CTTI Quality by Design Metrics Framework

This Quality by Design (QbD) Metrics Framework[1] helps key stakeholders in clinical research organizations to identify measures that quantify[2] the outcomes of QbD implementation and guide continuous improvement efforts. Selection of metrics should consider organizational context. Included in this document are detailed descriptions of nine example metrics (summarized below), each tied to an expected outcome of applying QbD principles to clinical trial design, as well as general considerations for selecting and tracking metrics over time.

Example Metric	Formula	Desired Trend Across Related Studies	Measurable At*	Related QbD Objective
Study Complexity – Endpoints	(# Endpoints Defined in Protocol)	↓ Decrease	▶ Draft study concept ▶ Draft protocol ▶ Final protocol	Streamlining
Percentage of Important Risks Mitigated by Modifying Study Design	[(# of Important Risks Mitigated by Modifying Study Design) / (Total # of Important Risks Identified During Study Design)] x 100%	↑ Increase	▶ Draft protocol ▶ Final protocol	Both
Rate of Patient Enrollment	(# Patients Enrolled) / (# Sites) / (Patient Recruitment Period)	↑ Increase	▶ Intervals until enrollment complete	Streamlining
Rate of Important Protocol Deviations	(# Important Protocol Deviations) / (# Patient Visits)	↓ Decrease	▶ Intervals during study conduct	Fewer 'Errors that Matter'
Rate of Missed Assessments for Key Endpoints	(# Missed Assessments) / (# Expected Assessments)	↓ Decrease	▶ Intervals during study conduct	Fewer 'Errors that Matter'
Rate of Early Terminations	(Total # Early Terminating Patients) / (Total # Enrolled Patients)	↓ Decrease	▶ Intervals during study conduct	Fewer 'Errors that Matter'
Patient Satisfaction with Study Participation	Average (Net Promoter Score)	↑ Increase	▶ Early/mid study conduct ▶ Study closeout	Streamlining
Rate of Avoidable Protocol Amendments	(# Avoidable Substantial Protocol Amendments During "Active" Phase of Study)	↓ Decrease	▶ Study closeout	Both
Number of Major and Critical Audit Findings	(# Critical Audit Findings) + (# Major Audit Findings)	↓ Decrease	▶ Study closeout	Both

**Bullet colors correspond to stages of study planning. For an overview of how QbD applies at each stage, see CTTI's QbD Documentation Tool.*

[1] *This Quality by Design Metrics Framework was developed by the Clinical Trials Transformation Initiative (CTTI) in collaboration with CluePoints.*

[2] *It may also be feasible to use this tool as a starting point for industry-level benchmarking efforts. Additional work would first be needed to standardize metrics across organizations.*

CONSIDERATIONS FOR SELECTING AND TRACKING METRICS

This section outlines one potential approach to identifying a set of quantitative, QbD-related impact metrics to track over time. Although the quantitative impact metrics identified via this tool can be used on their own, they can also be used in conjunction with CTTI's QbD Maturity Model for a broader assessment of how well QbD is being implemented.

Step 1: Select Relevant Metrics

Engage all stakeholders[3] to identify a set of QbD-related metrics to track over time, considering:

a) What are your primary objectives – for example, to show return on investment, or to measure improvement over time?

b) What metrics will be most informative, relative to those primary objectives?

c) Of the highly-informative metrics, which will be sufficiently feasible to analyze on an ongoing basis?

d) Are all selected metrics directly tied to anticipated outcomes of a QbD process (avoiding 'errors that matter' and/or streamlining to increase quality and efficiency)?

ERRORS THAT MATTER

CTTI defines "quality" in clinical trials as the absence of errors that matter to decision making—that is, errors which have a meaningful impact on the safety of trial participants or credibility of the results (and thereby the care of future patients). See additional details in CTTI's Quality by Design Recommendations.

This document includes nine example metrics that, for many organizations, will meet the criteria above (for additional background, see Appendix: Metrics Framework Development Process). Please note, however, that the examples are not intended to be exhaustive, many can be calculated in multiple ways, and not all metrics need to be used. Ideally, an organization will collect only the subset of metrics that are most meaningful and relevant with respect to its objectives for measuring the impact of QbD.

Step 2: Identify Meaningful Comparators

All of the example metrics in this Framework should be calculated at the study level[4], and then compared to one or more of the following:

- Relevant historical data (e.g., same therapeutic area and study phase)
- Concurrent studies for which QbD concepts are not explicitly applied
- Earlier studies for which QbD concepts were applied (i.e., to examine improvement over time)

In each case, it is important to ensure meaningful comparisons are made, considering factors such as disease/condition, primary purpose (e.g., marketing authorization), and study size, and to examine the range of values of baseline comparators for metrics of interest (not just mean or average values).

Step 3: Evaluate Progress Over Time

Effective implementation of Quality by Design should lead to improvement on relevant metrics over time: look for an overall trend toward improvement across a series of studies, recognizing that not every study will improve on every metric. Caution should be taken by organizational leadership to avoid creating incentives with unintended negative consequences. For example, measuring the impact of Quality by Design on rate of patient enrollment should not lead to expectations that every study will have faster enrollment than the last; and safeguards must be established to allow teams to strategically accept and plan for slower enrollment when doing so is critical to the quality of the study.

---

[3] See CTTI's Perspectives for QbD Discussions and Potential Champions for examples of internal organizational perspectives and external perspectives that may be valuable to include in selecting relevant metrics.

[4] Some metrics will also lend themselves to additional levels of analysis, such as drill-downs to examine impact by site, or aggregation to compare across therapeutic areas.

Reduced Study Complexity – Endpoints

Measurable at draft study concept, draft protocol, final protocol

DESCRIPTION	FORMULA
Number of endpoints defined per study protocol	Metric = (# of Endpoints Defined in the Study Protocol)

# of Endpoints Defined in the Study Protocol — Total number of endpoints defined in the study protocol including primary, secondary, and tertiary/exploratory, etc.

Formula Notes

- Consider excluding from this metric any tertiary/exploratory endpoints that require no additional work to collect. Evaluating study complexity should be done with the goal of supporting discussions about whether the planned complexity of a study is justified, and whether the study is as simple and streamlined as it can be.

Considerations for Metric Usage:

- It is recognized that study complexity can be measured in many ways, and a number of other factors might be considered, such as the number of procedures and assessments, number of sites and countries, eligibility criteria, etc. It is strongly encouraged to evaluate these other contributing factors to drive greater overall understanding of the study design factors associated with operational outcomes.
- While applying QbD should generally encourage reduced study complexity, there are likely to be instances in which greater-than-typical complexity scores are entirely appropriate based on the product under study (e.g. gene therapy).
- As with many of these metrics, it is important to compare studies at similar research stage. For example, a larger number of endpoints may be appropriate in the exploratory studies that academic research centers often run; whereas, an organization may expect that global phase 3 studies be streamlined to include the smallest number of endpoints, assessments, visits, countries, etc., that allow the study to accomplish its main objective.
- One potential use of this metric is to compare the number of endpoints proposed initially for a study, with the final number of endpoints in the approved study protocol. Depending on the study design process, this may or may not be meaningful. A pre/post comparison may be meaningful, for example, if protocol designers start by creating a 'wish list' of potential endpoints, and then narrow the list through consultation with additional internal stakeholders (e.g., clinical operations) and external stakeholders (e.g., patients, sites). However, some studies will be designed with the opposite approach, in which only a single primary endpoint is accepted as a given, and adding any other endpoints requires careful justification and cross-stakeholder consultation; in such cases, a pre/post assessment of protocol complexity may not be as meaningful.

Considerations for Data Retrieval:

- Retrieve from final protocol.
- If there is intent to compare the number of endpoints proposed vs. number of endpoints in the final protocol, then plans must be established to record proposed endpoints.

Increased Percentage of Important Risks Mitigated by Modifying Study Design

Measurable at draft protocol, final protocol

DESCRIPTION	FORMULA
Percentage of important risks mitigated by modifying study design	Metric = [(# of Important Risks Mitigated by Modifying Study Design) / (Total # of Important Risks Identified During Study Design)] x 100%

# of Important Risks Mitigated by Modifying Study Design — The subset of important risks that were addressed, in whole or in part, by modifying the study design.

Total # of Important Risks Identified During Study Design — Total number of risks to critical-to-quality factors that were identified during the design of the study.

Formula Notes

- An "important risk" is defined as the potential for errors that have a meaningful impact on the safety of trial participants or credibility of the results. An important risk should be directly tied to an identified critical-to-quality factor (CTQ). Identifying important risks and CTQs for a given study requires discussion by the broad range of stakeholders.
- Note that this metric is not assessing the total number or percentage of risks that were eliminated. Rather, it is intended to assess and demonstrate what portion of risks are being addressed in some manner through updates to the study design – and thereby encourage discussions between study designers and their operational colleagues from the earliest stages of study planning, which can often allow for elimination of important risks entirely, and can reduce the temptation to 'monitor quality in' after the protocol is near-final.

Considerations for Metric Usage:

- This metric requires documentation of QbD processes during protocol design, particularly related to the identification of critical-to-quality factors (CTQs), associated risks, and any mitigations implemented for those risks. This also aligns with the recommendations provided in ICH E6 (R2) (section 5.0.5) to document and communicate risk-based quality management activities.
- An increasing percentage of risks that are mitigated via study design modifications—and fewer mitigated through downstream protocol amendments—may be an indicator that a successful QbD process is in place and functioning as intended. It is also important to recognize, however, that this metric may become less meaningful over time, as study teams apply lessons learned from prior studies to develop higher quality designs from the start.

Considerations for Data Retrieval:

- CTQs, associated risks, and mitigation strategies can be captured in CTTI's QbD Documentation Tool.

Improved Rate of Patient Enrollment

Measurable at intervals until enrollment complete

DESCRIPTION	FORMULA
Rate of patient enrollment per site-month	Metric = (# of Patients Enrolled) / (# of Sites) / (Patient Recruitment Period)

# of Patients Enrolled	Total number of patients enrolled into the study, across all sites
# of Sites	Total number of sites that were initiated and had the opportunity to recruit patients
Patient Recruitment Period	[(Date of Last Patient Enrolled) - (Date of First Patient Enrolled)]

Formula Notes

- For studies where informed consent is followed by a screening period, we recommend assigning as enrollment dates those dates on which patients completed the screening phase and were successfully moved into the "active" study phase (e.g., exposed to study treatment).
- This metric is also applicable to single-site studies. Note that it simplifies the metric formula to: Metric = (# of Patients Enrolled) / (Patient Recruitment Period). This enables research sites to track and assess their rate of patient enrollment per month (for example).

Considerations for Metric Usage:

- It is important NOT to use a target or predicted enrollment rate as the baseline comparator for this metric. Such an approach would fail to distinguish true recruitment challenges from targets that were inaccurate or unrealistic.
- Take care not to create 'perverse incentives' that encourage faster enrollment at the expense of enrolling appropriate patients that fully meet eligibility criteria. Although it can be more challenging to capture accurately, one option would be to adjust this metric to capture the number of patients who were both enrolled *and evaluable*.
- In addition to evaluating the rate of patient enrollment, it may be valuable to evaluate efficiency (e.g., recruitment cost per patient). Enrollment rate alone may nevertheless serve as an adequate proxy for overall efficiency, especially for regulatory submission studies (e.g., faster but more expensive recruitment may provide greater return on investment in terms of expected net present value than would slower but less expensive recruitment).
- Consider assessing screen failure rate as a supporting or additional metric.

Considerations for Data Retrieval:

- Relevant data are typically available from study management systems, including CTMS and EDC.

Reduced Rate of Important Protocol Deviations

Measurable at intervals during study conduct

DESCRIPTION	FORMULA
Rate of important protocol deviations per patient visit	Metric = (# of Important PDs) / (# of Patient Visits)

# of Important PDs — Total number of protocol deviations (PDs) reported during the study that were considered important

# of Patient Visits — Total number of patient visits conducted during the study across all sites

Formula Notes

- The term "important" is chosen to align with the definition of “important protocol deviations” provided in the ICH E3 Q&A document, which supplements guidance provided in ICH E3, section 10.2. An excerpt from the ICH E3 Q&A definition is provided here:

  > “Important protocol deviations are a subset of protocol deviations that may significantly impact the completeness, accuracy, and/or reliability of the study data or that may significantly affect a subject's rights, safety, or well-being. For example, important protocol deviations may include enrolling subjects in violation of key eligibility criteria designed to ensure a specific subject population or failing to collect data necessary to interpret primary endpoints, as this may compromise the scientific value of the trial.”

- “# of Patient Visits” is proposed as a denominator to enable normalization of this metric based on a standard unit of study conduct (patient visits) common to most study designs. The opportunity for PDs to occur is generally proportional to the amount of study activity conducted, and patient visits represent a common "unit of study activity". While not all patient visits represent the same amount of activity, this normalization represents an effective method of assessing this metric at aggregate levels.

Considerations for Metric Usage:

- None

Considerations for Data Retrieval:

- PDs are typically logged (and coded) in a Clinical Trial Management System (CTMS) and/or Electronic Data Capture (EDC) system during the study.
- For academic studies, PDs also typically get reported into an electronic IRB system (per the institution’s policy).
- Important PDs are generally compiled and reported in the clinical study report (CSR) - per ICH E3 guidance.

Reduced Rate of Missed Assessments for Key Endpoints

Measurable at intervals during study conduct

DESCRIPTION	FORMULA
Proportion of expected patient assessments of interest that were missed	Metric = (# of Missed Assessments) / (# of Expected Assessments)

# of Expected Assessments Total number of key endpoint assessments across all patients in the study that were expected to be performed per protocol

# of Missed Assessments The subset of key endpoint assessments that were expected according to the protocol, but were missed, regardless of reason.

Formula Notes

- For purposes of this formula, a "missed assessment" would include assessments missed either because the entire patient visit was skipped, or because the required assessment was simply missed during a conducted patient visit. It would NOT include assessments missed due to early termination of a patient from the study.

Considerations for Metric Usage:

- The formula above is agnostic of the how, by whom, or where the key endpoint assessment was completed—provided that it was completed per protocol requirements—and can therefore accommodate a wide range of evidence generation approaches (e.g., traditional trial, decentralized trial, digital trial). An alternative approach to this metric would be to only count assessments that were missed at conducted patient visits, and to exclude assessments missed because of skipped patient visits. This definition would focus on scenarios in which the site had the opportunity to perform the assessment but failed to do so (i.e., where the patient was available at a visit). It also may be pragmatically easier to derive this form of the metric, rather than trying to account for skipped visits in addition to assessments absent from completed visits. Note, however, that this approach would not be applicable to the growing number of digital health trials that may include few or no site visits.

Considerations for Data Retrieval:

- Data may be available in systems including electronic data capture (EDC), electronic clinical outcomes assessment (eCOA), electronic patient reported outcomes (ePRO), lab data, etc. – whichever system is being used to collect key endpoint assessments.

Lower Rate of Early Terminations

Measurable at intervals during study conduct

DESCRIPTION	FORMULA
Proportion of enrolled patients that terminated the study early	Metric = (Total # of ET Patients) / (Total # of Enrolled Patients)

Total # of ET Patients — Total number of patients who were fully enrolled into the study, and who terminated participation in the study prior to assessment of key endpoints, for any reason other than due to safety concerns

Total # of Enrolled Patients — Total number of patients fully enrolled into the study, after completing any screening or wash-out periods, etc.

Formula Notes

- The formula definition recommends excluding those early terminations that were due to safety concerns, since it is expected that such terminations are driven primarily by the investigational product under study rather than by protocol design factors that could be influenced by QbD.
- The assessment of key endpoints often occurs at an interim timepoint (visit) for patients, followed by a longer-term follow-up period. For such scenarios, this definition proposes to count only those early terminations that occurred prior to that interim timepoint/visit. This enables assessment of the most impactful early terminations.

Considerations for Metric Usage:

- In addition to examining overall early termination rates, we suggest examining the reasons for early termination. As with avoidable protocol amendments, it is important to identify root cause issues and assess opportunities to proactively address these issues – not only for the study in question but for future studies as well.

Considerations for Data Retrieval:

- Identification of and reason for patient early terminations is typically logged in an eCRF form commonly referred to as "Subject Disposition". It should then be available in the EDC database for the study, and also in associated SDTM/SAS datasets prepared with the final study report. Attention needs to be paid to selecting only those terminations that occurred prior to key endpoint measurement (per the above metric definition).

Increased Patient Satisfaction with Study Participation

Measurable at early/mid study conduct, study closeout

DESCRIPTION	FORMULA
Average Net Promoter Score of clinical trial participants	Metric = Average (Net Promoter Score)

Average (Net Promoter Score) The average net promoter score measured across all patients who participated in the study. The **Net Promoter Score** is assessed as, "On a 0-10 scale, how likely is it that you would recommend participation in a clinical research study to a friend or family member?"

Formula Notes

- Net Promoter Score will typically be measured in an online or paper survey, which may include additional questions (see Considerations for Metric Usage, below).

Considerations for Metric Usage:

- Ensure the survey is provided to each participant as soon as possible upon completion of his or her last study visit. Plans should also be in place to ensure participants who leave the study early can complete the questionnaire.
- Consider providing the questionnaire at multiple time points. For example, if planned enrollment is two years, the questionnaire could be distributed at year 1 and year 2. In this way, the metric can also be used as an interim indicator (vs. lagging).
- To understand overall satisfaction scores, consider providing an expanded questionnaire to a random sample of participants. As a rough guide, aim to ask no more than about 10 questions in total, in order to minimize burden and increase response rate. Most questions should be closed-ended (e.g., Likert scale) and may address topics[5] such as:
  - Whether the informed consent process prepared patients for what happened in the study
  - Ease of taking study medication
  - Impact of participation on daily life
  - Whether the number of visits and visit length were reasonable
  - Whether the number and length of assessments/procedures and any questionnaires were reasonable
  - Reasonableness of out-of-pocket expenses
- If additional understanding of patient response is desired, consider inviting survey respondents to provide contact information for participation in a telephone interview.

Considerations for Data Retrieval:

- An appropriate electronic survey platform and/or database will need to be set up.

[5] *Ideas for specific questions can be found in a variety of sources. See, for example, "Research Participant-Centered Outcomes at NIH-Supported Clinical Research Centers" (Kost et al. 2014), and "Study Participant Feedback Questionnaire" (TransCelerate Biopharma, Inc.).*

Lower Rate of Avoidable Protocol Amendments

Measurable at study closeout

DESCRIPTION	FORMULA
Total number of avoidable protocol amendments per study	Metric = (# of Avoidable Substantial Protocol Amendments During "Active" Phase of Study)

# of Avoidable Substantial Protocol Amendments During "Active" Phase of Study

The number of protocol amendments that are considered "avoidable" and "substantial" (see Formula Notes below), and that occur between the date of first patient enrolled and the date of final patient visit (or assessment) for the key endpoint in the study (e.g., prior to long-term follow-up phase)

Formula Notes

- We consider "substantial" protocol amendments as "any change to a protocol ... requiring internal approval followed by approval from the institutional or ethical review board or regulatory authority", following Getz et al. 2016.[6]
- We consider "avoidable" protocol amendments as inclusive of both "completely avoidable" and "somewhat avoidable" amendments as used in Getz et al. 2011,[7] with causes including protocol design flaws, inconsistencies and/or errors in the protocol, recruitment difficulties, and investigator/site feedback.

Considerations for Metric Usage:

- In addition to examining the number of avoidable, substantial amendments, we suggest examining what was changed, determining how the underlying issues could have been anticipated and proactively addressed, and incorporating this analysis into the design of future trials.
- Other definitions of "substantial" and "avoidable" are viable. However, it is critical that definitions be consistently applied (e.g., within an organization or functional group) for this metric to be meaningful.

Considerations for Data Retrieval:

- The number and rationale for protocol amendments will typically be captured in protocol amendment summary documents stored in study files, as well as in a "Version History" section of the final protocol version.

[6] *Getz, K. A. et al. (2016) 'The Impact of Protocol Amendments on Clinical Trial Performance and Cost',* Therapeutic Innovation & Regulatory Science*, 50(4), pp. 436–441. doi: 10.1177/2168479016632271.*

[7] *Getz, K. A. et al. (2011) 'Measuring the Incidence, Causes, and Repercussions of Protocol Amendments',* Drug Information Journal*, 45(3), pp. 265–275. doi: 10.1177/009286151104500307.*

Reduced Number of Major and Critical Audit Findings

Measurable at study closeout

DESCRIPTION	FORMULA
Number of major and critical audit findings	Metric = (# of Critical Audit Findings) + (# of Major Audit Findings)

# of Critical Audit Findings Conditions, practices or processes that adversely affect the rights, safety or well-being of the subjects and/or the quality and integrity of data.

# of Major Audit Findings Conditions, practices or processes that might adversely affect the rights, safety or well-being of the subjects and/or the quality and integrity of data.

Formula Notes

- The word “audit” in the above definition is intended to refer to both sponsor-led audits and inspections conducted by regulators.
- The metric definition uses the terms “Major” and “Critical” since they are commonly recognized categories to indicate more significant (serious, impactful) audit findings. Definitions follow the EMA grading of inspection findings, which also distinguish “Minor” findings: conditions, practices or processes that would not be expected to adversely affect the rights, safety or well-being of the subjects and/or the quality and integrity of data.
- While the specific classification approach for audit/inspectional finding varies by sponsor organization and by regulatory agency, the above definitions provide a good frame of reference for understanding the proposed intent of this metric: to understand the extent to which important risks of ‘errors that matter’ were identified and addressed during study design, rather than attempting to ‘catch’ the errors later.

Considerations for Metric Usage:

- To normalize this metric across a broader range of studies, one option may be to use the following adjusted formula: (# of critical and major audit findings)/(# of site audits). Audits/inspections and associated findings are of course not restricted to investigative sites; rather, this represents a pragmatic approach to help achieve consistent measurement.
- Academic studies typically do not have such formalized inspection/audit processes as with industry-sponsored studies, which may present challenges for assessing this metric. Consideration should be given to establishing a “self-inspection” framework that would enable better identification, documentation and tracking of issues in the conduct of these studies.

Considerations for Data Retrieval:

- The number and reason for audits will typically be captured in an audit database or module of a broader electronic Quality Management system.

Appendix: Metrics Framework Development Process

Overview

The example metrics in this tool were developed by CTTI's multi-stakeholder QbD Adoption project team in collaboration with CluePoints. In brief, the development process was to iteratively generate a list of potential metrics, based on the team's expert knowledge and literature scans; exclude metrics that proved inappropriate or non-relevant to QbD upon inspection; and qualitatively evaluate the remaining metrics' value and feasibility with respect to QbD. Consideration was also given to whether metrics concepts would be better measured via other tools, such as the QbD Maturity Model.

Examples of Additional Metrics Concepts to Consider

Through the course of this work, a number of additional ideas for QbD impact metrics were suggested by diverse stakeholders. Examples are provided below as ideas to consider for further development, though it is possible that some or all will prove challenging to implement in practice.

1) **Process improvement metrics**, such as:
   a) Number of stakeholders or stakeholder groups engaged in study design discussions
   b) % of new studies (by year) that conducted patient and patient advocacy organization engagement/feedback as part of study design process
   c) % of new studies (by year) that conducted assessment of CTQs and associated risks as part of study design process
   d) Investigator satisfaction with QbD process
2) **Shorter-term impact metrics**, such as:
   a) Extent to which identification of CTQs, risks, and potential mitigation strategies resulted in 'significant' changes to the protocol aims or methods
   b) Number of IRB comments/requirements made during protocol review
   c) Faster site recruitment
3) **Longer-term impact metrics**, such as the impact of QbD on:
   a) Funding rates for academic studies
   b) Number of peer-reviewed publications
   c) Regulatory clearance or coverage decisions (e.g., cycle time to obtain regulatory approval)

Examples of Excluded Metrics

Provided below are brief examples of metrics that were identified but excluded from this tool, as well as the rationale for exclusion. Similar considerations may be relevant for organizations developing their own metrics.

Potential Metric	Reason Excluded	Detailed Rationale
Reduced number of adverse events	Inappropriate	To the extent that adverse events are caused by the investigational medical product, an improved study design (i.e., QbD) should have no ability to reduce the number of adverse events.
Increased site reputation	Low feasibility	Although site reputation is important and potentially influenced by study design, it is unlikely to be an effective metric of QbD impact because so many other factors (e.g., clinic hours, friendliness of staff) also play a major role in site reputation.
Accelerated infusion of quality culture	Better measured with other tools	It is a foundational QbD recommendation that organizations should create a culture that values and rewards critical thinking and open dialogue about quality. However, organizations can best assess the infusion of quality culture using CTTI's QbD Maturity Model.