The Software Precertification (Pre-Cert) Pilot Program: Tailored Total Product Lifecycle Approaches and Key Findings

September 2022

Table of Contents

Executive Summary

Software as a Medical Device (SaMD) is increasingly being adopted throughout the healthcare sector. These devices are developed and validated differently than traditional hardware-based medical devices in that they are developed and designed iteratively and can be designed to be updated after deployment to quickly make enhancements and efficiently address issues, including malfunctions and adverse events. In 2017, FDA recognized that the current device regulatory framework, enacted by Congress more than 40 years prior and incrementally updated since then, had not been optimized for regulating these devices.[1] The pilot explored innovative approaches to regulatory oversight of SaMD developed by organizations that have demonstrated a robust culture of quality and organizational excellence and that are committed to monitoring real-world performance of their products once they reach the U.S. market. While the pilot was focused on SaMD, what we learned is relevant to medical device software[2] in general and the latter term is used throughout the report. With this report, FDA is concluding this important first step, marking the completion of the pilot (see Figure 1).

To inform the pilot and its stakeholders, FDA issued a number of documents to describe its vision and approach, the plan for exploring whether the approach could provide reasonable assurance of safety and effectiveness when compared with the traditional regulatory paradigm, and the plan for implementing the pilot and the eventual Pre-Cert Program using current regulatory authorities: the Working Model, Test Plan, and Regulatory Framework, respectively.[3]

---

[1] Shuren, J., Patel, B., & Gottlieb, S. (2018). FDA Regulation of Mobile Medical Apps. JAMA, 320(4), 337-338. https://doi.org/10.1001/jama.2018.8832.

[2] FDA's regulatory oversight of medical device software applies to software that meets the definition of "device" in section 201(h)(1) of the Federal Food, Drug, and Cosmetic Act (FD&C Act) to include "an instrument, apparatus, implement, machine, contrivance, implant, in vitro reagent, or other similar or related article, including any component, part, or accessory, which is – (A) recognized in the official National Formulary, or the United States Pharmacopoeia, or any supplement to them, (B) intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals, or (C) intended to affect the structure or any function of the body of man or other animals, and which does not achieve its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of its primary intended purposes. The term 'device' does not include software functions excluded pursuant to section 520(o)" of the FD&C Act.

[3] Developing a Software Precertification Program: *A Working Model v1.0,* available at https://www.fda.gov/media/119722/download; Software Precertification Program: 2019 Test Plan, available at https://www.fda.gov/media/119723/download; and Software Precertification Program: Regulatory Framework for Conducting the Pilot Program within Current Authorities, available at https://www.fda.gov/media/119724/download. See also Digital Health Software Precertification (Pre-Cert) Program, available at https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-software-precertification-pre-cert-program.

The pilot focused on exploring the viability of this approach, as laid out in the Working Model, under FDA's current authorities. More specifically, the pilot explored whether certain methods for evaluating safety and effectiveness that leveraged information at multiple, opportune points throughout the total product lifecycle (TPLC) could be used to efficiently and successfully assess medical device software safety and effectiveness. The pilot also explored the value of potential tools that could enhance TPLC review processes while fostering efficient, consistent, and transparent regulatory decisions.

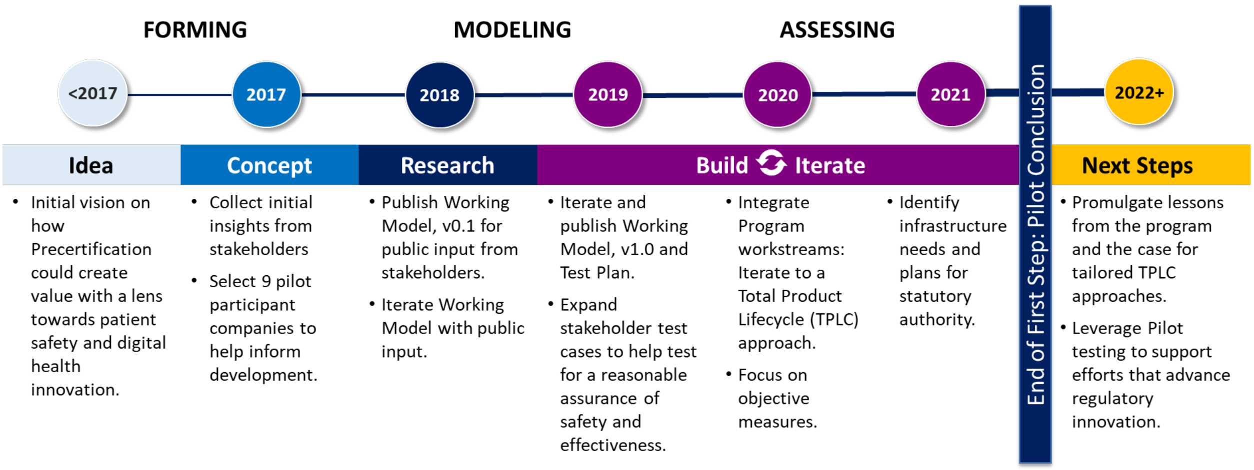

Figure 1. Overview of Pre-Cert Pilot Program Progression

While the pilot enabled FDA to explore innovative techniques and approaches to regulatory oversight with stakeholders, FDA encountered challenges with implementing the proposed approach under our current statutory authorities. Additionally, FDA found that limiting participation to 9 pilot participants[4], combined with the need to limit formal implementation of any approaches to De Novo classification[5], led to few devices being available for consideration under the pilot. In addition, for devices where De Novo classification was appropriate, the approach would have resulted in establishing pilot-specific special controls that would apply not only to the pilot participant's device but also to all devices where substantial equivalence could be demonstrated to the pilot participant's device, whether or not they were developed by pilot participants or other manufacturers. As a result, FDA was simultaneously unable to pilot the program approaches with a broad sample of devices while also being unable to limit the scope of any resulting device classifications. Further, FDA could not require pilot participants to provide information under the pilot that was not otherwise already required under existing statute. Pilot participants nonetheless engaged voluntarily and provided significant additional information to support the pilot, although it was challenging to

---

[4] The Paperwork Reduction Act (PRA) is a law governing how federal agencies collect information from the American public (see About the PRA at https://pra.digital.gov/about/). Under 5 CFR 1320.5, an agency shall not conduct or sponsor a collection of information unless, among other requirements, the Office of Management and Budget has approved the proposed collection of information. However, 5 CFR 1320.3(c) defines collection of information to include, among other things, identical reporting "imposed on **ten or more persons**" (emphasis added).

[5] For details, see the Regulatory Framework for Conducting the Pilot Program within Current Authorities, available at https://www.fda.gov/media/119724/download.

harmonize this information across pilot participants to develop consistent, repeatable methodologies.

Despite these challenges, the pilot provided key insights and furthered FDA's understanding of the concepts proposed in the Working Model. In particular, the pilot excellence appraisals enabled FDA to better understand the practices that pilot participants and others use in designing, developing, and managing digital health products.

We are not fully capitalizing on these capabilities and approaches for software in the current statutory and regulatory framework for medical devices. Based on these observations from the pilot, FDA has found that rapidly evolving technologies in the modern medical device landscape could benefit from a new regulatory paradigm, which would require a legislative change.

Given the challenges faced during the pilot, FDA has determined that the approach described in the Working Model is not practical to implement under our current statutory and regulatory authorities. However, the pilot informed what new statutory authorities could support a future regulatory paradigm that builds on these concepts. The pilot reinforced that a systems-based approach that leverages structured objective data can support a learning regulatory system that benefits from data-driven insights to provide efficient and consistent regulatory decisions. Such a system could better enable least burdensome paradigms that provide a reasonable assurance of safety and effectiveness for medical device software. Appropriate new legislative authority would be necessary to support the development and implementation of a new regulatory paradigm. In the meantime, FDA, with leadership from CDRH's Digital Health Center of Excellence, will continue to develop policies and tools within current authorities to improve the efficiency and effectiveness of regulatory oversight, including through collaborative engagement with the public, such as the Medical Device Innovation Consortium (MDIC).

Background

The digital health sector continues to grow as interoperable computing platforms, sensors, and software improve. In particular, software is increasingly being used in the treatment and diagnosis of diseases and conditions, including aiding clinical decision-making, and managing patient care. From fitness trackers to mobile applications, to drug delivery devices that track medication adherence, software-based tools can provide a wealth of valuable health information and insights.

For digital health technologies that meet the definition of a device in section 201(h) of the FD&C Act, FDA identified not only the potential for these devices to improve public health, but also that FDA's regulations are not optimally suited to the manner in which these devices are designed, validated, and improved over time.[6] Specifically, the current statutory framework for medical devices is not well suited to the faster cycles of innovation and the speed of change sometimes necessary to provide reasonable assurance of safety and effectiveness of rapidly evolving devices. The current framework relies on rigid device classifications,[7] resulting in requirements that are not narrowly tailored to the device, nor are they simple to amend based on new information, including device improvements.

The FDA launched the Software Precertification (Pre-Cert) Pilot Program ("the pilot") in July 2017 to explore the possibility for innovative approaches to regulatory oversight of medical device software developed by organizations that demonstrate and maintain a robust culture of quality and organizational excellence and who will monitor real-world performance of their products once they reach the U.S. market.[8] After enrolling 9 pilot participants[9] and conducting a public workshop[10] for all stakeholders, in 2019, FDA issued a number of documents to describe its vision and approach. These documents included the plan for exploring whether the approach could provide a reasonable assurance of safety and effectiveness when compared with the traditional regulatory paradigm, the plan for implementing the pilot, and the plan for the eventual program using current regulatory authorities: the Working Model, Test Plan, and Regulatory Framework, respectively.[11] This report officially concludes the pilot and provides an

---

[6] See FDA's notice, Fostering Medical Innovation: A Plan for Digital Health Devices; Software Precertification Pilot Program, available at https://www.federalregister.gov/documents/2017/07/28/2017-15891/fostering-medical-innovation-a-plan-for-digital-health-devices-software-precertification-pilot.

[7] In 1976, Congress amended the FD&C Act to establish a comprehensive system for the regulation of medical devices intended for human use. Section 513 of the FD&C Act (21 U.S.C. 360c) established three categories (classes) of devices, reflecting the regulatory controls needed to provide reasonable assurance of their safety and effectiveness. The three categories of devices are class I (general controls), class II (special controls), and class III (premarket approval). The classification of a device within one of these three categories determines its regulatory pathway for premarket authorization, if applicable, and postmarket controls that are necessary to provide reasonable assurance of safety and effectiveness (e.g., general controls and/or special controls).

[8] The Digital Health Innovation Action Plan announced the launch of the Pre-Cert Pilot Program, available at http://www.fda.gov/media/106331/download.

[9] The Paperwork Reduction Act (PRA) is a law governing how federal agencies collect information from the American public (see About the PRA at https://pra.digital.gov/about/). Under 5 CFR 1320.5, an agency shall not conduct or sponsor a collection of information unless, among other requirements, the Office of Management and Budget has approved the proposed collection of information. However, 5 CFR 1320.3(c) defines collection of information to include, among other things, reporting "imposed **on ten or more persons**" (emphasis added).

[10] See Public Workshop - Fostering Digital Health Innovation: Developing the Software Precertification Program, available at https://wayback.archive-it.org/7993/20201222110749/https://www.fda.gov/medical-devices/workshops-conferences-medical-devices/public-workshop-fostering-digital-health-innovation-developing-software-precertification-program.

[11] Developing a Software Precertification Program: *A Working Model v1.0,* available at https://www.fda.gov/media/119722/download; Software Precertification Program: 2019 Test Plan, available at https://www.fda.gov/media/119723/download; and Software Precertification Program: Regulatory Framework for Conducting the Pilot Program within Current Authorities, available at https://www.fda.gov/media/119724/download. See also Digital Health Software Precertification (Pre-Cert) Program, available at https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-software-precertification-pre-cert-program.

overview of FDA's experiences during the pilot, including the lessons learned and challenges with testing and implementing the concepts presented in the trio of documents published in 2019 using our current authorities. A new regulatory paradigm to optimize FDA's ability to regulate medical device software and improve public health outcomes would require a legislative change.

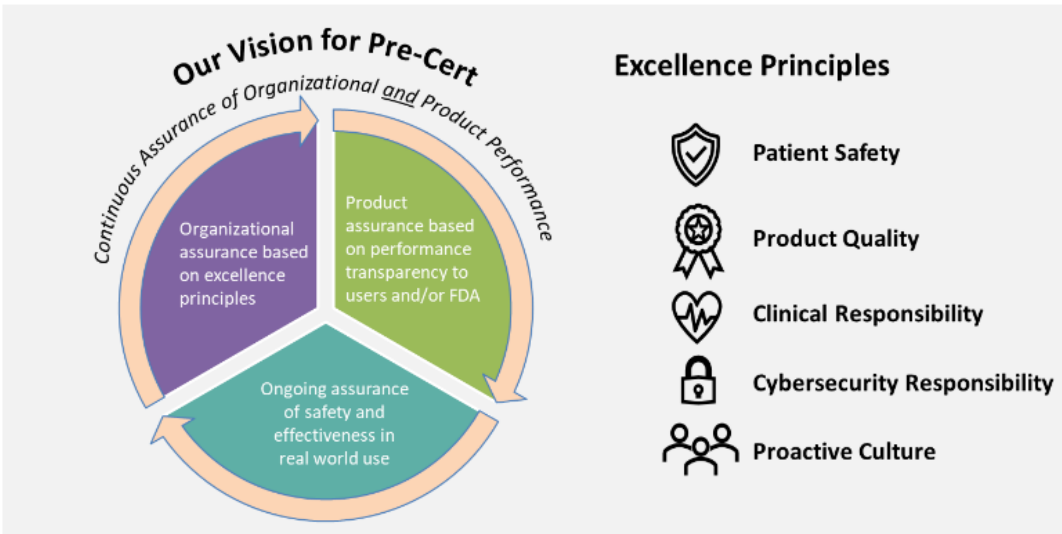

Figure 2. Pre-Cert Program Vision

The Pre-Cert Pilot: A First Step to Inform Development of an Adaptive Regulatory Approach

The Working Model described the goal of the Pre-Cert Pilot as to help develop a "tailored, pragmatic, and least burdensome" approach to regulatory oversight. This approach is based on an assessment of a manufacturer's culture of quality and organizational excellence, ability to develop safe and effective devices, and ability to continuously monitor key performance indicators related to organizational excellence and product performance across the TPLC to verify the continued safety and effectiveness of their devices. Figure 2 provides an overview of the Pre-Cert Program Vision and Figure 3 provides an overview of the Pre-Cert Program Approach throughout TPLC, both of which are further described in the Working Model.

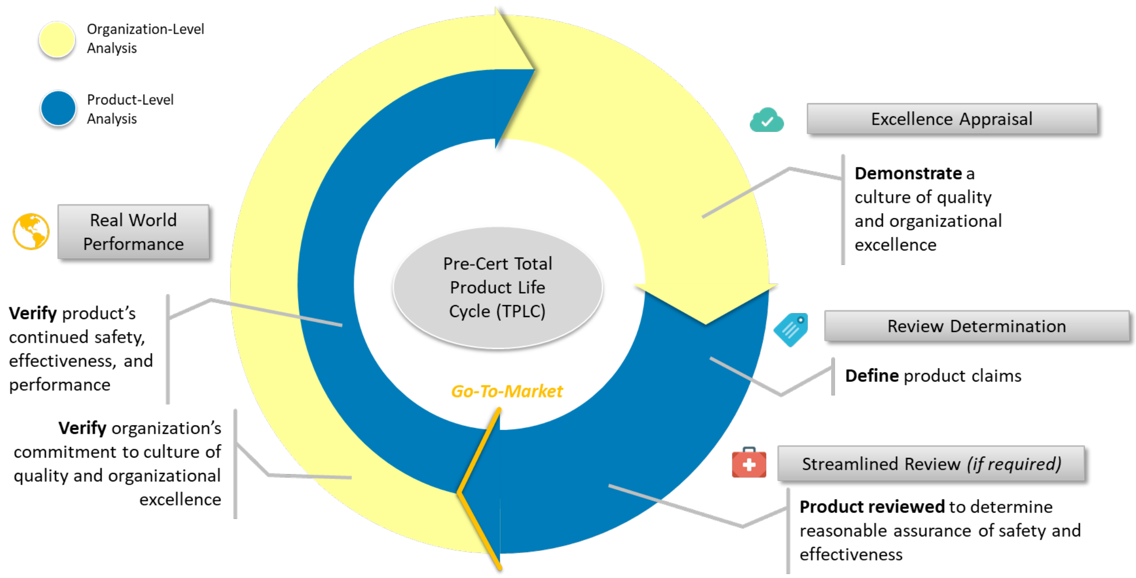

Figure 3. Pre-Cert Program Approach throughout TPLC

Retrospective and Prospective Testing Goals in the Pilot

As part of the pilot, FDA investigated the feasibility of this vision within FDA's current laws and regulations. Specifically, the Test Plan explained that FDA would test the Pre-Cert Program approach by:

1) Internally conducting retrospective tests of SaMD regulatory submissions that have been previously reviewed, and
2) Prospectively conducting tests of SaMD regulatory submissions with volunteers using the De Novo premarket authorization pathway as described in the Regulatory Framework document and comparing the premarket review and regulatory decision to "mock" packages created using the Pre-Cert Program approach.

FDA's goals with the retrospective and prospective tests were twofold:

- Validate elements traditionally included in a device-specific premarket submission using voluntarily provided information on the Excellence Appraisal elements; and
- Test whether the "mock" packages containing information from an Excellence Appraisal and Streamlined Review provided a sufficient basis for determining reasonable assurance of safety and effectiveness as compared to the traditional regulatory paradigm, which in these test cases was a De Novo review. (*Note that testing did not impact the final marketing authorization decision – all premarket submissions proceeded through traditional review pathways.*)

Retrospective Testing Summary

In the retrospective review, FDA found that an Excellence Appraisal summary for use in lieu of certain other premarket software documentation traditionally included in premarket submissions should either be a concise statement of precertification without the expectation of further premarket review of the Excellence Appraisal information, or a detailed report of the Excellence Appraisal process and results to be reviewed in-depth in the context of the device subject to review. High-level summaries between these two extremes were the least viable approach and prompted more questions than they answered, particularly for those reviewers who did not have first-hand familiarity with the pilot Excellence Appraisal process. Retrospective testing is further detailed in the Software Precertification Program 2019 Mid-Year Update.[12]

Prospective Testing Summary

FDA developed and refined the Excellence Appraisal approach with pilot participants based on the feedback from retrospective testing. An update on the Excellence Appraisal approach development and refinement is detailed in Developing the Software Precertification Program: Summary of Learnings and Ongoing Activities: 2020 Update.[13] However, in continuing pilot Excellence Appraisals and conducting premarket reviews during prospective testing, FDA did not find the De Novo submission-based approach outlined in the Test Plan to be the optimal test method for the pilot. Below, FDA reports the lessons learned from validation of the Excellence Appraisal elements and use of the Pre-Cert Program approach for premarket review as a basis for demonstrating a reasonable assurance of safety and effectiveness.

Validation of Excellence Appraisal Elements

The pilot investigated processes for obtaining insight into quantifiable[14] Key Performance Indicators (KPIs)[15] at an organizational level, and whether this information could be used to help streamline premarket reviews. The Excellence Appraisal activities explored the following elements, which are further described in Section 4.2 of the Working Model:

---

[12] Software Precertification Program 2019 Mid-Year Update, available at https://www.fda.gov/media/129047/download.

[13] Developing the Software Precertification Program: Summary of Learnings and Ongoing Activities: 2020 Update, available at https://www.fda.gov/media/142107/download.

[14] "Quantifiable," as defined in Hester, P., Ezell, B., Collins, A., Horst, J.A., and Lawsure, K. (2017). A Method for Key Performance Indicator Assessment in Manufacturing Organizations. *International Journal of Operations Research*. 14(4), 157-167. https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=918303.

[15] "Key Performance Indicator," as defined in Weiss, B.A., Horst, J.A., and Proctor, F.M. (2013). Assessment of Real-Time Factory Performance through the Application of Multi-Relationship Evaluation Design. *NIST Interagency/Internal Report (NISTIR)*, National Institute of Standards and Technology, Gaithersburg, MD. https://doi.org/10.6028/NIST.IR.7911.

- Leadership and Organizational Support
- Transparency
- People
- Infrastructure and Work Environment
- Risk Management: A Patient Safety Focus
- Configuration Management and Change Control
- Measurement, Analysis, and Continuous Improvement of Processes and Products
- Managing Outsourced Processes, Activities, and Products
- Requirements Management
- Design and Development
- Verification and Validation
- Deployment and Maintenance

Conducting Excellence Appraisals

FDA conducted pilot appraisals in-person and through remote interactions. Through these pilot appraisals, FDA sought to assess the association between the organization's software design, development, verification, and validation processes and the organization's general business processes and KPIs. In particular, FDA found that an organizational appraisal and ongoing monitoring focused on KPIs that demonstrate how processes used for the development and maintenance of medical device software are effective and sustained over time provided a consistent, holistic view of the organization's software development processes. Through the pilot, FDA found that this was more informative than a review of the organizational procedures that govern the creation of such data alone. Further, while in-person pilot appraisals went into considerably more depth than remote interactions, FDA found that remote interactions and data reviews were a valuable mechanism for assessing a culture of quality and organizational excellence.

During these pilot appraisals, FDA was able to better understand the practices that pilot participants and others use in designing, developing, and managing digital health products, including:

- Flexible and agile approaches to the software development lifecycle;
- Employment of modern configuration management and version control platforms;
- Robust, automated testing;
- Workflows around continuous integration and delivery;
- Automated infrastructure provisioning, orchestration, and patching;
- New tools and statistical techniques for observational and decentralized trials;
- Enhanced capabilities to conduct A/B testing of different versions of a device in production; and
- More frequent and more detailed software telemetry information that can provide real-time information about product use and malfunctions.

Based on these observations from the pilot, FDA concluded that a new regulatory paradigm, which would require a legislative change, could benefit from certain capabilities and attributes:

- The ability to keep pace with the speed of technology innovation, leveraging information that exists across the TPLC to provide timely assurance of safety and effectiveness of devices, including modified devices, for public health.
- The ability to objectively and continually assess an organization's ability to deliver devices with a commitment to a culture of quality and organizational excellence.
- Ongoing visibility into Key Performance Indicators (KPIs), Real-World Performance (RWP) metrics, and other data that are transparent and objective, enabling timely and targeted actions to resolve issues, creating opportunities to prevent adverse events, and increasing regulatory compliance.
- Regulatory decision support tools that clearly and consistently communicate FDA regulatory policies, which support frameworks for transparent organizational appraisals and communication of device performance by manufacturers to advance safe and effective use of devices by users.

A streamlined premarket review process that leverages regulatory decision support tools (such as those discussed above), uses structured data, and can be adjusted based on postmarket performance, facilitates the delivery of timely, pertinent information, and fosters efficient scientific reviews that ensure the safety and effectiveness of medical devices for the U.S. market. Because FDA sought not to overly constrain the design of the appraisal process too early in the pilot, initial appraisals were not based on pre-specified criteria or KPIs and were instead intended primarily to inform later criteria and KPI selection. This resulted in appraisal outputs that were less specific than would be necessary to adequately inform regulatory decision-making. Should these appraisals become a routinely used method and tool for providing information on an organization's culture of quality and organizational excellence, it will be necessary for FDA to clearly define the elements to be evaluated during an appraisal in advance. This could allow manufacturers to assess how they align with the elements, identify the appropriate tools and resources for the necessary activities to meet these elements, and determine the appropriate KPIs and processes for collecting metrics on the KPIs. Similarly, FDA could develop consistent approaches to conduct appraisals and perform ongoing monitoring by establishing specially trained and dedicated appraisal teams to ensure a consistent and repeatable approach with clear, reliable, and standardized outputs.

FDA also found that software development environments differ based on the specific clinical and technological considerations associated with the devices under development. The pilot appraisals reinforced that the appraisal methods need to accommodate different approaches to medical device software development to balance flexibility with standardization. Based on the lessons learned from the pilot, if appraisals become a routinely used method and tool, the collection of standardized, structured data during an appraisal and ongoing monitoring by FDA or by an FDA-accredited third party could facilitate a consistent and efficient view of an organization's culture of quality and organizational excellence and promote product quality outcomes, including outcomes related to device safety and effectiveness.

Excellence Appraisal Elements – Key Performance Indicators

In the pilot, FDA also investigated using KPIs to assess an organization's capability to develop and maintain safe and effective medical device software throughout the TPLC.

To determine potential KPIs and metrics for such KPIs, FDA began by aggregating metrics from a range of sources, including from the Pre-Cert Program Public Workshop,[16] initial site visits with pilot participants, pilot Excellence Appraisals, comments received from the Pre-Cert Program public docket,[17] and literature. Through FDA's research from these various sources and the pilot, FDA identified 250 KPIs and metrics. Despite the different approaches to software development by each organization, FDA was able to categorize the commonly used KPIs into 9 categories that generally aligned with the Excellence Appraisal elements in the Working Model. Specifically, the 9 KPI categories were:

1. Complaints
2. Data Quality
3. Defects
4. Device Activations / User Adherence
5. Regression Testing
6. Releases
7. Rollbacks
8. Services
9. Security

Should KPIs, and metrics to monitor such KPIs, become a routinely used method and tool, these KPI categories could be useful to assess an organization's capability to develop and maintain safe and effective medical device software throughout the TPLC. FDA further details observations from pilot appraisals in this report, including observations regarding a multidisciplinary appraisal approach (Appendix A, Table 1), organizational processes that embody Excellence Appraisal principles (Appendix A, Table 2), and detailed KPIs and metrics (Appendix B, Table 1) that FDA observed were useful to understand and assess an organization's culture of quality and organizational excellence.

---

[16] See Public Workshop - Fostering Digital Health Innovation: Developing the Software Precertification Program, available at https://wayback.archive-it.org/7993/20201222110749/https://www.fda.gov/medical-devices/workshops-conferences-medical-devices/public-workshop-fostering-digital-health-innovation-developing-software-precertification-program.
[17] See Fostering Medical Innovation: A Plan for Digital Health Devices; Software Precertification Pilot Program public docket, available at https://www.regulations.gov/comment/FDA-2017-N-4301-0001.

Evaluating Reasonable Assurance of Safety and Effectiveness in the Pre-Cert Pilot

The pilot also identified how information is currently used in premarket submissions to make regulatory decisions, investigated the variability of how information is used and opportunities to address this variance, and generated ideas to support Streamlined Review approaches. Based on the outcomes from Streamlined Review during retrospective and prospective testing, modern mechanisms that support clear and consistent communication of device information could facilitate efficient device review activities.

Throughout the pilot, FDA consistently observed the need for structured data and modern regulatory support tools. Though FDA is developing new tools to address these needs, as a result of observations and feedback during the pilot, FDA found that standardized, structured data formats could facilitate the development and use of new manufacturer and FDA regulatory support tools; additionally, it could foster clearer communication between FDA and stakeholders. These regulatory support tools could assist by integrating review workflows and submission content in a more navigable and contextualized manner, which is needed to advance TPLC review practices and account for the increasing complexity of medical device software and their documentation in marketing submissions.

In a separate but related exploration during the pilot, FDA considered tailoring the elements of a Streamlined Review to the medical device and organization, consistent with the pilot's investigation of tailored organizational appraisal approaches. FDA found that organization-level information about product development processes can be considered in a cross-cutting way for multiple devices; however, given the range of potential device indications and technologies, and given the nuances of devising appropriate cybersecurity and clinical validation approaches across this range, FDA found that further development is needed before being able to identify low-risk devices where an organizational appraisal alone could be relied upon without further premarket review of the device. In particular, FDA found that organizational appraisals would not be sufficient to take the place of device-specific clinical performance reviews and cybersecurity reviews for all moderate-risk devices, and device-specific reviews of these elements are of particular value for higher-risk and novel devices. A future approach could build on features of the current regulatory system, where a Quality Management System and other general and, in some cases, special controls provide a reasonable assurance of safety and effectiveness for certain low to moderate risk devices. Such a future approach could incorporate a robust organizational appraisal process for devices of all risk levels that could reduce the need for individual device reviews and enable FDA's remaining individual device reviews to be more streamlined, focusing primarily on a device's cybersecurity, clinical performance data, and other device-specific information that cannot be gleaned through an organizational appraisal.

The Need for New Statutory Authority

The faster cycles of innovation and the speed of change for medical device software would benefit from a new regulatory approach. The challenges faced in completing the Test Plan for the pilot under FDA's current authorities underscore the potential benefits of moving beyond the same medical device legal framework that FDA has operated under since 1976.

Ultimately, the approach to regulating novel, swiftly-evolving medical device software must foster, not inhibit, innovation, while continuing to provide reasonable assurance of safety and effectiveness. These aspects are not mutually exclusive. A flexible, risk-based approach to regulation could allow FDA to tailor regulatory requirements more efficiently for devices based on the latest science, the benefits and risks posed by devices, their real-world performance, and their contribution to promoting health equity. It could leverage the capabilities of evolving medical device software so that health care providers, patients, and users can benefit from advancement and innovation, and so that risk for such devices can be reduced through swift software and cybersecurity updates throughout the TPLC, when needed. New legislative authority establishing such an approach could be supplemental to, and not replace, the established regulatory pathways.

Acknowledgments and Additional Information

FDA recognizes all stakeholders, including pilot participants,[18] test participants, and contributors to the public docket for their input and involvement that helped inform the pilot exploration.

- FDA's Digital Health Program: https://www.fda.gov/medical-devices/digital-health-center-excellence
- FDA's Pre-Cert Pilot Program: https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-software-precertification-pre-cert-program
- Contact: DigitalHealth@fda.hhs.gov
- Report Citation: *The Software Precertification (Pre-Cert) Pilot Program: Tailored Total Product Lifecycle Approaches and Key Findings*. U.S. Food and Drug Administration. September 2022.

[18] See Fostering Medical Innovation: A Plan for Digital Health Devices; Software Precertification Pilot Program public docket, available at https://www.regulations.gov/comment/FDA-2017-N-4301-0001.

Some of the information that FDA has received from pilot participants includes commercial information that is privileged, confidential, or trade secret, which cannot be disclosed (see 21 CFR 20.61). The information in Appendix A and B reflects FDA's findings during the pilot about the types of information that may be useful or informative, although not necessarily comprehensive, when assessing an organization's ability to develop safe and effective medical device software. The observations and findings are general and not specific to any individual company or pilot participant.

Appendix A – Observations from Pilot Excellence Appraisals: Excellence Principles and Key Performance Indicator Objectives

Through the prototyping and testing of the Excellence Appraisal concept, FDA explored whether organizations could demonstrate expertise and a consistent record of implementing their software development, monitoring, and maintenance processes in a manner that results in safe and effective medical device software. Further, FDA investigated whether such data related to these organizational processes could be leveraged across device reviews without repeated submission to FDA for each device. This exploration was informed by three fundamental premises:

(1) Consistent and rigorous application of appropriate software development, monitoring, and maintenance processes results in high quality, safe, and effective medical device software;
(2) Some information currently reviewed in premarket submissions describes software development processes and practices that are applicable to devices across the organization and are not solely device-specific; and
(3) A regulatory framework that ensures continuous and consistent application of, and promotes continuous improvements to, organizational processes can be a more effective and efficient regulatory approach than repeated, standalone device-specific reviews of this information.

Throughout the pilot, FDA explored the Excellence Appraisal concept by conducting several in-person pilot appraisals, reviewing pilot appraisal summaries, and developing an appraisal guide to iterate upon the learnings from the pilot appraisals. FDA also developed an appraisal reporting framework, including potential appraisal queries, in which Excellence Appraisal domains and elements[19] were associated explicitly with

[19] See Section 12 Appendix – Proposed Organizational Elements to Demonstrate Excellence Principles, Developing a Software Precertification Program: A Working Model v1.0, available at https://www.fda.gov/media/119722/download.

generalized quality assurance processes and KPIs. This appendix summarizes FDA's observations on KPIs that support the Excellence Appraisal concept.

Over multiple interactions with pilot participants over several years, FDA identified a number of commonly reported KPIs[20] associated with organizational processes that embodied the Excellence Appraisal principles.[21] The identified KPIs aligned with medical device software safety and effectiveness throughout the device TPLC, and cut across industry, as such KPIs were used by the broad spectrum of diverse organizations in the pilot. In parallel assessments of organizations in the pilot and review of their products, FDA observed that organizations used and relied on KPIs associated with those processes and measures that generate high quality products.

Figure 2- Organizational processes that embody Excellence Appraisal principles observed during appraisals

Through the pilot appraisals, FDA also observed varying levels of organizational process maturity—from the presence and use of organizational processes, to a more mature level, where changes were made to a device based on KPI data. Some levels of organizational process maturity include:

---

[20] See Appendix B – Observations from Pilot Excellence Appraisals: Key Performance Indicators.
[21] See Section 3.4 Software Precertification Program Overview, Developing a Software Precertification Program: A Working Model v1.0, available at https://www.fda.gov/media/119722/download.

1. Organizations were able to demonstrate that documented protocols or processes were **followed**.
2. Organizations were able to identify that KPIs collected and monitored from organizational processes were used to **support** device safety and effectiveness.
3. Organizations were able to identify specific instances in which KPIs were used to **improve** organizational processes and device safety and effectiveness.

To continue to evaluate the Excellence Appraisal concept and organizational process maturity, FDA participated in discussions with pilot participants. This included sessions with individuals and teams performing multiple functional activities at the organization, including software and security engineering, clinical affairs, usability/user experience, quality, regulatory, customer service, human resources, and program/product management. Meeting with the organization's subject matter experts and reviewing the organization's relevant KPIs helped FDA characterize the organization's objectives for the reported KPIs.

Examples of the organizations' subject matter experts that participated in the discussions are included in Table 1. Table 1 also includes the Excellence Principle, examples of the primary and supporting roles with expertise in the corresponding Excellence Principle, and the relevant KPI objectives corresponding to the Excellence Principle. Table 1 is not intended to be comprehensive, but illustrative of the different lenses of the subject matter experts through which KPI objectives were viewed. FDA also understands that not every organization may have the roles as listed; the example roles below are intended to illustrate that a multidisciplinary approach to development, monitoring, and maintenance of organizational processes in alignment with the Excellence Principles may be helpful throughout the device TPLC.

Table 1 – How KPI objectives can be shaped by a multidisciplinary appraisal approach in alignment with the Excellence Principles

Multidisciplinary Appraisal Approach		
Excellence Principle	**Organization Participants**	**Objectives**
Product Quality	***Primary Roles:*** Software Engineering, Information System Security Engineering, Information Security, Usability/User Design ***Supporting Roles:*** Program Management, Security Control	Provide perspective for: • Measuring and monitoring product quality. • Learning about an organization's documented quality metrics and models, particularly around mitigation of defects that affect device safety and effectiveness.

		• Learning about the standard operating procedures (SOP) involved in service downtime and patch releases.
Cybersecurity Responsibility	***Primary Roles:*** Information System Security Engineering ***Supporting Roles:*** Software Engineering, Security Control	Provide perspective for: • Learning about an organization's process to identify potential threats and vulnerabilities, communicate them, and develop countermeasures. • Understanding an organization's security metrics and testing standards.
Patient Safety	***Primary Roles:*** Quality Professional, Information Security, Regulatory, Usability/User Design ***Supporting Roles:*** Safety Engineering, Security Control, Usability	Provide perspective for: • Verifying device compliance and established quality and risk management processes. • Learning about an organization's policies around internal customer/staff and external customer/patient data safety measures.
Proactive Culture	***Primary Roles:*** Quality Improvement, Quality Engineering, Human Resources, Customer Service, Data Analytics ***Supporting Roles:*** Program Management, Usability	Provide perspective for: • Assessing that the proactive programs in an organization go beyond basic-level SOP training. • Verifying quality data collection and analysis for qualitative and quantitative reporting.
Clinical Responsibility	***Primary Roles:*** Quality Professional, Clinical, Data Analytics ***Supporting Roles:*** Software Engineering, Usability	Provide perspective for: • Understanding the clinical staff's role in an organization. • Learning about the interrelationships and communication between data elements, testing, reporting, and external linked devices/software (interoperability).

The pilot demonstrated that understanding of the good practices, culture, and maturity of an organization could help to improve FDA's oversight of medical device software across the TPLC. While each organization is different, FDA observed common practices for developing safe and effective medical device software that produce commonly reported KPIs which can provide evidence of the long-term performance of an organization and its devices. Learnings derived from internal and external feedback of the pilot suggest that a regulatory approach integrating organizational process information calls for consistent, generalizable KPIs. As such, it was important to identify areas of focus for KPI objectives that are impactful, common, and can be assessed in a consistent manner across the breadth of diverse manufacturers.

Appendix List A below identifies the descriptive KPI objectives that FDA observed were useful to understand and assess an organization dedicated to developing safe and effective medical device software during the pilot. In addition, FDA found that these KPI objectives were utilized by the majority of organizations in the pilot. The KPI objectives are grouped into focus areas, where each focus area is grouped by a high-level process and how the process ensures device safety and effectiveness.

Appendix List A: Descriptive KPI Objectives

1. *Processes engage the right people, at the right times, to the right degree: "We use knowledgeable, qualified, and multidisciplinary teams throughout the TPLC."*

- 1.1. Touchpoints among different stakeholders and contributors are established across the TPLC.
- 1.2. Management oversees and is accountable for all aspects of the product throughout the TPLC.
- 1.3. Product requirements are defined by multidisciplinary expertise throughout the TPLC.
- 1.4. Product risk assessment activities include qualified multidisciplinary expertise including clinical, quality, cybersecurity, engineering, human factors, etc.
- 1.5. Vendor and supplier qualifications are established to identify external vendors, contractors, and partners for activities affecting product safety and effectiveness throughout the TPLC.
- 1.6. Product failures are evaluated by all relevant multidisciplinary Subject Matter Experts (SMEs).
- 1.7. Employees and externally sourced contributors are qualified for their respective work unit duties with a focus on product safety and effectiveness (e.g., product development, clinical expertise, risk assessment, etc.) throughout TPLC.
- 1.8. SMEs qualifications are reviewed and updated regularly.

1.9. Organization uses processes to ensure SMEs, including employees and consultants, are adequately tasked to contribute and are accountable for appropriate stages of the product lifecycle.
1.10. Continuity of developing safe and effective products is ensured through planned redundancies and overlap among SMEs and contributors.
1.11. The voice of internal and external stakeholders is integrated into product requirements.
1.12. Human factors engineering is integrated into the TPLC.

2. *Development process results in well-characterized software: "Our software behaves as expected."*

2.1. Safety and effectiveness, regulatory, and market access requirements are established early in the product development process and reviewed throughout the TPLC.
2.2. Engineering and security practices with a focus on software failing safely and visibly are implemented and monitored regularly.
2.3. Secure coding standards and practices to prevent, detect, and eliminate security vulnerabilities that could compromise software security are used.
2.4. Product cybersecurity risks are managed, including performing threat modeling, penetration testing, and specifying data security and encryption requirements, for data at rest and in transit.
2.5. Relevant data sources, including customer reports, public vulnerability sources, and private vulnerability sources, are used for product risk assessment including identifying product vulnerabilities.
2.6. User testing, including longitudinal testing, as appropriate, is conducted using subjects representing the intended end-user groups.
2.7. Probability of occurrence of harm and severity of each hazard and its harm is appropriately assessed and documented.
2.8. Product risk assessment includes the full range of use-case scenarios, including reasonably foreseeable misuse.
2.9. Impact of critical third-party vendors and suppliers is integrated in the risk assessment of the safety and effectiveness of the product.

3. *Deployment and monitoring process confirms well-characterized software in context of use: "Our software behaves as expected in the real-world."*

3.1. Deployment process ensures software is properly installed and configured to operate correctly and safely.
3.2. Products are monitored for safety and effectiveness.
3.3. Product new risks and mitigations are actively identified.

3.4. Product complaints, defects, vulnerabilities, bugs, and safety reports are addressed using a risk-based approach.
3.5. Product users have access to clear, contextually relevant information that is appropriate for the intended audience (such as health care providers or patients) including the product's intended use and indications for use, known limitations, user interface interpretation, and clinical workflow integration.
3.6. Products are traceable post-market to enable action when needed.
3.7. Product use-related (i.e., human factors) failures are evaluated.

4. *Patching process ensures timely resolution of issues across the entire installed base: "We can fix our software when it doesn't behave as expected."*

4.1. Good product software engineering and security practices related to code changes are implemented and monitored regularly.
4.2. Product software updates, including dependency updates (e.g., operating systems, ancillary software, etc.), are evaluated to identify any impact on product performance, intended use, and labeling.
4.3. Regression testing is conducted any time there is new coding, a software change, or an addition to the product software.
4.4. Regression testing is automated and occurs at appropriate points in the development, build, and integration process.
4.5. Regression testing is overseen and approved by appropriate SMEs.
4.6. Product updates are communicated to regulatory authorities per regulations and guidance.
4.7. Product backward compatibility is evaluated to ensure safety and effectiveness are not adversely impacted.
4.8. Product software rollback rationale, method, and timeline are risk-based.

5. *Update process ensures modifications meet user needs identified through real-world use: "We can identify and implement improvements to the expected behavior of our software."*

5.1. Product software defects are identified, tracked, prioritized, investigated, and corrected with validation in a timely and risk-based fashion throughout the TPLC.
5.2. User feedback on product quality and performance issues are actively sought.
5.3. Stakeholders and users are informed of product-related potential quality issues impacting safety, effectiveness, and security of the product, and of the steps needed to ensure safe and effective use of the product.
5.4. Corrective and Preventive Action (CAPA) plans identify actions taken to address the root cause of a problem, include date(s) when actions will be

completed, and include measurement of the action plan effectiveness or amendment to the plan if ineffective.

5.5. Product software quality problems are communicated to those responsible for determining the solution and for preventing future reoccurrence of the problem and are integrated into the CAPA process.

5.6. Systematic errors that cause and/or fail to detect and prevent product software defects, bugs, complaints, vulnerabilities, etc., are identified, investigated, and eliminated.

5.7. Risk assessment is integrated throughout the TPLC.

5.8. Reliable real-world product performance data is used in the development of new software versions or new product lines.

Appendix B – Observations from Pilot Excellence Appraisals: Key Performance Indicators

Through the pilot, FDA identified an initial set of possible KPIs that can provide insight into an organization's software development, monitoring, and maintenance processes, and the impacts these processes have on the safety and effectiveness of medical device software. FDA gathered KPIs that are commonly used in software development across industries from a variety of sources, including the Pre-Cert Program Public Workshop,[22] initial site visits with pilot participants, pilot Excellence Appraisals, comments received from the Pre-Cert Program public docket,[23] and literature. A lack of standardization in vocabulary and specific methods used to describe and calculate KPIs and metrics resulted in FDA's identification of more than 250 different, but overlapping, KPIs and metrics used by different organizations. While these KPIs and metrics were different, they often addressed similar, underlying information and needs for the organization.

The information in Appendix List B highlights examples of FDA's observations of organizations' formulas and data structures of KPIs and metrics from event logs. This type of information provided insights into the culture of quality and organizational

Figure 2 - Common KPIs associated with an organization's software development, monitoring, and maintenance processes, and their impacts on the safety and effectiveness of medical device software

[22] See Public Workshop - Fostering Digital Health Innovation: Developing the Software Precertification Program, available at https://wayback.archive-it.org/7993/20201222110749/https://www.fda.gov/medical-devices/workshops-conferences-medical-devices/public-workshop-fostering-digital-health-innovation-developing-software-precertification-program.
[23] See Fostering Medical Innovation: A Plan for Digital Health Devices; Software Precertification Pilot Program public docket, available at https://www.regulations.gov/comment/FDA-2017-N-4301-0001.

excellence of an organization developing medical device software. While FDA understands that different organizations may track different KPIs and metrics for a variety of reasons (for example, the need to track certain KPIs based on the device technology), based on the experience in the pilot, FDA observed that it was common for organizations developing medical device software to extract KPIs and metrics similar to those identified below. Organizations could extract these types of KPIs and metrics from existing event logs as a key component of their efforts to monitor their organization, processes, and the safety and effectiveness of their devices. FDA observed information characterizing KPIs and metrics which may not be reflected in the list below, such as measurement frequency (i.e., the interval between two measurement points), interpretation (i.e., how the calculated number can be interpreted), acceptable ranges and target values, KPIs/metrics governance, and rules for management escalation for KPIs/metrics out of range. Though not comprehensive, the information in the list below is a representation of the variety of examples of data log events and calculations for KPIs and metrics that FDA observed.

Appendix List B: Example KPI Formulas and Data Structures

1. *Complaints*

Organizations measured complaints, concerns, and questions from users to help provide leads in identifying product defects and to improve product performance.

- 1.1. Examples of derived measure(s)/formula(s):
  - 1.1.1. Rate of complaints: the number of complaints for the product per given time period divided by the number of units sold for the product per given time period
  - 1.1.2. Rate of open complaints: the number of open complaints per product per given time period divided by the total number of complaints for the product per given time period
  - 1.1.3. Complaint risk management efficacy: the number of occurred and previously identified product issues divided by the total number of issues for the product
- 1.2. Examples of observed base measures from data log events and calculations:
  - 1.2.1. Number of total complaints
  - 1.2.2. Number of units sold
  - 1.2.3. Number of product issues
  - 1.2.4. Average and quantiles of time that complaints spend in open state
  - 1.2.5. Number of overdue complaints
  - 1.2.6. Number of repeat complaints
- 1.3. Examples of observed attributes of interest:
  - 1.3.1. Complaint status, (e.g., open, closed, resolved)
  - 1.3.2. Problem classification(s) (e.g., safety, security, quality, alert, adverse event, third party, communication, training, usability, etc.)

1.3.3. Organization response (e.g., resolution, explanation, feature development, investigation, etc.)
1.3.4. Source(s) (e.g., patient, physician, user, professional, public, private, etc.)
1.3.5. Severity (e.g., negligible, minor, serious, critical)
1.3.6. Routing (e.g., team, third party, etc.)
1.3.7. Resolution (e.g., communication, action, regulatory reporting, etc.)
1.3.8. Product identifier(s) (e.g., version, brand name, device model, etc.)

2. *Data Quality*

Organizations measured or verified data retention and integrity, backups, and encryption, at rest and in transit, to ensure that critical data is not visible or had not been altered in an unauthorized manner by destructive malware, ransomware, malicious insider activity, or through inadvertent mistakes that could affect their ability to ensure that data is secure.

2.1. Examples of derived measure(s)/formula(s):
2.1.1. Ratio of data to errors: the number of data errors per data set per given time period divided by the count of the total number of items of the data set per given time period
2.1.2. Data transformation error rate: the number of data transformation operation fails per given time period divided by the total number of the data transformation units per given time period
2.2. Examples of observed base measures from data log events and calculations:
2.2.1. Number of data errors
2.2.2. Number of empty values
2.2.3. Average and quantiles of time data issue open
2.3. Examples of observed attributes of interest:
2.3.1. Systems and processes risk priority
2.3.2. Issue risk priority (e.g., high, medium, low)
2.3.3. Issue types (e.g., duplicate data, inaccurate data, inconsistent data, etc.)

3. *Defects*

Organizations measured defects throughout the TPLC to ensure any defects were resolved in a timely manner.

3.1. Examples of derived measure(s)/formula(s):
3.1.1. Defect density: the number of defects confirmed in software/module during a specific period of operation or development divided by the size of the software/module (size for the purposes of this metric is

usually counted per thousand (1000(K)) source lines of code also known as KSLOC)

3.1.2. Defect resolution success rate: the total number of resolved defects minus the total number of reopened defects divided by the total number of resolved defects

3.1.3. Critical defects rate: the number of critical defects for a given product divided by the total number of defects reported for the given product

3.1.4. Coding standards adherence rate: the number of units of code per product pre-release that comply with coding standards, divided by the total number of units of code for the product

3.2. Examples of observed base measures from data log events and calculations:

3.2.1. Number of defects

3.2.2. Average time defects stay open

3.2.3. Product source code size in KSLOC

3.3. Examples of observed attributes of interest:

3.3.1. Levels of priority (e.g., urgent, high, low, etc.)

3.3.2. Defect type (e.g., functional, performance, usability, compatibility, security, etc.)

3.3.3. Lifecycle phase found (e.g., planning, requirements, design, development, testing, implementation, maintenance, etc.)

3.3.4. Fault code (e.g., administrative, organizational, technical, unique design, testing, etc.)

3.3.5. Correction (e.g., third party, user interface, error handling, boundary, control flow, etc.)

3.3.6. Failure (e.g., new, reopened, duplicate, rejected, etc.)

4. *Device Activations / User Adherence*

Organizations evaluated the interactions between the user and the device on a rolling basis as a means to track device adherence.

4.1. Examples of derived measure(s)/formula(s):

4.1.1. Device activation rate: the number of products activated by users for a given product over a given time period divided by the number of the products sold over the given time period

4.1.2. User training completeness rate: the number of service desk calls from users due to inadequate training or onboarding divided by the total number of service desk calls

4.2. Examples of observed base measures from data log events and calculations:

4.2.1. Number of installations

4.2.2. Number of active installations

4.2.3. Number of users completing intended activity

4.3. Examples of observed attributes of interest:
   4.3.1. Activity completion category (e.g., i.e., successful, unsuccessful)
   4.3.2. Users (e.g., total, active, cancelled, etc.)
   4.3.3. Product identifier(s) (e.g., version, brand name, device model, etc.)

5. *Regression Testing*

Organizations selectively retested device software and measured regression failures to verify that modifications had not caused unintended effects and that the device software still complied with its specified requirements.

5.1. Examples of derived measure(s)/formula(s):
   5.1.1. Regression test failure rate: the number of failed tests re-executed by unit, module, or system not impacted by software change/update divided by the total number of test cases re-executed by unit, module, or system not impacted by software change/update
5.2. Examples of observed base measures from data log events and calculations:
   5.2.1. Number of failed retests
   5.2.2. Number of automated regression tests
   5.2.3. Number of total regression tests
5.3. Examples of observed attributes of interest:
   5.3.1. By function, module, KSLOC, code commits, merges, builds, and/or releases
   5.3.2. Regression tests methods (e.g.,i.e., automate, manual)

6. *Releases*

Organizations measured the efficiency of software releases and upgrades required to deploy services to consumers and/or resolve emergency, safety critical, and/or security issues.

6.1. Examples of derived measure(s)/formula(s):
   6.1.1. Deployment failure rate: number of deployments that have failed in a 30-day rolling period
   6.1.2. Time to restore service: time taken to recover from a failure in deployment
6.2. Examples of observed base measures from data log events and calculations:
   6.2.1. Deployment frequency (expressed per day, per week, or per month)
   6.2.2. Number of releases
   6.2.3. Number of failed deployments
   6.2.4. Number of critical incidents caused by a deployment
   6.2.5. Number of urgent releases
   6.2.6. Time spent per release/upgrade

6.2.7. Number of installed base affected
6.2.8. Number of installed base updated
6.3. Examples of observed attributes of interest:
6.3.1. Release type (e.g., emergency production patches, major coordinated, minor coordinated, major isolated, minor isolated, rollback, etc.)
6.3.2. Release classification(s) (e.g., safety, security, quality, adverse event, etc.)
6.3.3. Release priority (e.g., urgent, minor, etc.)

7. *Rollbacks*

Organizations measured the appropriateness of deployment strategies with active control mechanisms for user safety and security when rollbacks were needed due to a critical issue identified in the field following release.

7.1. Examples of derived measure(s)/formula(s):
7.1.1. Rollback downtime duration: overall duration (end time minus start time) of unavailability of the product being rollbacked per given time period
7.2. Examples of observed base measures from data log events and calculations:
7.2.1. Number of rollbacks
7.2.2. Number of installed base affected
7.2.3. Number of installed base updated
7.3. Examples of observed attributes of interest:
7.3.1. Problem classification(s) (e.g., safety, security, quality, alert, adverse event, third party, communication, training, usability, etc.)
7.3.2. Source(s) (e.g., patient, user, professional, public, private, etc.)
7.3.3. Routing (e.g., team, third party, etc.)
7.3.4. Resolution (e.g., communication, action, regulatory reporting, etc.)
7.3.5. Product identifier(s) (e.g., version, brand name, device model, etc.)

8. *Services*

Organizations measured the performance of the software and support services to ensure these were being delivered appropriately and in a timely manner.

8.1. Examples of derived measure(s)/formula(s):
8.1.1. Availability (proportion of time product is able to perform its function): up time divided by total time by product for a given time period
8.1.2. Rate of service level agreements (SLA) met: the number of SLA met divided by the total number of SLA within a given time period
8.2. Examples of observed base measures from data log events and calculations:

8.2.1. Percent of time software and support service fully functional

8.2.2. Average and quantiles of time software and support service request open

8.3. Examples of observed attributes of interest:

8.3.1. Types of support (e.g., phone and email support, live chat, self-service content, communities and forums, problem solving, technical support, etc.)

8.3.2. Problem code (e.g., false alarm, blank screen, failure to capture, software problem, software failure, display difficult to read, no display/image, false negative result, false positive result, false device output, inadequate instructions, interference, incorrect measurement, incorrect software programming calculations, failure to sense, etc.)

9. *Security*

Organizations measured the number of deployments required to resolve emergency, safety-critical, or security issues.

9.1. Examples of derived measure(s)/formula(s):

9.1.1. Security incident duration: average duration (end time minus start time) for each security incident

9.2. Examples of observed base measures from data log events and calculations:

9.2.1. Number of security incidents

9.2.2. Average and quantiles of time security incidents remain open

9.3. Examples of observed attributes of interest:

9.3.1. Risk (e.g., acceptable, conditionally acceptable, unacceptable)

9.3.2. Incident classification (e.g., cybersecurity incident, reportable, compromised system, etc.)

9.3.3. Priority (e.g., critical, high, moderate, low, etc.)

9.3.4. Condition (e.g., manual, human observation, automatic, cybersecurity alert, physical alarm, etc.)

9.3.5. Source (e.g., internal, public, private, etc.)

9.3.6. Routing (e.g., SMEs, incident management, etc.)

9.3.7. State (e.g., open, new, closed, etc.)

9.3.8. Resolution (e.g., service restoration, escalation, software redesign, system redesign, etc.)

9.3.9. Risk factor (e.g., hazard, condition or state of the environment, hazard exposure or duration, etc.)

9.3.10. Vulnerability category (e.g., design, implementation, configuration, etc.)

9.3.11. Product identifier(s) (e.g., version, brand name, device model, etc.)

Additional KPIs and metrics were observed during the pilot specific to certain TPLC roles and expertise including internal process compliance, static code analysis, and customer and employee satisfaction. Additional exploration is warranted to evaluate the extent to which KPIs and metrics, including those mentioned herein, can be automated, computable, and accessible in real-time (depending on the operational context), and standardized across the diverse spectrum of organizations developing medical device software.

References – KPIs and Metrics

1 Chapter 20: The Software Quality Landscape. Code Complete, Steve McConnell, Second Edition, Microsoft Press, 2004

2 Puppet State of DevOps 2017

3 982.1-2005 IEEE Standard Dictionary of Measures of the Software Aspects of Dependability

4 HSCC Cybersecurity Working Group, Medical Device and Health IT Joint Security Plan, January 2019, https://healthsectorcouncil.org/the-joint-security-plan

5 Benjamin C. Dean, Metrics for Organizational Security Metrics, Metricon X, March 2019; http://www.securitymetrics.org/attachments/Metricon-X-Dean-Metrics-for-Organization-Practices.pptx

6 Boiney, L. G., Connolly, J. L., Skorupka, C. W., Krueger, S. E., Summers, A. Cyber Operations Rapid Assessment (CORA): Examining the State of Cybersecurity Assessment Methodologies and Introducing a New Alternative. MITRE Technical Paper, February 2016. https://www.mitre.org/publications/technical-papers/cyber-operations-rapid-assessment-cora-examining-the-state-of

7 1061-1992 IEEE Standard for a Software Quality Metrics Methodology

8 21 CFR 820, Quality System Regulation

9 Peng, W. W., and Wallace, D. R., NIST Special Publication 500-209, Software Error Analysis, April 1993, https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication500-209.pdf

10 de Vaulx, F., Simmon, E., and Bohn, R. NIST Special Publication 500-307 Cloud Computing Services Metrics Descriptions, April 2018, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.500-307.pdf

11 Flater, D. NISTIR 8289 Quantities and Units for Software Product Measurements, March 2020, https://nvlpubs.nist.gov/nistpubs/ir/2020/NIST.IR.8289.pdf

12 Cawthra, J., Ekstrom, M., Lusty, L., Sexton, J., Sweetnam, J. NIST Special Publication 1800-25 Data Integrity: Identifying and Protecting Assets against Ransomware and Other Destructive Events, December 2020, https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1800-25.pdf