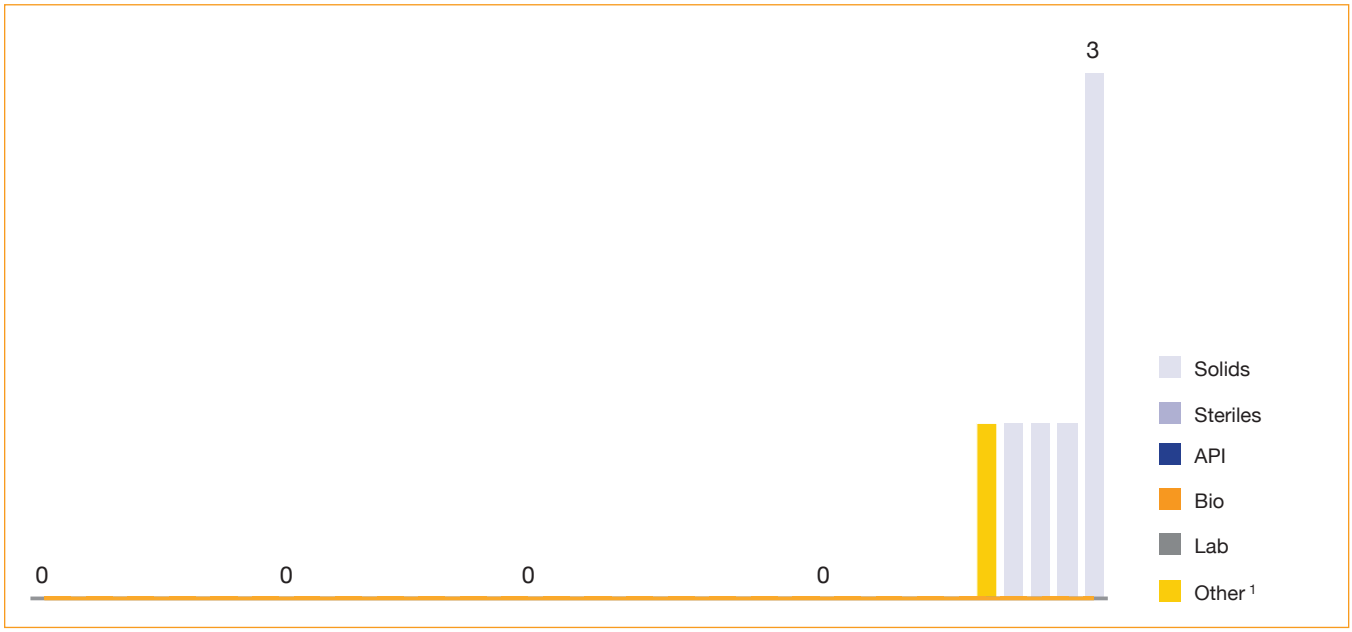
- This is non-applicable for some drug substance sites
- Time was short as we had no recalls, not sure if this reflects situation with recalls
- We should not limit ourselves to US market

Rx
Gx
OTC
Drug subst.

[1] On a scale from 1 - easiest to 4 - most difficult
[2] Retrospective + current

Recall events class I and II US

of events



[1] Other includes Creams, Liquids and Other

Appendix 5

Recalled lots US

Recalled lots US

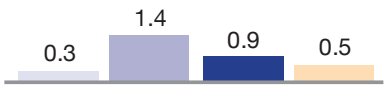
Recalled lots US correlates with:
Not tested

Preferred frequency of reporting

MONTHLY

Effort difficulty[1]

Time consumed [h][2]
Overall



Comments

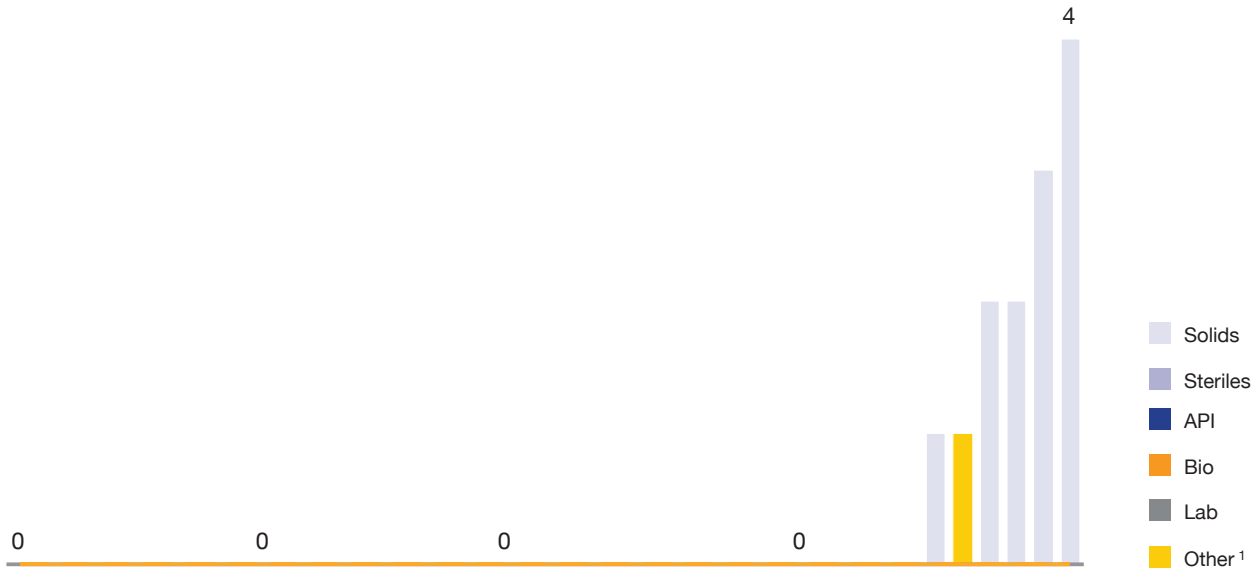
- This is non-applicable for some drug substance sites
- Time was short as we had no recalls, not sure if this reflects situation with recalls
- We should not limit ourselves to US market

Rx
Gx
OTC
Drug subst.

[1] On a scale from 1 - easiest to 4 - most difficult
[2] Retrospective + current

Recalled lots US

of lots



[1] Other includes Creams, Liquids and Other

Appendix 5

Total complaints rate

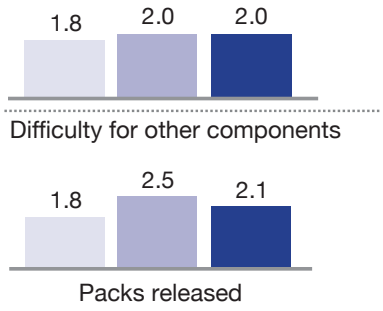
Total complaints rate defined as ratio of total complaints to total packs released

Total complaints rate correlates with:
Action limit excursions OOS product (lag)

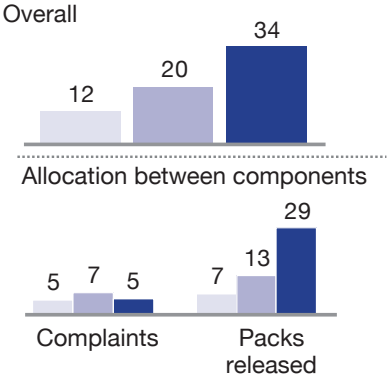
Preferred frequency of reporting

MONTHLY

Effort difficulty
Total complaints difficulty[1]



Time consumed [h][2]
Overall



Comments

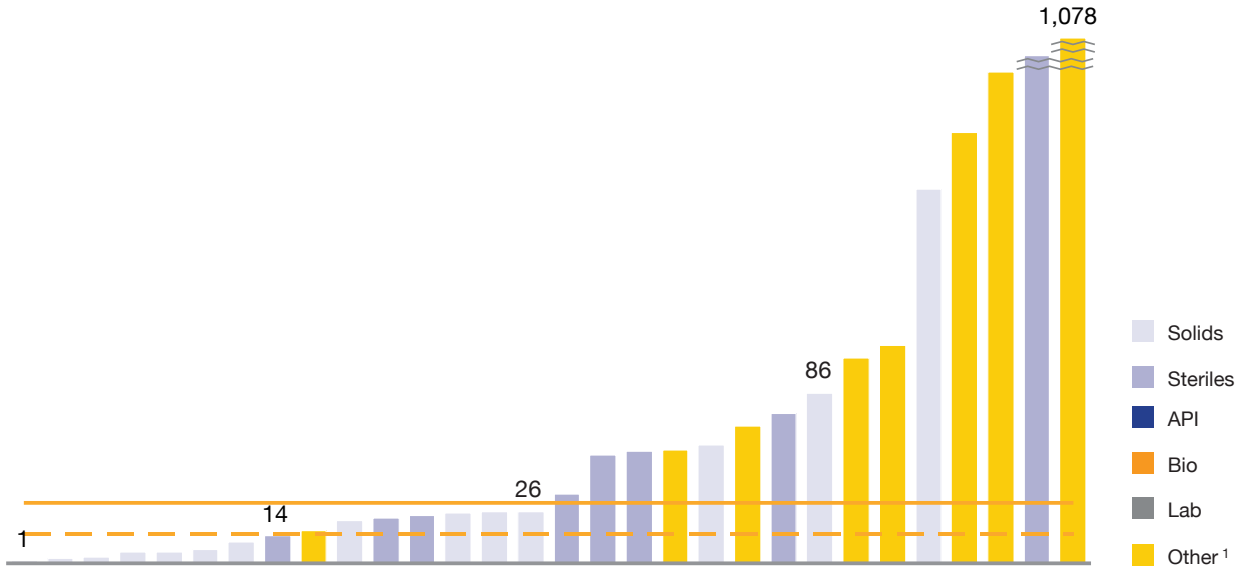
- 1 plant did not manage to collect packs data
- Had trouble to report packs due to different system set-up
- Difficult to split between technologies
- Time consuming due to many products (few comments)
- Some complaints cannot be allocated to product at all
- Need more clarity – e.g. whether include market only or also internal complaints, if bulk complaint should be reported by bulk of packaging site, how to compare complaints on bulk with packs (few comments)
- Little value added in that, confirmed complaints would be better

Rx
Gx
OTC

[1] On a scale from 1 - easiest to 4 - most difficult
[2] Retrospective + current

Total complaints rate

Per million packs



[1] Other includes Creams, Liquids and Other

Appendix 5

Critical complaints rate

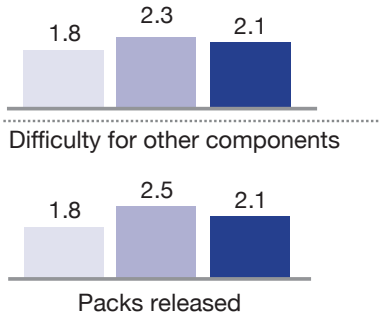
Critical complaints rate defined as ratio of total complaints to total packs released

Critical complaints rate correlates with:
Deviations rate Lot acceptance rate OOS product (with lag)

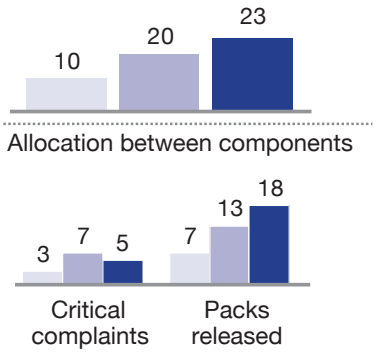
Preferred frequency of reporting

MONTHLY

Effort difficulty
Critical complaints difficulty [1]



Time consumed [h][2]
Overall



Comments

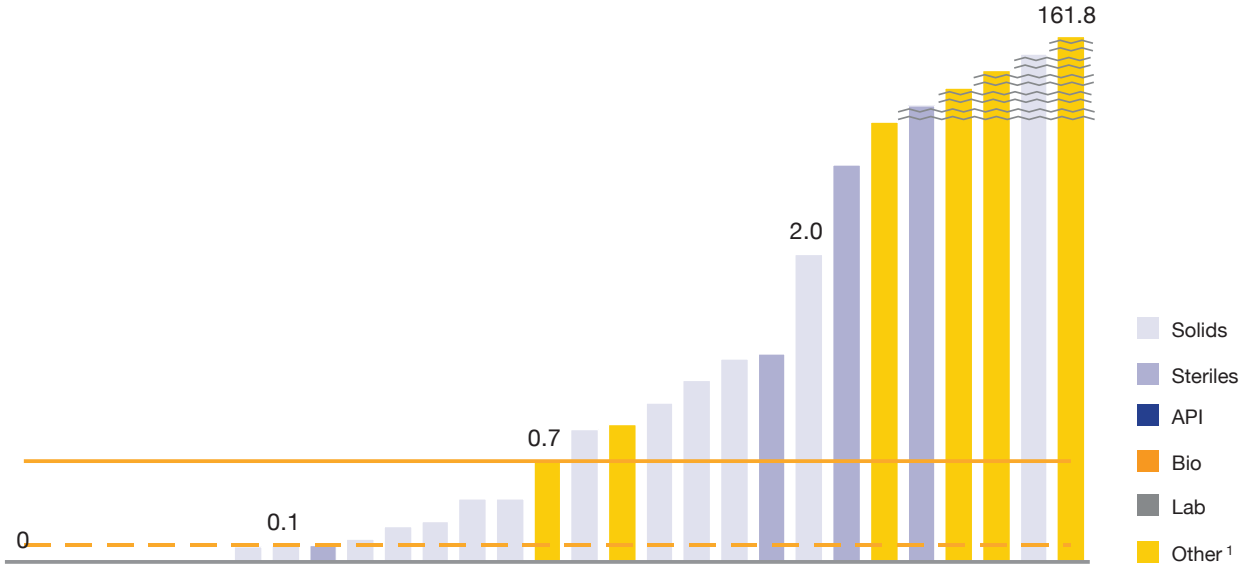
- 1 plant did not manage to collect packs data
- Had trouble to report packs due to different system set-up
- Difficult to split between technologies (few comments)
- Time consuming due to many products (few comments)
- Some complaints cannot be allocated to product at all
- If we have a complaint on bulk how do we report corresponding packs?
- Had to manually assess each complaint for criticality
- Little value added in that, confirmed complaints would be better (few comments)

Rx
Gx
OTC

[1] On a scale from 1 - easiest to 4 - most difficult
[2] Retrospective + current

Critical complaints rate

Per million packs



[1] Other includes Creams, Liquids and Other

Appendix 5

APQR on time

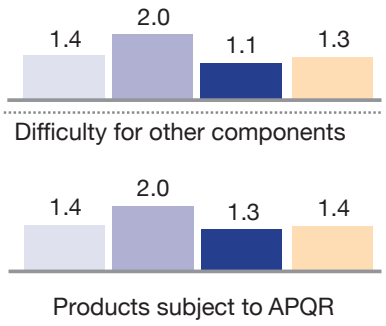
APQR on time defined as ratio of APQR on time to all APQR closed in reporting period

APQR on time correlates with:
CAPAs effective

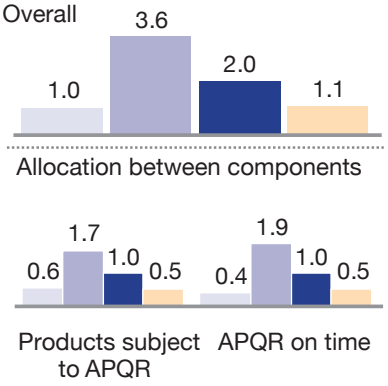
Preferred frequency of reporting

ANNUALLY

Effort difficulty
APQR on time difficulty[1]



Time consumed [h][2]
Overall



Comments

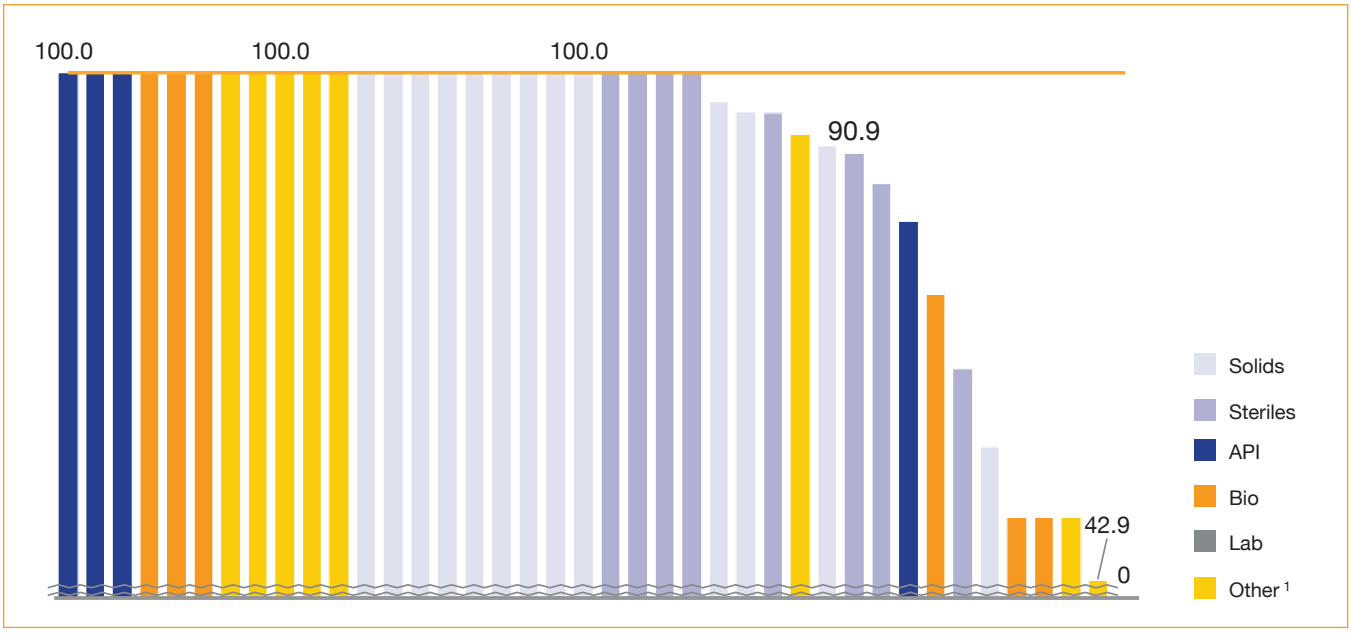
- We group our products for APQR (few comments)
- Period may need to be adjusted for this point, products evaluated are different than ones produced in the period

Rx
Gx
OTC
Drug subst.

[1] On a scale from 1 - easiest to 4 - most difficult
[2] Retrospective + current

APQR on time

Percent



[1] Other includes Creams, Liquids and Other

Appendix 5

CAPA effectiveness

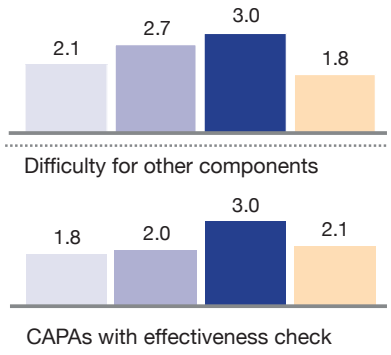
CAPA effectiveness defined as ratio of CAPAs evaluated as effective to all CAPAs evaluated for effectiveness in the period

CAPA effectiveness correlates with:
APQR on time

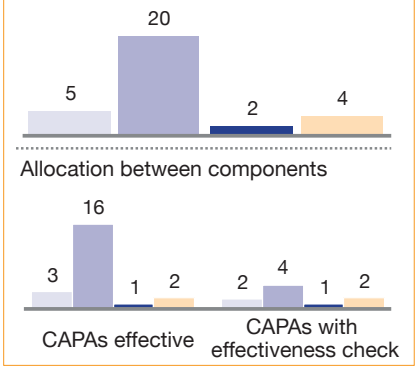
Preferred frequency of reporting

QUARTERLY

Effort difficulty
Effective CAPA difficulty[1]



Time consumed [h][2]
Overall



Comments

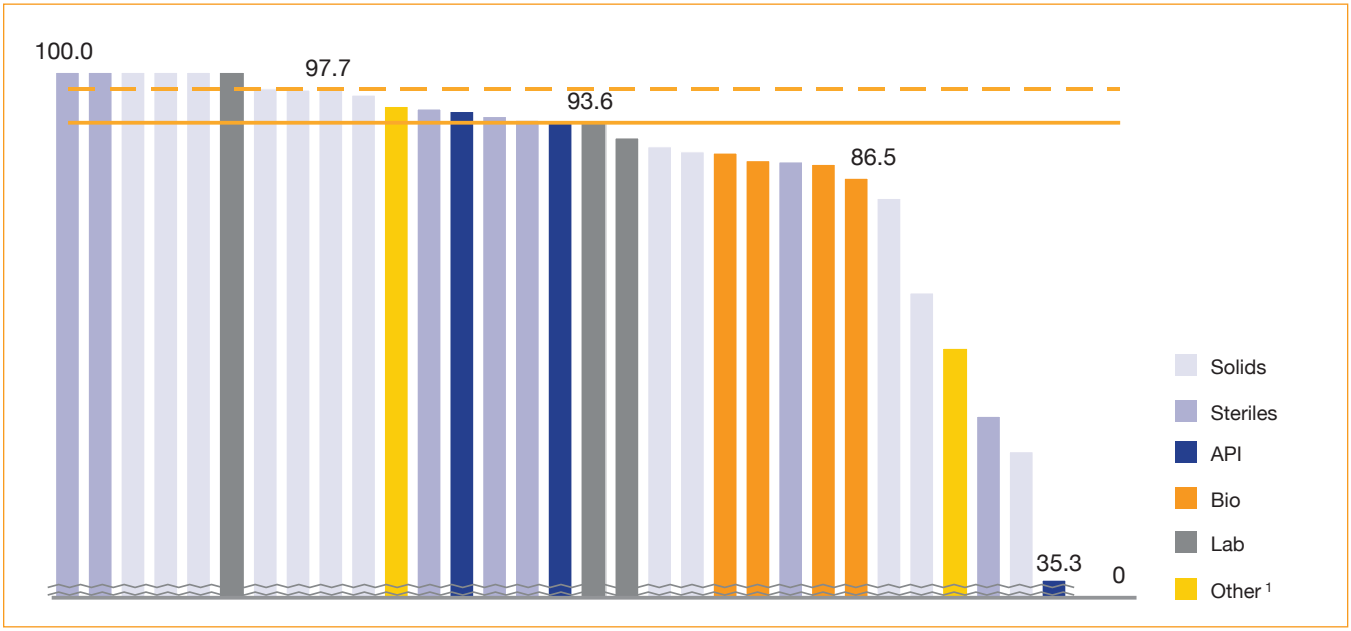
- 27% of plants did not collect enough data for this metric
- Manual extraction of data from e.g text files (few comments)
- We evaluate group of CAPAs as a whole and not single CAPAs
- Need clarity how to include partially effective CAPA
- Recurrence of the "quality issue" is secondary. CAPA target a specific root cause and to be effective should address the given root cause which subsequently prevents recurrence of the "quality issue"

Rx
Gx
OTC
Drug subst.

[1] On a scale from 1 - easiest to 4 - most difficult
[2] Retrospective + current

CAPA effectiveness

Percent



[1] Other includes Creams, Liquids and Other

Appendix 5

Recurring deviations

Recurring deviations defined as ratio of recurring deviations (with 1 year look-back period) to all deviations closed

Recurring deviations correlates with:

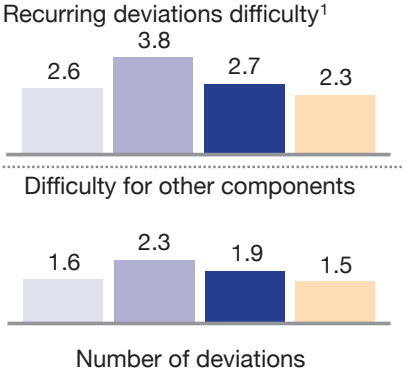
Lot acceptance Recalls Culture

Preferred frequency of reporting

MONTHLY

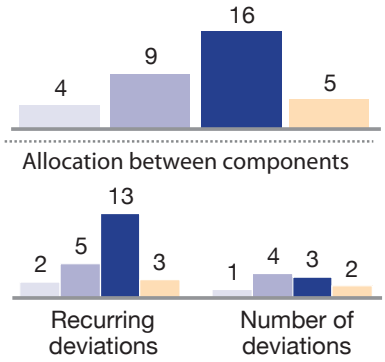
Effort difficulty

Recurring deviations difficulty[1]



Time consumed [h][2]

Overall



Comments

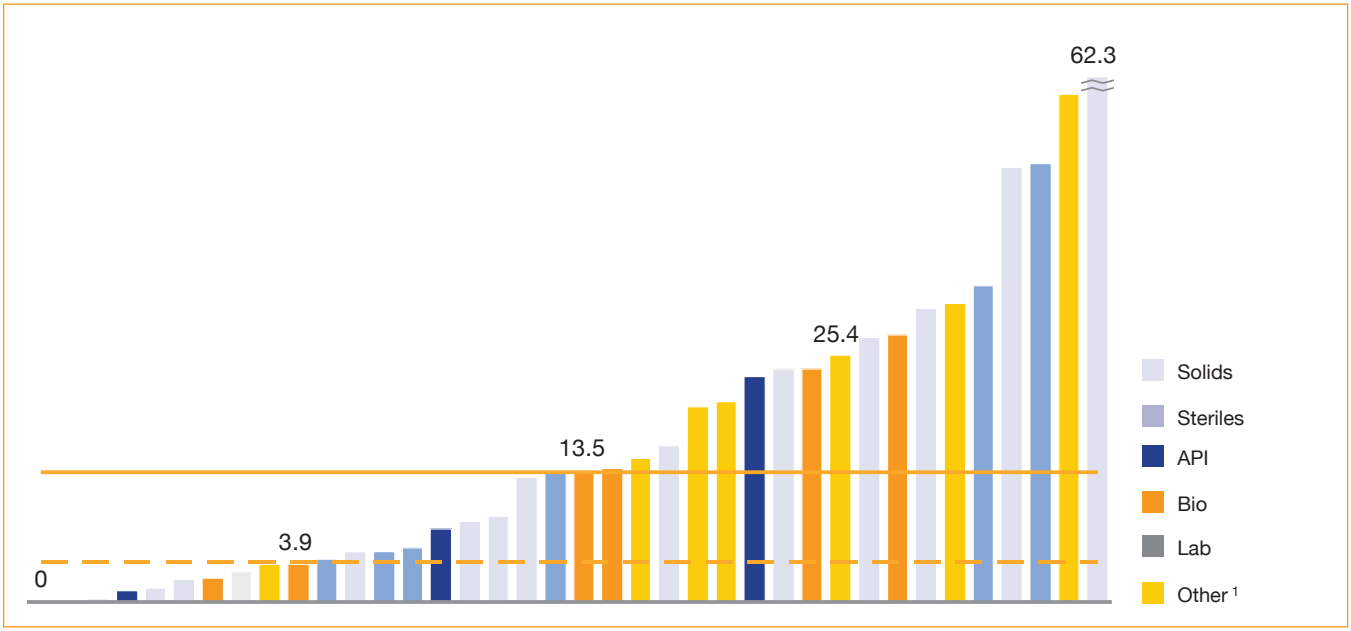
- 15% of sites did not report it at all
- Deviation def. is very broad and includes almost everything, should be narrowed, also adding OOS here confuses the picture
- Need more clarity if e.g. RM or buffer deviations are to be counted
- Issues with data pulling (e.g. separate systems for OOS and deviations and need for aggregation, manual checking) (few comments) recurring
- Deviation should be limited to the same deviation (same specific departure from approved process or instructions) and independent of "root cause". The given definition of "deviation" is so broad many things that meet that definition likely do not require root cause analysis

Rx
Gx
OTC
Drug subst.

[1] On a scale from 1 - easiest to 4 - most difficult
[2] Retrospective + current

Recurring deviations

Percent



[1] Other includes Creams, Liquids and Other

Appendix 5

Media fills successful

Media fills successful defined as media fills successful to all media fills

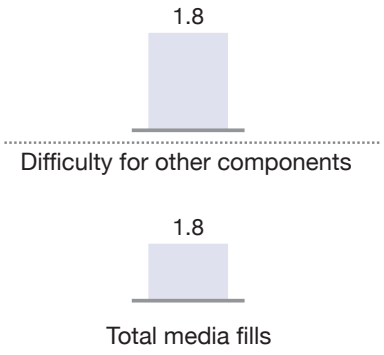
Media fills successful correlates with:

Preferred frequency of reporting

ANNUALLY

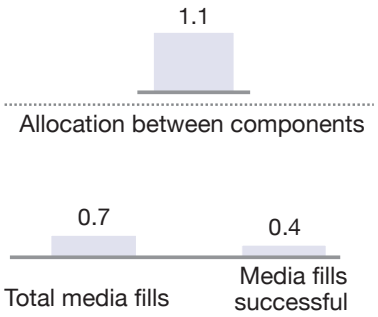
Effort difficulty

Media fills successful difficulty [1]



Time consumed [h][2]

Overall

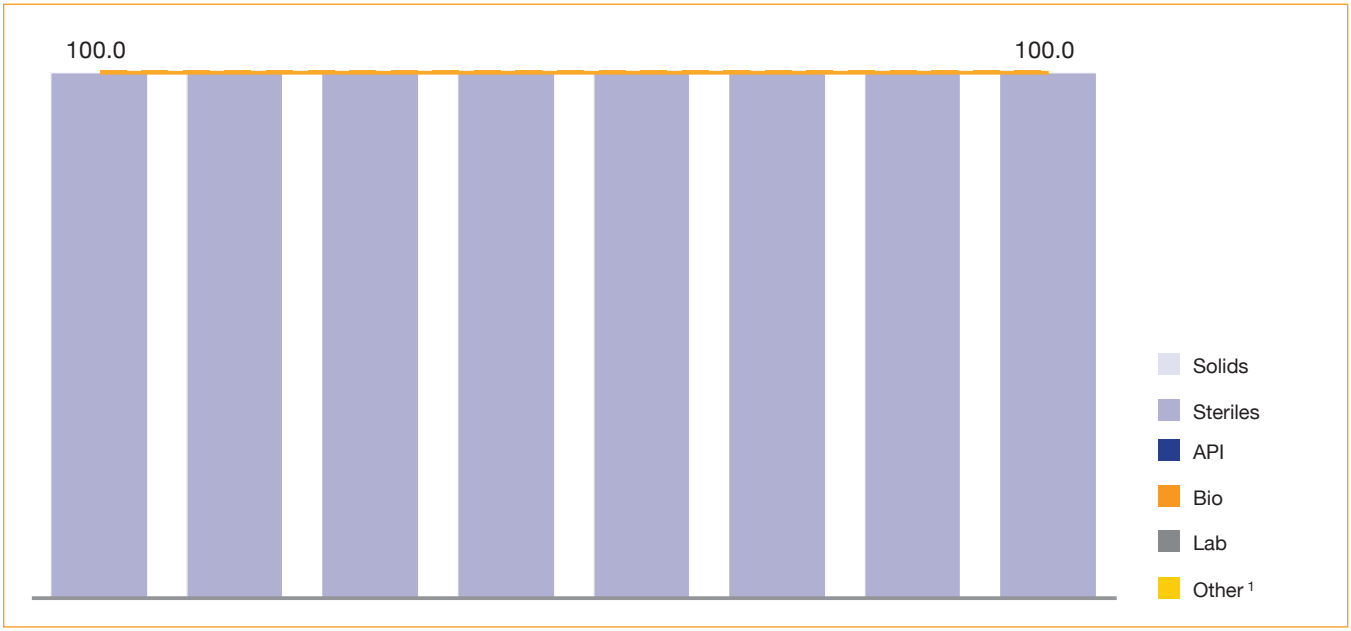


Rx

[1] On a scale from 1 - easiest to 4 - most difficult
[2] Retrospective + current

Media fills successful

Percent



[1] Other includes Creams, Liquids and Other

Appendix 5

Lots rejected due to action limit excursions

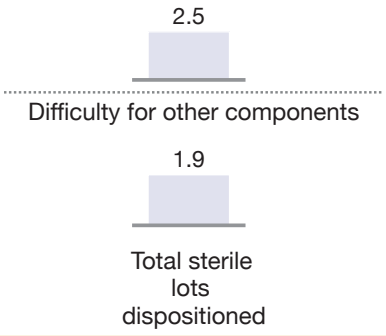
Lots rejected due to action limit excursions defined as ratio of lots rejected due to action limit excursions to total sterile lots dispositioned

Lots rejected due to action limit excursions correlates with:
-

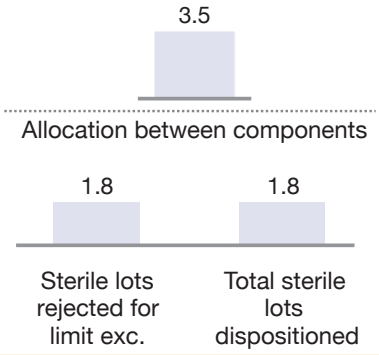
Preferred frequency of reporting

QUARTERLY

Effort difficulty
Sterile lots rejected for action limit - difficulty[1]



Time consumed [h][2]
Overall

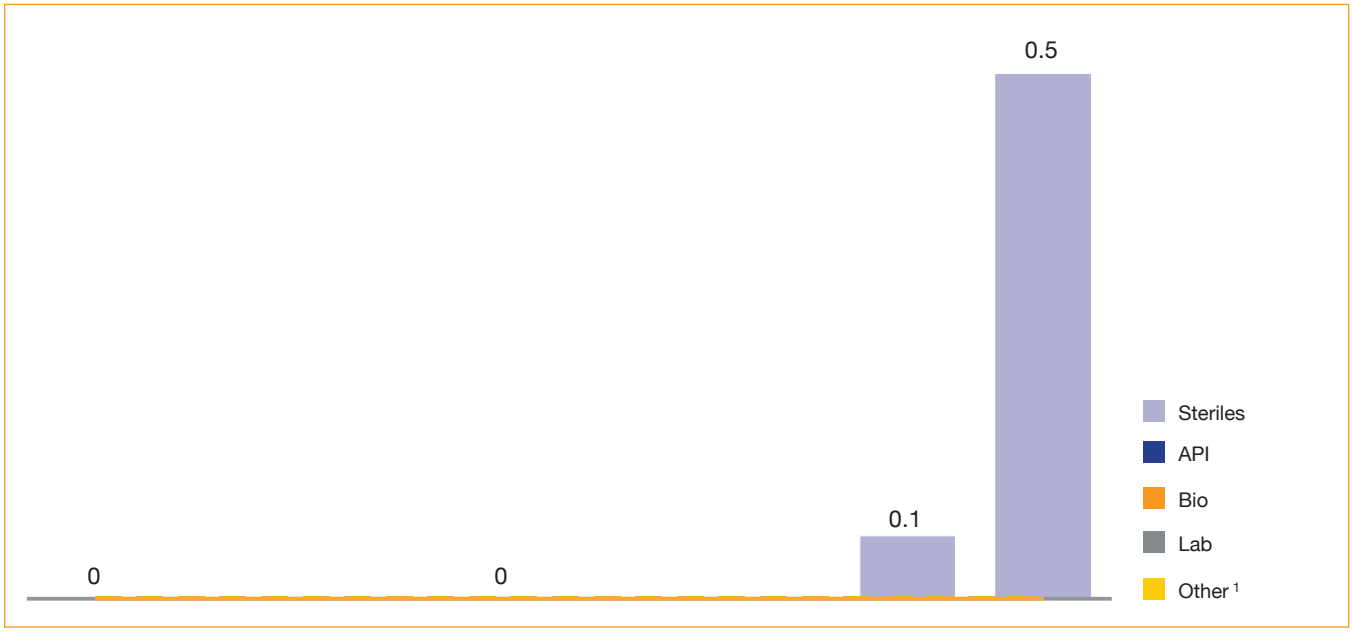


Rx

[1] On a scale from 1 - easiest to 4 - most difficult
[2] Retrospective + current

Lots rejected due to action limit excursions

Percent



[1] Other includes Creams, Liquids and Other

Appendix 5

Investigations related to action limit excursions

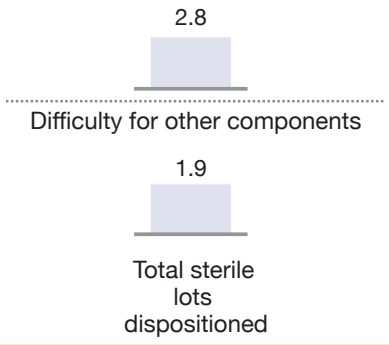
Investigations related to action limit excursions defined as ratio of lots with action limit excursions to total sterile lots dispositioned

Investigations related to action limit excursions correlates with:
Total complaints Culture (capabilities)

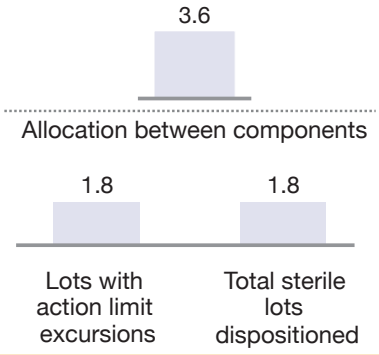
Preferred frequency of reporting

MONTHLY

Effort difficulty
Investigations difficulty[1]



Time consumed [h][2]
Overall

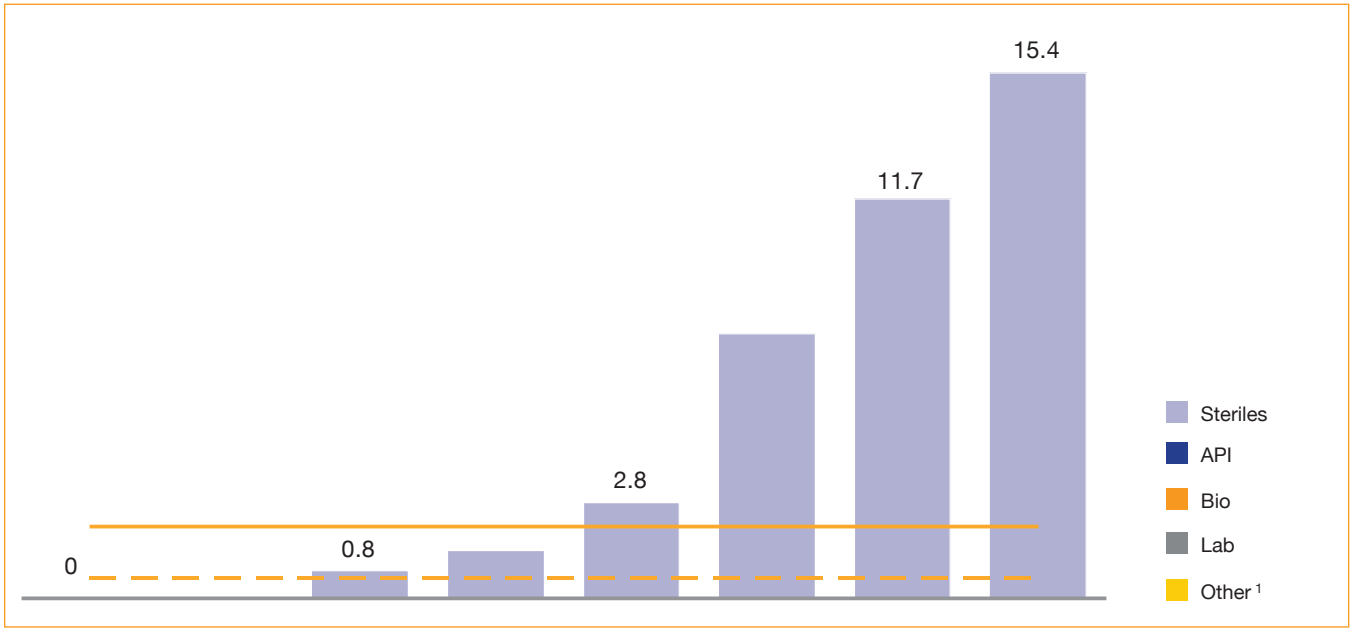


Rx

[1] On a scale from 1 - easiest to 4 - most difficult
[2] Retrospective + current

Investigations related to action limit excursions

Percent



[1] Other includes Creams, Liquids and Other