MDIC Case for Quality Program

Medical Device Quality Metrics

Best Practices Document for Metrics Identified Across the

Total Product Lifecycle

August 1, 2016

For more information on MDIC or Case for Quality please see www.MDIC.org/cfq

Table of Contents

Background

The FDA CDRH Office of Compliance launched the Case for Quality initiative in 2011 to explore with industry professionals how to shift the historical focus from one of compliance and enforcement action to one of device quality. This initiative was launched after an in-depth review of device quality data, which demonstrated a lack of improvement in the risk to patient safety and the total number of enforcement actions taken by FDA year after year. Through a series of national forums, CDRH engaged an array of stakeholders —industry, healthcare providers, patients, payers, academia, and investors — to collaboratively discuss ways to advance the industry by identifying and promoting practices that result in high-quality devices and adapting FDA regulatory approaches to align with those practices.

As a result of the FDA National Forum hosted by Xavier University, FDA and Xavier launched the Quality Measures Initiative in September 2014 with a Work Group comprised of industry professionals and FDA officials (Appendix A). The goal of this initiative was to identify ways for industry to proactively and predictively measure the risk to its own product quality, which would therefore enable industry to focus on improving product quality commensurate with the need. The outcome of the FDA/Xavier work was presented to FDA and the Medical Device Innovation Consortium (MDIC) in May of 2015, and consisted of 17 measures identified across the Total Product Lifecycle. In June of 2015, MDIC adopted the FDA/Xavier initiative and requested that a defined subset of measures be converted into metrics. The FDA/Xavier Work Group chartered Phase II of the measures initiative to convert three measures into quantifiable metrics – one from pre-production, one from production, and one from post-production. Once developed, these metrics would be piloted by volunteer organizations in order to analyze the effectiveness of the metrics to predictively measure risk to product quality. The Work Group recognized the need to develop Best Practices for each metric to describe how the metrics could be implemented in a way that would inform decisions and trigger action.

The intent of the FDA/Xavier work was to arm industry with practical metrics to implement commensurate with the needs of the business and complexity of the products, such that the Right-First-Time mentality could be shifted as close to the initial days of development as possible. Importantly, the metric calculations are provided as a guide for industry to adjust in a way that makes sense for the products and business in question, while maintaining the intent of the metric. The Best Practices described herein provide foundational principles to use when applying all three metrics, as well as a process for enterprise-wide continual improvement.

Effective Implementation Principles

INTERPRET IN CONTEXT: the goal of a robust metrics program is to help drive continual improvement by enabling an organization to focus on operational areas requiring improvement, commensurate with the need. Metrics should be used to determine the degree to which a process is under control or is improving. However, the number itself is not significant without the context of many factors, including (but not limited to) product complexity, performance history, trends, targets, and action levels. Therefore, metrics need to be interpreted by assessing the root causes and determining the resulting risk of each failure point, which can then inform decisions and trigger action that lead to preventing recurrence. The point is not to make the metric number itself look good, but to analyze the data behind the metric to understand if an issue exists, to determine the underlying causes contributing to the signal, and to identify how to control the contributing variables.

ALIGN WITHIN CURRENT USE OF METRICS: Many organizations may already use the metrics identified through the FDA/Xavier Metrics initiative described herein, or may even use more complex analyses. The intent of the metrics provided herein is not to reduce the rigor and breadth of an organization's internal approach, but to propose a manageable system of metrics across the Total Product Lifecycle, as well as best practices for metric implementation.

APPLY INTERNALLY: These metrics have been developed in order to provide an assessment of the risk to product quality within an organization, not to provide organization to organization comparisons out of context. In order to provide focus to an inspection and demonstrate the maturity of its quality processes and systems, an organization may discuss with regulators the actions it voluntarily took in response to metric signals.

EVALUATE RISK: Metric analyses should address risk associated with the signal, either quantitatively or through qualitative commentary to explain the context and associated risk. Utilizing a risk analysis allows for identification of the impact of failures, prioritization of actions, and coordination of resources throughout the organization to ensure areas of high risk receive appropriate attention. For example, a Right-First-Time (RFT) result of 90% would enable a firm to focus attention on the 10% failure and analyze the data based on risk. *(NOTE: risk profiles vary from organization to organization, and then over time within an organization. Each organization must determine risk levels based on the severity of the impact, the probability of occurrence (including an historical understanding of the failure modes), and the dectability of*

the failure. These elements guide the development of mitigation strategies such that they are commensurate with the need.

EVALUATE WITH OTHER METRICS: The metrics identified by the FDA/Xavier Work Group are often analyzed along with other data to drive decisions. As a result, each of the metrics provided herein include examples of other data to assess in parallel or in aggregation.

CONSIDER THE IMPLICATIONS ACROSS THE TOTAL PRODUCT LIFECYCLE (TPLC): Ideally, metrics assist in the detection of underlying root causes that need to be addressed in order to prevent recurrence. The ultimate goal of a robust metrics program is continual improvement throughout the TPLC, such that the root cause of the failure is taken back to the earliest stages of development as possible in order to improve the outcome for any current and future product.

Pre-Production Metric and Best Practices for Implementation

Background for the Pre-Production Metric. Identification of a metric to assess the effectiveness of the research and development process proved to be challenging for the FDA/Xavier Work Group. Fortunately, the Work Group was divided into three separate teams such that one team was specifically charged with exploring this phase of the Total Product Lifecycle. Without this concerted focus, the overall team likely would have focused only on the production and post-production phases of the product lifecycle, since these areas are most commonly tracked. The Pre-Production team found that typical trial-and-error operations within the discovery phase of development made it challenging to track "failures" that reflect true failure. As a result, the team developed a filter such that the assessment of the rate of development failure would begin once development work is considered complete. By applying this filter, the team recognized that device transfer is the step that occurs after the development work is considered complete. Interestingly, however, it was found that most organizations do not track the rate of development failure that occurs during device transfer, and therefore, do not have the ability to improve the rigor of the development system.

Description of the Pre-Production Metric: If the research and development process is conducted with rigor, then the rate of change to the product and/or process during device transfer should be minimal. As a result, the pre-production metric is designed to track the number of changes that occur during the transfer stage that were triggered by product and/or process inadequacies. Assessing this metric enables organizations to track the frequency and volume of changes that could possibly have been avoided by a more robust research and development system. By tracking the metric, an organization has information that can inform the decisions of senior leaders related to potential improvements needed to the research and

development process. For example, upon review, it might be recognized that the rigor of the Voice of the Customer could be improved, or a more thorough evaluation of literature could be conducted, or even still, Human Factors studies could be improved. Additionally, tracking changes related to product and/or process inadequacies provides an indication of the overall time and cost of getting a product to a mature state in the market.

Goal of the Pre-Production Metric: to drive the Right-First-Time (RFT) mindset in the research and development phase such that post-design transfer changes due to inadequate product/process development are not needed.

The Pre-Production Metric:

$$\frac{\text{Total \# of changes (product \& process across projects)}}{\text{Total \# of projects}}$$

AND/OR

$$\text{Total \# of changes (product \& process for each project)}$$

Note: each organization must determine what constitutes a project

Definitions:

1. Changes. In order to compare the output of one research and development group to another such that successful practices can be more easily identified, harmonization of the change classification across the organization will be necessary. This harmonization creates the opportunity to aggregate data for higher level analyses. Organizations need to develop a mechanism to classify the significance of varying levels of changes based on risk. This will vary from organization to organization, and will change over time. Each organization must then identify which types of changes will be included in the metric, but triaging root causes to those pertaining to product and/or process inadequacies is what will enable the organization to identify opportunities for systemic improvements to the rigor of the research and development processes.

   The Work Group recommends separation of change into categories such that the failures can be triaged in a meaningful way. Examples of change category groupings include:

   - Product changes (including sub-assembly)

- Administrative
- Training
- Raw material components (Supplier)
- Process changes
- Planned or unplanned changes

Examples of category definitions include the following:

- Planned – Enhancements and New Features already considered during the design phase but not implemented prior to launch.
  - Examples: Realignment of Production Process Steps for efficiency (Lean Manufacturing), Adding Product Feature (Software features to product) if said feature does not change the product performance.
- Unplanned – Corrections implemented due to unanticipated VOC feedback, complaints, or any other unforeseen required changes.
  - Examples: Quality issue driven by Complaints, Returns, CAPA, Production yields, Nonconformance reports, etc.

2. Projects. The term "project" was used instead of "product" since a product can consist of many (sometimes thousands) of components. Therefore, "project" encompasses all aspects of the product being transferred from development to production.

Best Practices for Implementation of the Pre-Production Metric

1. Consistent application of the metric.

- Although not historically tracked by organizations, this metric is ideally calculated during the device transfer stage of the total product lifecycle. By demonstrating that device transfer occurs Right-First-Time, confidence in production performance is increased with a decrease in risk to product quality. The outcome of this metric establishes a baseline for performance during the production phase, and can be used to inform senior leaders of any systemic improvements needed in the development process.
- In order to establish the appropriate baseline for each product (which will vary from product to product), organizations should compare the results of this metric against the risk profile for each product.
- It is recommended that data be collected for Planned and Unplanned changes while evaluating the reason for each type of change. It is important to not thwart improvement in an attempt to just improve the metric result. It is equally important to not hide unplanned changes in the planned category, or to tag unplanned changes as innocuous improvements. After triage, only those changes that are truly related to

inadequacies associated with the product and/or process development should be included in this metric.

- In determining target values, acceptable ranges and action limits (threshold), each organization must consider the complexity of product, impact to patient population, risk of failure, business impact, etc. It is important to recognize that some products naturally have a higher rate of change than others (e.g. complex vs. simple products).

2. Time period and Frequency of data collection.
    a. During transfer, data should be collected for each transfer trial for each product to demonstrate the rate of success.
    b. During production, data should be assessed for trends at a frequency that statistically makes sense for the product in question, and includes information regarding historical performance.
    c. It is recommended that data be collected for a minimum of 12 months after release on the market and/or until steady state is reached. By trending this data periodically (e.g. quarterly), organizations can track how long it takes to reach product maturity. Importantly, successful practices can be gleaned from those products reaching a mature state sooner than others.
    d. Steady state could be governed by risk assessments and post production performance metrics (e.g. assessment of design FMEAs, or rate of change related to product/process inadequacies, or complaint rate remains consistent for consecutive quarters, etc.).
    e. In order to track how long it takes for a product to reach a mature state, or steady-state, it is recommended that tracking the rate of changes related to product and/or process inadequacies continues in production and after the product is released on the market. Importantly, the outcome of this metric is to be fed back into the pre-production phase such that improvements can impact future product development.
3. Usefulness of the metric.
    - Tracking the number of changes (related to product and/or processes inadequacies) per project over time can be an indication of the effectiveness of the design controls process.
    - Outputs from key processes can be incorporated as feedback loops that are inputs to the change management process and new product development process. (E.g. CAPA Governance, Internal Audit Process, Risk Management, Design Review, etc.).
    - Inputs to the change management process (CAPA, Complaints, Risk Management, etc.) should be trended to identify an appropriate target value for this metric. Recognition of

the interconnectivity of these metrics is critical for a robust quality management process.

- The data can be used by organizations to focus decisions and actions on areas that need attention.
- It is anticipated that the overall metric result should go down from product generation to generation since improvement opportunities should be better anticipated when developing a new generation of an existing or older product.
- The proposed metric can be useful to measure relative performance between internal business groups.
- Other metrics that could be used in conjunction with this one: Near Miss, Right-First-Time, Production Yields, Non-Conformance rates, Scrap Trends, Process and Design Validation Results, Design Review, Post Market Complaint incidence per million (CIPM).

Production Metric and Best Practices for Implementation

Background of the Production Metric. The FDA/Xavier Work Group was divided such that a team was dedicated to the exploration of which metric could be tracked during production that would drive continual improvement and could inform decisions and trigger actions such that the Right-First-Time mentality could be shifted as close to the initial days of development as possible. Many commonly tracked metrics were assessed, as well as innovative ways to look at the production process. In the end, the team decided to go forward with a commonly tracked metric, but to specify how best to leverage the output.

Description of the Production Metric: The Production metric calculation that was chosen is the Right-First-Time in production metric that many organizations already track. However, the team recognized that by triaging the root causes such as to isolate those related to product and process inadequacies, an organization can continue assessing the effectiveness of the development process. By tracking this metric during production in this way, senior leaders are armed with more information related to the robustness of the development process, and can more adequately identify opportunities for systemic improvement. Traditionally, the production Right-First-Time metric is tracked to assess the efficiency of the production operation. Rarely is this information fed back into the research and development system in an effort to decrease failure risk for future products. Senior leaders can better allocate resources across the enterprise (i.e., human, capital, etc.) to be focused on areas that will result in true improvement, and therefore, a reduction in risk to product quality.

Goals of this Metric:

- To gather nonconformance information during production operations related to inadequate product/process development.
- To inform senior leaders of the nonconformance rate in production related to inadequate development, such as to drive improvement in the overall research and development process for future products.
- To drive a Right-First-Time mindset throughout the organization such that failures occurring during production due to inadequate product/process are reduced or eliminated.

The Production Metric:

$$\frac{\text{\# of units mfg. Right-First-Time within or across lots}}{\text{\# of units started}}$$

Definitions:

1. Right-First-Time (RFT). A production run is considered RFT when the unit has been manufactured without the occurrence of a defect/non-conformance.

2. Start. The definition of "Start" is to be determined by each organization based on an assessment of the timing of the critical steps for the finished good. Each organization is best suited for determining the most meaningful "start" timing for its own product, which could be associated with the finished product and/or a number of production sub-processes.

3. Unit. For the purpose of this metric, the term unit is meant to be either a finished good, in-process material, sub-component, or other. "Units" can be interchanged with "lots" if more appropriate. It may be useful to consider the finished good in its entirety (including raw materials, sub-components and finished goods), or treating each sub-component and finished good lot separately.

In order to calculate the rate of Right-First-Time, the number of non-conforming units is subtracted from the number of units started to determine the number of Right-First-Time units in the numerator, which is then divided by the number of units started.

Example: Over the course of one (1) month, Organization A initiated production of 4,000 lots. Of the 4,000 lots initiated, 3,500 were produced Right-First-Time (RFT). Therefore, the Right-First-Time rate was calculated to be 87.5%.

(3,500 lots produced RFT / 4,000 lots initiated) * 100% = 87.5%

Best Practices for Implementation of the Production Metric

1. Consistent application of the metric.

   Each organization must consistently determine what is included in the tracking of non-conformances related to this metric. The Work Group recommends the following for consideration, but again, the key is to ensure consistent tracking of meaningful inputs:

   - Planned and unplanned rework. After triage, the organization can determine if the amount of planned rework is consistent with history and is acceptable. Any changes noted should be investigated to determine if there is an increase to patient safety risk.
   - Scrap and set-up scrap. Similar to planned rework, after triage, the amount of scrap and set-up scrap should be assessed for consistency or improvement compared to expectations and history. Any changes noted should be investigated to determine if there is an increase to patient safety risk.
   - All product-impacting non-conformances, which could include (but not limited to):
     - Sub-components / In-process material not meeting specification
     - Finished goods not meeting specification
     - Manufacturing in-process output not meeting acceptance criteria
     - Manufacturing process steps not following the procedure
     - Manufacturing process parameters not run within pre-defined acceptable ranges
   - It is felt that most non-product impacting non-conformances should be included until triage indicates appropriately that they truly are non-product impacting.

- It is not intended that this metric will include raw material/component failures. However, the Work Group does encourage organizations to have critical suppliers adopt this metric.

2. Time period and Frequency of data collection.

   - This metric is meant to track performance over time – monthly or as statistically appropriate, since trending this metric can illustrate shifts in production performance.
   - Applying pre-determined action limits, targets or control limits can be useful in identifying when action may be needed. It is expected that different thresholds exist across products, so each product should be assessed independently.

3. Usefulness of the Metric.

   This metric is a powerful tool for assessing production efficiencies, as well as comparing production facilities, lines and/or products during the production phase of the total product lifecycle (reminder to consider the data in context when making comparisons, especially considering product complexity, maturity, historical trends, etc.). However, as noted previously, if the root causes of failure are triaged consistently, then the power of the tool grows considerably. For example, all root causes related to product and/or process inadequacies can inform senior leaders such that actions can be triggered to initiate systemic changes to the development process. This metric can be a feeder into the CAPA system, triggering corrective actions to prevent recurrence, and therefore increase efficiency while decreasing risk to product quality.

   Additional insight can be gained by combining the Production Right-First-Time metric with post-production metrics (i.e. complaints, MDRs, recalls, etc.). For example, if the Production RFT rate is very high, yet many complaints are received for failures, then this indicates internal processes are not sufficient for detecting inadequacies or even failures prior to release. Also, if the Production RFT rate is low, then this could signal a higher risk to product quality if not all failures were caught and handled appropriately. A higher complaint rate, or risk of recall could be experienced.

   Importantly, in order to evaluate multiple metrics simultaneously, the time period must be consistent, and similar terms in the calculations must be defined consistently (or the inconsistencies must be understood).

Post-Production Metrics and Best Practices for Implementation

Background of the Post-Production Metric: The third team of the FDA/Xavier Work Group explored a vast array of post-production metrics to determine which metric could best assess risk to product quality once the product was on the market. The Work Group needed to keep in mind the vast array of product types manufactured and released by medical device companies across the industry, so a comprehensive list of post-production metrics needed to be considered.

Description of the Post-Production Metric: The Work Group found a number of post-production metrics to be important in painting a holistic picture of product performance on the market, so multiple indicators are included in the final metric that are commonly tracked by organizations: service records, installation failures, complaints, medical device records, recalls by number of units involved, and total number of recalls. The Work Group included calculations for each individual indicator that are commonly employed so as to ensure all organizations have a clear understanding of how to calculate the metrics. Additionally, however, the Work Group provided a more advanced calculation for each indicator that importantly incorporates the risk profile of the product. Too often, industry focuses resources on high volume issues alone, without having a way to quickly pinpoint high risk issues that may occur less frequently.

Goals of the Post-Production Metric:

- To analyze key post-market surveillance data to eliminate, reduce, and prevent future on-market failures. This assessment feeds into an overall Quality Management System (QMS) performance assessment and holistic QMS scorecard.
- To triage root causes of the on-market indicators that are related to inadequate product/process development.
- To drive a Right-First-Time mindset in product and process development such that failures occurring during post-production related to inadequate product/process development are reduced or eliminated for future products.

The Post-Production Metrics:

In the equations below, the selected period should be the same for the numerator and denominator (e.g., rolling 12-months for Complaints vs. rolling 12 months for products sold over the same time period).

- Service Records: Records per product / total # of units in service (for the period)

- Includes unplanned service and in-warranty service events. Does not include preventative maintenance.
- Note: Since some organizations track service records as complaints, they should be excluded from the service record count so as not to double count them. However, organizations are encouraged to track and trend the rate and causes of services records.

- Installation failures: # of installation failures / total # of installations (for the period)
  - Includes "out-of-box" failures.
  - Note: Since some organizations track installation failures as complaints, they should be excluded from the installation failure count so as not to double count them. However, organizations are encouraged to track and trend the rate and causes of installation failures.

- Complaints: Complaints for the product per period / units sold for the product per period
  - The final number may be adjusted if the initial triage indicates that the complaint is not a valid complaint.
  - Per period – need to define the period, such as rolling year or calendar year. (e.g., complaints received in 2015 vs. products sold for the calendar year, regardless of whether the complaints were received for products sold prior to 2015.
  - Number of units sold can be replaced by total number of products in use or by total uses in the period (e.g., capital equipment, such as instruments or defibrillators).

- MDRs: MDRs for the product / units sold (for the product)
  - Need to include MDRs, MDVs, and AERs.
  - Count only initial (and not follow-up) MDR submissions.

- Recalls (units): # of units involved in the recalls (for the period) – worldwide
  - Include the number of units involved in recalls, field corrective actions, and safety alerts (measure of absolute number).

- Recalls (total): # of recalls (for the period) – worldwide
  - Include the number of recalls, field corrective actions, and safety alerts (measure of absolute number).

Advanced Post-Production Metric Calculations.

An example calculation of how to move from a simple Complaint Rate to a Complaint Risk Score is provided in Appendix B. Predetermined risk ranges can be used to establish color coding of High, Medium and Low risk so as to provide a quick visual identification of critical areas needing attention and mitigation. This same process could be repeated for each indicator in the Post-Production metric where severity can be categorized as follows with appropriate/commensurate weighting factors:

- Complaints: Catastrophic, Critical, Marginal, Negligible
- Service Records: Catastrophic, Critical, Marginal, Negligible
- MDR's: Death, Serious Injury, Malfunction
- Recalls: Class 1, Class 2, Class 3

The Work Group intentionally did not define severity rankings for each category of each Post-Production indicator, since risk is highly dependent on product complexity, product performance trends, and company culture (not intended to be an all-inclusive list). Therefore, each organization must determine its own Risk Scores.

Best Practices for implementation of the Post-Production metric

The FDA/Xavier Work Group has developed a multi-step approach for options on how to implement use of the post-production metrics:

1. Each applicable post-production indicator should be calculated, tracked and trended as an important part of any post-market surveillance program. Calculations have been provided for each indicator.
2. Advanced Calculations are provided in Appendix B that incorporate product risk profile information into the calculation of each Post-Production indicator. Importantly, this enables organizations to quickly identify areas of highest risk across the product in question, as well as across all products.
3. Comparative analyses can be conducted through mechanisms such as dashboards, score cards or heat map tools. These analyses can give senior leaders an enterprise-wide view of risk to product quality, and is discussed in the "Enterprise-wide Continual Improvement" section of this document.

Usefulness of the metric.

It is believed by the FDA/Xavier Work Group that tracking and trending each applicable metric provided herein is an important aspect of any post-market surveillance program. However, organizations should consider incorporating a more advanced calculation based on product risk profiles in order to ensure resources are focused on areas of greatest risk to product quality, and therefore, patient safety.

Importantly, the Work Group believes that viewing the Post-Production indicators in aggregate could lead to the identification of risk that otherwise might not be recognized. For example, medium risk scores for complaints, or service records might not trigger action limits. However, an organization might respond differently if it was recognized that complaints, service records, MDRs and installation failures all had medium risk scores in a given time period. Therefore, an assessment of all post-market indicators in the calculation could provide a more holistic assessment of risk to product quality that could otherwise have been missed.

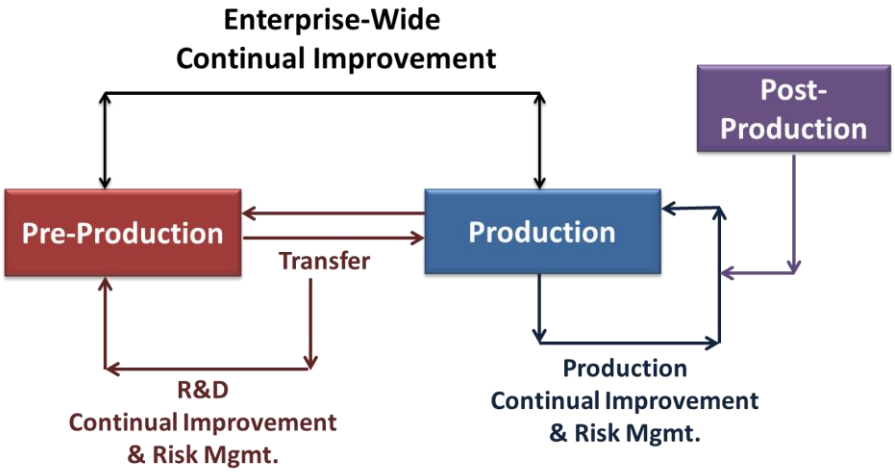
An assessment across products enables the organization to focus its resources on the products driving the largest number of complaints and MDRs. The metrics can be applied to a representative product from each product family to determine otherwise hidden trends and differences.

Importantly, this type of calculation should be conducted commensurate with the need, and can be useful when there is a large diversity in product profiles. The output of the advanced calculations in aggregation could be used by senior leaders to inform decisions and trigger action, and is therefore covered in more depth in the “Enterprise-wide Continual Improvement” section of this document. This metric is intended to be monitored over time, such that the improvements and regressions can be assessed for learnings that can improve systemic practices.

Enterprise-wide Continual Improvement

Background for Enterprise-wide Continual Improvement. The FDA/Xavier Work Group quickly recognized that the value of any metric is not tied to the number itself, but rather, to how organizations use the output from the metric to inform decisions and trigger action. Additionally, the Work Group recognized that the metrics defined within this document need to be assessed along with other metrics and sources of information to provide a more holistic view of the overall risk to product quality. As detailed herein, the Work Group assessed the Total Product Lifecycle in order to drive the Right-First-Time mentality as close to the initial days of development as possible. If this mindset is achieved, then risk to product quality and therefore patient safety can be realized, in addition to increased production efficiencies, cost reduction, and preservation of Brand equity.

Description of Enterprise-Wide Continual Improvement Process. The enterprise-wide continual improvement process enables a comparative analysis of products through the use of tools such as heat maps, dashboards, and/or score cards. The process represents the highest level of analysis to allow senior leaders to keep their fingers on the pulse of the performance of their total product portfolio, and is therefore referred to as the Enterprise-wide Continual Improvement process. The diagram below portrays the critical feedback loop mechanisms, and importantly how the enterprise-wide continual improvement loop links the production and post-production phases of the product lifecycle back to the development phase. This feedback loop enables systemic improvements to be made to the rigor of product development.



Goal of the Enterprise-Wide Continual Improvement Process: To inform Enterprise-wide Continual Improvement across the Total Product Lifecycle by using the Pre-Production, Production and Post-Production metrics to enrich knowledge management across the entire enterprise. Additionally, it is important to drive the Enterprise-Wide Review to the lowest points in the organizational structure as possible such as to operate with an understanding of "why" at all levels. The assessment of data must be performed consistently across products, and the data must consistently be used to make decisions. In the case of Senior Management review, employees should assess and evaluate the data prior to providing it to Senior Management in order to be able to explain the data context and propose an appropriate course of action for Senior Management approval. These recommendations would be further accompanied by an understanding of risk impact to product quality, patient safety, organizational finances, and brand equity. The data should be provided to Senior Management with an analysis of the underlying root causes and any contributing causes; a description of what action has already been taken at various stages throughout the TPLC; and a proposal of what other action could be taken to mitigate any additional cost of poor quality.

The following best practices provide examples of how to implement these kinds of analyses, and should be implemented commensurate with the need.

Examples of How to Implement. Some organizations may use more sophisticated heat maps; others may use dashboards or scorecards that are suitable for the complexity of their operations and the maturity of their organization. Provided in this section are examples of how data can be aggregated to provide a Big Picture view of the risk to product quality, such that resources are focused on the right area.

1. Aggregation of data across the Total Product Lifecycle.

The example provided in Appendix B was described in the Post-Production metric section of this document, and demonstrates how to establish a risk score for complaints over time. Appendix C takes that example further and demonstrates how a risk severity categorization of each of Post-Production metric can be established along with ranges of acceptability. These ranges can be color coded to provide an immediate visual queue as to areas of high risk. The example provided in Appendix C demonstrates how to incorporate the Pre-Production metric and Production metric into an overall quantification of risk to product quality.

Assessment of the table in Appendix C demonstrates how reviewing metrics in isolation can lead to an incomplete understanding of risk to product quality. The following are examples of this point:

- There were no recalls in Q3, so the recall risk profile is green. However, taken into context with the other post-production metrics yields a risk profile of orange, which is the second highest risk categorization.
- A Q2 pre-market risk profile of green is rolled into an overall risk profile of red across the Total Product Lifecycle.
- The Q1 red risk profiles of Pre-Production and Production outweigh the green risk profile of Post-Production to yield an overall risk of orange.

This type of analysis demonstrates the power of reviewing metrics in aggregation versus isolation.

2. Dashboard for new product launches.

A dashboard can be a simple graphic that can be utilized to evaluate metric performance, as demonstrated below.

	Pre-Production	Production	Post Production
Product A	Value 1 ⟷	Value 4 ↓	Value 7 ↓
Product B	Value 2 ⟷	Value 5 ↓	Value 8 ↑
Product C	Value 3 ↑	Value 6 ⟷	Value 9 ⟷

Dashboard Creation:

- Generate a standard table containing (X) number of rows corresponding to each product being evaluated and (Y) number of columns corresponding to each metric being tracked
- Populate the table with the data for each metric for each product. In order to conduct a meaningful comparison across products, the metric terms in the calculations must be defined and applied consistently.
- The frequency for reporting the metric results must be established and again, harmonized across the metrics.

- Identify change over time utilizing a visual indicator such as color coded symbol. (e.g. no change- horizontal double arrow, increase - arrow up, decrease – arrow down)

Interpreting the data within a product and driving action:

- Individual metric data can be evaluated against the previous time point to understand performance and identify changes. Shifts up or down need to be further examined to understand the cause and identify any required actions to improve or sustain performance.

- Individual metric data can be evaluated against an established goal to understand performance. Goals can be established based on anticipated design or process improvements for new product. Ex: Product B was designed with specific product enhancements in comparison to the predicate product. A goal for RTF and post market may be established in advance to determine if the changes are effective.

- Data can be compared across the metrics within a product to identify interdependencies/correlations. This evaluation may provide predictive analysis for future performance or identify areas of improvement. Ex: Improvements in the production Right-First-Time metric may correlate to an improvement in post market performance. This correlation may be utilized to establish a future performance expectation for post market based on continuous improvements within RFT.

Interpreting data across products:

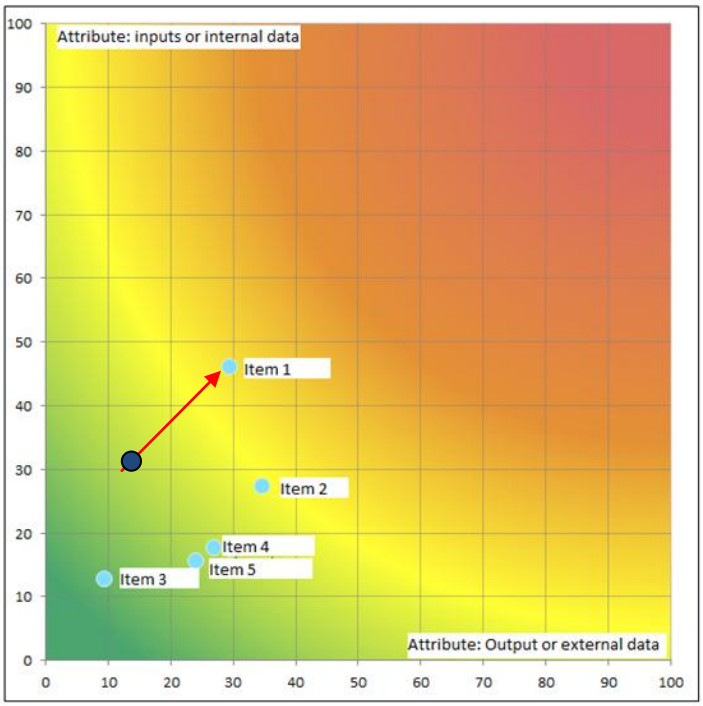
- Individual metric data can be compared across products to understand differences in performance. Differences between products can be further evaluated to understand the cause and identify areas of alignment or best practices to ensure optimal product performance.

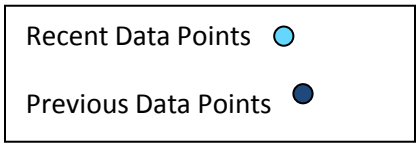
- Individual metric data can be evaluated against a predicate/similar product at the same time point. Ex: Product A had X performance 6 month after launch. New Product B at 6 months post launch can be compared to the baseline of Product A. Product B has a reduction in RFT, which is therefore shown in red since that is not desired. Although there was a decrease in the number of post-production issues for Product B, the risk of was higher, therefore resulting in a red arrow from the increased risk. Product C has an increase in the pre-production measure, but the effects have not been noticed in production or post-production measures. There may be an opportunity in product C to reduce or improve the occurrences of issues in production and post-production by

understanding what is occurring with Product C in regards to the pre-production rate of product change metric.

3. Heat Map.

A heat map is an effective tool that can be utilized to evaluate individual product performance against an established scale and provide direct comparisons across products.





Heat Map Creation:

- Generate standard xy graph by establishing the data for each axis. This can be an individual metric or compilation of data (typically a compilation of data)
- Establish the relative performance/gradient across the graph (Ex: create a color coded green, yellow, and red gradient based on risk)
- Establish the frequency for reporting the results, and plot the data for each product

  An example: assess external signals for a product versus internal signals in order to establish the overall risk to product quality.

- External signals would include the components in the Post-Production metric. Internal signals would include Design Changes and Right-First-Time in production.

- To combine data together and include all products on the same graph, each metric can be turned into an index. For each metric, the organization determines a range from 1-100: 1 being the "best" and 100 being the "worst" in terms of metric performance. Based on actual product performance a score from 1-100 is assigned. Ex: product with RTF of 10% is considered worst (score of 100) and 90% is considered best (score of 1). Any performance between those values is assigned a linear score between 1-100.

- The 1-100 scores for each metric are added together for a total score for that axis (External Signals or Internal Signals), which is then plotted on the graph.

- The graph is segmented by a risk gradient or grid that is predetermined based on severity or the significance of the 1-100 score of each metric.

Interpreting data and driving action:

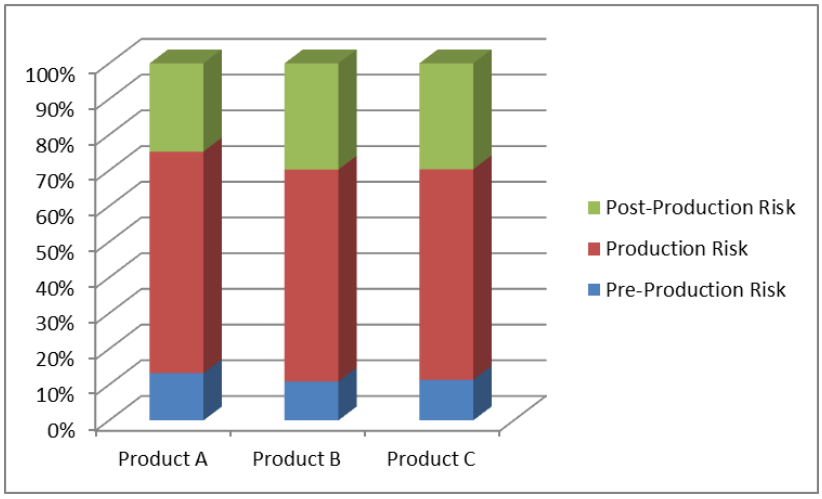
- By plotting product performance on a heat map, organizations are able to recognize potential risks associated with an individual product performance and differences across product. The purpose of this is to foster a deeper dive on why the difference exists. In the heat map above, item 1 (product 1) has shifted in performance from a more acceptable risk to an area of potential concern. This is depicted by the shift in the graph from the previous data comparison to the recent data comparison. The other products in the heat map have remained stable in performance compared to the previous assessment. Questions to consider:
  - What are the key drivers both positive and negative for an individual product's placement on the map?
  - What are the key differences between the best and worst performing products?
  - What actions can be taken to move the performance in a positive direction?

- At a defined interval (recommended six months or as statistically appropriate for the data being analyzed), reevaluate and identify changes in performance. Questions to consider:
  - What are the key drivers behind shifts in the data – within product time points and between products?

- Has the product moved into a lower "risk" range and is it sustainable?
- Are there any shifts impacting multiple products and is that indicative of a systemic change?
- Were planned changes effective at improving product risk scores?
- Why is Item 1 (product 1) behaving differently and shifted in performance from the previous assessment period?

4. Stacked Bar Charts.

Stacked bar charts with all three metrics (Pre-Production, Production and an aggregation of Post-Production) can demonstrate to what degree each metric is contributing to the whole. This can be used to assess the proportions across product families, or trend changes over time for a single product.



Without much effort, the stacked bar chart shown above would lead an organization to recognize that all three products experience significantly more risk in production than the other two areas.

Drawing Conclusions

Organizations need to be careful to not draw unsubstantiated conclusions from the data. This section provides examples of factors that can influence the metric outcomes that are often not considered, realized or recognized. These factors include: supplier performance, validation/verification data, resourcing (staff turnover), training, preventative maintenance, and cultural region, just to name a few.

The table below provides scenarios for changes over time across the metrics described within this document. The scenarios are followed by examples of misinterpretation, and a shift in thinking on what could be influencing the metric result.

1	**Design Changes up, RFT down, Post market down** Scenario: Post-launch design changes being driven by production issues, where changes do not affect product form, fit or function. First assumption might be: inadequate development of production processes during design control stage At second look, it might also represent: (1) changes that are also product impacting, but where the effects appear later in the product lifecycle or at a time point beyond that of the assessment period, or (2) cause of reduced RFT units is unrelated post-launch design changes
2	**Design Changes up, RFT down, Post market up** Scenario: Some post-launch design changes being driven both by production issues and customer complaints First assumption might be: inadequate design of product and of production processes during design control stage At second look it might also represent: (1) inadequate product design (as reflected by increased post market issues, but where the increase in RFT might be due to (a) aging equipment (right equipment but beyond usable life) or (b) turnover in production personnel, or (c) unplanned environmental issues (such as loss of power during production) (2) inadequate production process design unrelated to complaints; where complaints might represent change in user's expectations for the product (product does not meet needs of customer)
3	**Design Changes up, RFT up, Post market down** Scenario: Post-launch design changes being implemented as preventative measures against future RFT or complaint issues First assumption might be: incremental changes being implemented as preventative measures

	At second look it might also represent: (1) changes that are also product impacting, but where the effects appear later in the product lifecycle or at a time point beyond that of the assessment period
4	**Design Changes up, RFT up, Post market up** Scenario: Post-launch design changes being driven by customer complaints (i.e., wrong design, right production process). First assumption might be: inadequate development of product during the design control stage At second look it might also represent: (1) changes in source of critical materials that change the cosmetic appearance or performance of the device; (2) in the case of instrumentation changes (i.e., changes in parts specifications) in response to inadequate preventive maintenance or servicing programs
5	**Design Changes down, RFT down, Post market down** Scenario: Few or no post-launch design changes, few complaints but RFT rate is lower than targeted First assumption might be: Right product design; potentially inadequate process design (i.e., RFT rates suggest additional design changes may be necessary to fix production processes) At second look it might also represent: (1) product is in its earliest phase of post-launch life cycle (i.e., no problems seen yet), or (2) RFT rate is being affected by change in quality of a raw material or component (i.e., correct design and specifications) and where defective units can be identified and removed during production, (3) aging equipment affecting RFT rate, (4) turnover in production personnel effecting RFT rate
6	**Design Changes down, RFT up, Post market down** Scenario: Few or no post-launch design changes, most production units are RFT, few complaints First assumption might be: Right product design and right production process design is driving low complaint rate (i.e., state of control) At second look it might also represent: (1) product and the amount of data from metrics

	are in their earliest phase of post-launch lifecycle and not yet capable of detecting problems or trends
7	**Design Changes down, RFT down, Post market up** Scenario: Few or no post-launch design changes, many complaints, RFT rate is lower than targeted First assumption might be: Potentially inadequate product and process design and where future changes may be necessary At second look it might also represent: (1) product is in its earliest phase of post-launch life cycle where RFT and complaint rates reflect production personnel and customer burn-in periods, or (2) RFT rate being affected by change in quality of a raw material or component (i.e., correct design and specifications) and where effects of the changes were not detected by in-process or final testing, but were detected by the customer, (3) aging equipment affecting RFT rate that could be related or unrelated to complaints, (4) turnover in production personnel effecting RFT rate and quality of product released to customers
8	**Design Changes down, RFT up, Post market up** Scenario: Few or no post-launch design changes, RFT rate is positive, but complaints indicate a problem First assumption might be: Right process design but potentially wrong product design At second look it might also represent: (1) product is in its earliest phase of post-launch lifecycle and where complaint rate indicates customer burn-in period as might be expected with the introduction of new or first-of-kind technology, or (2) complaints reflect customer's changing demands/expectations for product performance

Note that the above scenarios require the organization to determine true root cause in order to prevent recurrence. For example:

- Complaints reflecting customer's changing demands: the rigor of and process for Voice of the Customer, or Human Factors trials or clinical trials needs to be assessed to ensure proper inputs are considered in the device design.

- Turn-over rate in production: this is not an acceptable reason for failure. The training program needs to be capable of properly preparing employees for the work they are to perform without error.
- Aging equipment affecting RFT: the preventative maintenance program needs to be reevaluated to determine the true acceptable life of the equipment to ensure the equipment is capable of consistently manufacturing the device.

Appendix A: FDA/Xavier Work Group Members

The following is a listing of the industry professionals and FDA officials who participated during all or part of the September 2014 – August 2016 work.

First Name	Last Name	Title	Company Name
Paul	Andreassi	Interim Vice President	CR Bard
Karen	Archdeacon	Compliance Officer - New England District	FDA
Pat	Baird	Director, Engineering	Baxter Healthcare
Kathleen	Bardwell	Sr. VP and Chief Compliance Office	STERIS Corporation
Anupam	Bedi	Director of Quality	AtriCure
Pankit	Bhalodia	Director	PwC
KB	Bheda	Manager	PwC
Steve	Binion	Director Regulatory Affairs/Corporate Clinical Development	BD
Robin	Blankenbaker	Divisional Quality Operations Leader	W.L. Gore & Associates
Rafael	Bonilla	Quality Engineer	ScottCare Corp.
Gina	Brackett	Compliance Officer	FDA
Kate	Cadorette	Manager, Internal Audit	STERIS Corporation
Patrick	Caines	Dir, Quality & Global post market surveillance	Baxter Healthcare
Tony	Carr	Vice President, Global Quality Systems	Boston Scientific
Ross	Carter	Experienced Associate	PwC
Kara	Carter	Senior Director, QA Operations	Abbott Vascular Division
Vizma	Carver	Founder and CEO	Carver Global Health Group
Aaron	Dunbar	Quality Systems Manager	Boston Scientific
Ryan	Eavey	Senior Manager, Quality Systmes	Stryker
Joanna	Engelke	Senior Vice President Global Quality	Boston Scientific
Tom	Haueter	Director, Quality and Regulatory Affairs	Clinical Innovations
Chris	Hoag	Director of Global CAPA and Quality eSystems	Stryker
Jeff	Ireland	VP Core Quality Services, Cardiac & Vascular Group	Medtronic
Greg	Jones	AVP Healthcare	BSI
Bryan	Knecht	Quality Systems Manager	AtriCure
Jonathan	Lee	Senior Associate	PwC
John	Lewis	Director, Quality Operations	Meridian Bioscience, Inc.
Bill	MacFarland	Director, Division of Manufacturing Quality, CDRH	FDA

First Name	Last Name	Title	Company Name
Kristin	McNamara	Senior Advisor	FDA
Rhonda	Mecl	Supervisory CSO	FDA
Brian	Motter	VP Quality and Compliance, Diabetes	J&J MD&D
Ravi	Nabar	Sr. Director Supplier Quality Management	Philips
Scott	Nichols	Director	Abbott
Steven	Niedelman	Lead Quality Systems and Compliance Consultant	King & Spalding LLP
Pete	Palermo	VP Quality Assurance	CR Bard
Luann	Pendy	Vice President Global Quality	Medtronic
Marla	Phillips	Director	Xavier University
Greg	Pierce	President and Founder	Engisystems
Susan	Rolih	Executive Vice President, Regulatory and Quality Systems	Meridian Bioscience, Inc.
Barbara	Ruf	Director, Corporate Quality, Management Controls	Zimmer Biomet
Joe	Sapiente	VP Global Quality Operations	Covidien/Medtronic
Brian	Schultz	Vice President, Quality and Regulatory Affairs	Fisher & Paykel Healthcare Limited
Gin	Schulz	VP Quality Assurance	CR Bard
Benjamin	Smith	Vice President, Global Quality System & Compliance	Biomerieux
Kristy	Spoon	Quality Systems Manager	Cook
Katie	Sullivan	Manager	Meridian Bioscience, Inc.
Isabel	Tejero	Lead Consumer Safety Officer	FDA
Shelley	Turcotte	WW Director Quality Information Systems	DePuy Synthes
Sam	Venugopal	Partner	PwC
Marta	Villarraga	Principal Biomedical Engineering	Exponent
Monica	Wilkins	Divisional Vice President of Quality and Business Support	Abbott

Co-Leaders of the Initiative:

- Kristin McNamara FDA
- Marla Phillips Xavier University

The Work Group would like to express gratitude to PricewaterhouseCoopers (PwC) for leading the pilot study, analysis and summary.

Appendix B: Example – Complaint Metric Analysis with Risk Profile

The following is an example of how to generate a Complaint Risk Profile Score.

Step 1.

Prior to receiving complaints, severity classifications of complaints need to be determined, along with a weighted value for each severity level.

Severity Classification	Severity Weighted Value	Severity Definition
Catastrophic	50	Potential for Death
Critical	30	Potential for Serious Injury
Marginal	19	Potential for Non-Serious Injury
Negligible	1	Minor Customer Annoyance, Cosmetic Issue, No injury to patient

Step 2.

Once the severity classifications and weighted severity values have been determined, each complaint is assessed for severity. For example, 1 critical complaint received out of 1,000 complaints would be multiplied by a severity weighted value of 30, then is multiplied by 100:

1 critical complaint/1,000 complaints received x severity weighted value of 30 x 100 = 3

This calculation is repeated for each complaint in each period, where the total is the sum of severity from each classification of complaints:

1st Quarter Product a Risk Profile Score:
1000 Complaints for period
1 Critical Complaint/1000 Complaints X 30 X 100 = 3
3 Marginal Complaints/1000 Complaints X 19 X 100 = 5.7
996 Minor Complaints/1000 Complaints X 1 X 100 = 99.6
Total Risk Profile Score = 108.3

2nd Quarter: Product A Risk Profile Score:
1000 Complaints for period
2 Critical Complaints/1000 Complaints X 30 X 100= 6
4 Marginal Complaints/1000 Complaints X 19 X 100 = 7.6
995 Minor Complaints/1000 Complaints X 1 X 100 = 99.5
Total Risk Profile Score = 113.1

3rd Quarter: Product A Risk Profile Score:
500 Complaints for period
2 Catastrophic Complaints/500 Complaints X 50 X 100 = 20
4 Marginal Complaints / 500 Complaints X 19 X 100 = 15.2
494 Minor Complaints/ 500 Complaints X 1 X 100 = 98.8
Total Risk Profile Score = 134

4th Quarter: Product A Risk Profile Score:
2500 Complaints for period
50 Marginal Complaints / 2500 Complaints X 19 X 100 = 38
2490 Minor Complaints/ 2500 Complaints X 1 X 100 = 98
Total Risk Profile Score = 136

Step 3.

The rate of each complaint is an important factor in the overall risk determination, since higher volumes of lower level risk can be categorized as a higher risk based on increased opportunities for greater risk. The complaint rate should be calculated in a way that allows for quick detection of changes when comparing one period to another. As an example, the number of complaints per number of units released can be multiplied by any factor (consistently for all) such that the reported rate is a whole number. In the table below, the complaint rates were multiplied by 1,000.

Quarter	Number of Complaints	Number of units released	Complaint Rate Total	Multiplier	Reported rate
1st	1,000	10,000	0.1	1,000	100
2nd	1,000	15,000	0.067	1,000	67
3rd	500	18,000	0.028	1,000	28
4th	2,500	10,000	0.25	1,000	250

Step 4.

The Complaint Rate from Step 3, and the Overall Risk Profile Score from Step 2 are tabulated in a way that allows for a quick assessment of complaint risk over time.

	1st Quarter Product A	2nd Quarter Product A	3rd Quarter Product A	4th Quarter Product A
Complaint Rate	100	67	28	250
Complaint Risk Profile Score	108.3	113.1	134	136

Interpreting the Data.

Upon review of the data above, the overall complaints per units sold decreases from 1st quarter to 2nd quarter. The risk profile of the complaints is increasing in the second quarter, which demonstrates an increase in the number of higher severity complaints within the population of the complaints reported during the period. This is an indicator to further investigate why the risk profile is increasing within the complaints.

The complaints per units sold continues to decrease into the third quarter, which is significantly lower than the 1st quarter. However, this is the second increase in the number of higher severity complaints. Overall the first three quarters demonstrate a decreasing trend in the complaints; however, we are seeing an increase in the severity of the complaints. This

increasing trend in severity provides an indicator to further investigate the risk increase to determine root cause.

In the 4th quarter there is an increase in the complaints per units sold, as well as an increase in the complaint risk profile. This increase in both complaint numbers and risk profile would require further investigation of the details of the complaints to determine root cause. In this scenario, there is a significant number of complaints rated as marginal in the risk profile, and a higher number of complaints overall.

Note: Overall an organization would need to define the detailed analysis and verify weighted values and multipliers to ensure the data is providing appropriate signals to determine when further investigation is required.

Appendix C: Example – Total Product Risk Profile

Production Phase	PRODUCT A	2016 Q1	2016 Q2	2016 Q3	2016 Q4	Scoring Criteria
Post-Production	Complaints	1000	1000	500	2500	Severity
	Catastrophic	0	0	2	0	50
	Critical	1	2	0	0	30
	Marginal	3	4	4	50	19
	Negligible	996	995	494	2490	1
	Complaint Risk Profile	108.3	113.1	134.0	137.6	Optimal ≤ 100, Acceptable 100-110, Moderate 111-135, Poor > 135
	Service Records	100	125	85	90	Severity
	Catastrophic	0	0	1	0	50
	Critical	2	6	5	9	30
	Marginal	12	15	18	20	19
	Negligible	86	79	76	71	1
	Service Record Risk Profile	37.4	54.4	123.6	28.8	Optimal ≤ 30, Acceptable 31-50, Moderate 51-100, Poor > 100
	MDR's	10	15	12	8	Severity
	Death	0	1	1	0	1000
	Serious Injury	1	0	1	3	500
	Malfunction	9	9	8	6	1
	MDR Risk Profile	50.9	100.9	150.8	150.9	Optimal ≤ 60, Acceptable 61-90, Moderate 91 - 115, Poor > 115
	Recalls	0	1	0	0	Severity
	Class 1	0	0	0	0	1000
	Class 2	0	1	0	0	500
	Class 3	0	0	0	0	250
	Recall Risk Profile	0	500	0	0	Acceptable = 0; Mitigation Level 1 = 1-250; Mitigation Level 2 = 251 - 500; Mitigation Level 3 = 501 - 1000
Total - Post Production	Post Production Total	196.6	768.4	408.4	317.0	Optimal 0 - 250; Acceptable = 251 - 400; Moderate = 401 - 700; Poor > 700
Pre-Market	change rate related to product/process inadequacies * 1000	400	5	67	150	Optimal ≤ 50, Acceptable 51 - 100, Moderate 101 - 200, Poor > 200
Total - Pre-Market	Pre Production Total	400	5	67	150	Optimal ≤ 50, Acceptable 51 - 100, Moderate 101 - 200, Poor > 200
Production	Right-First-Time rate	88.9	96.5	98.2	93.5	
	(only including product/process failures)					
Total - Production	Production Total	88.9	96.5	98.2	93.5	Optimal ≥ 98, Acceptable 96-97.9, Moderate 93-95.9, Poor <93
Overall Risk	Total (Pre + Prod + Post)	686	870	574	561	Red zone ≥ 750, Orange zone 570 - 749, Yellow zone 250 - 569, Green zone < 250

Sample calculations are shown on the following page.

Sample Calculations for how to determine the individual Risk Profiles for each metric category:

1. **Complaints Risk Profile** for Q1 = Sum of (total catastrophic complaints/1000 * severity multiplier * 100) + (total critical complaints/1000 * severity multiplier * 100) + (total marginal complaints/1000 * severity multiplier * 100) + (total negligible complaints/1000 * severity multiplier * 100)

   Complaints Risk Profile for Q1 = 0 + 3 + 5.7 + 99.6 = 108.3

2. **Service Records Risk Profile** for Q1 = (total catastrophic service records/1000 * severity multiplier * 100) + (total critical service records/1000 * severity multiplier * 100) + (total marginal service records/1000 * severity multiplier * 100) + (total negligible service records/1000 * severity multiplier * 100)

   Service Records Risk Profile for Q1 = 0 + 6 + 22.8 + 8.6 = 37.4

3. **MDR Risk Profile** for Q1 = MDR Risk Profile for Q1 = (total deaths/1000 * severity multiplier * 100) + (total serious injuries/1000 * severity multiplier * 100) + (total malfunctions/1000 * severity multiplier * 100)

   MDR Risk Profile for Q1 = 0 + 50 + 0.9 = 50.9

4. **Recalls Risk Profile** for Q1 = (total Class 1 Recalls/1000 * severity multiplier * 100) + (total Class 2 Recalls/1000 * severity multiplier * 100) + (total Class 3 Recalls/1000 * severity multiplier * 100)

   Recalls Risk Profile for Q1 = 0 + 0 + 0 = 0