Guiding Principles for conducting Clinical Trial for Machine Learning-enabled Medical Devices (MLMD)

The Ministry of Food and Drug Safety (MFDS) Korea and the Health Sciences Authority (HSA) Singapore have collaboratively released guiding principles for conducting clinical trial for machine learning-enabled medical device (MLMD)[1]. The purpose of these guiding principles is to address the unique challenges posed by MLMD in clinical studies. The MFDS and HSA aim to facilitate the development and assessment of MLMD, ensuring that they meet rigorous standards for safety and effectiveness. In addition, the guiding principles can facilitate the efficient entry of MLMD into the market, while ensuring that patient safety and clinical efficacy remain paramount.

In addition to the Guiding Principles listed in this document, the sponsors are to ensure that the clinical trials adhere to local laws and regulations governing medical research, human subject, and data protection. Compliance with these legal frameworks is imperative to ensure the ethical conduct of trials, safeguard the rights and well-being of participants, and maintain the integrity and privacy of the data collected.

1. Clinical Trial Design

The design of a clinical trial is of paramount importance, shaping the validity, reliability, and ethical conduct of the study. It encompasses critical elements such as the following:

- Selection of clinical trial design configuration (e.g. single-arm, parallel, crossover, etc.);
- Formulation of statistical hypothesis;
- Determination of study population characteristics;
- Randomization and blinding strategies;
- Definition of control groups;
- Identification of primary and secondary endpoints;
- Sample size calculation; and
- Statistical analysis planning.

A well-structured trial design not only ensures the scientific rigor and integrity of the study but also safeguards the welfare of participants and facilitates the generation of robust evidence to inform medical decision-making. By configuring the trial design to align with the research objectives and intended use of the medical device, including context of use within the clinical workflow, researchers can effectively address key scientific questions, minimize unwanted bias, and maximize the clinical relevance and impact of the trial outcomes.

When conducting a clinical trial, we should consider the trial's objectives, product attributes, approach to demonstrating safety, clinical performance and/or effectiveness (such as superiority, equivalence, non-inferiority) of a medical device. In the context of

[1] Machine Learning-enabled Medical Device (MLMD): A medical device that uses machine learning, in part or in whole, to achieve its intended medical purpose. Reference: IMDRF/AIMD WG/N67:2022 Machine Learning-enabled Medical Devices: Key Terms and Definitions

retrospective clinical trials (e.g. using available dataset), the application of parallel or crossover designs can be considered, depending on the specific purpose and objectives of the trial. One should also take note that retrospective clinical trials do not facilitate the assessment of additional factors that could influence the MLMD's performance, such as, device's usability, unintended consequences within the intended clinical workflow, etc. Manufacturer may consider additional studies (e.g. usability study) to address these limitations of retrospective clinical trials.

2. Patient and test dataset selection

Ensuring that clinical study participants or testing datasets are representative of the intended patient population is vital for the validity and generalizability of trial results. To achieve this, clear inclusion and exclusion criteria must be established during the clinical trial design, aligning with the intended use and indication for the medical device. These criteria should encompass the target population, disease groups, disease frequency, gender, and other pertinent factors to accurately reflect the patient population of interest.

When utilizing retrospective or prospective test dataset, it is imperative that they remain independent of the training datasets used during the device's development process.

The determination of an adequate sample size and employing appropriate statistical methods, is crucial for both prospective and retrospective clinical trials. Factors such as the target disease, trial purpose, endpoints, statistical power, and other relevant considerations should be carefully taken into account to ensure the robustness and reliability of the trial findings.

Minimizing unwanted bias in clinical studies for machine learning-enabled medical devices is crucial in both prospective and retrospective study designs. One of the key considerations is to implement randomization and blinding techniques to reduce unwanted bias in patient assignment to different groups and in the assessment of outcomes. This helps ensure that the study results are not influenced by preconceived notions or preferences.

3. Selection of Clinical Reference Standard and Interpretation of Clinical Data

Reference Standard is defined as "*An objectively determined benchmark that is used as the expected result for comparison, assessment, training, etc."* according to IMDRF/AIMD WG/N67:2022 Machine Learning-enabled Medical Devices: Key Terms and Definitions.

Reference standard is typically selected based on existing established recognized clinical guidelines. In the event where there are no clear clinical guidelines due to the novelty of the use case, clinical experts, together with other relevant domain expert, may collectively establish a reference guideline for this novel use case. An additional study may need to be conducted to establish the novel clinical association[2]. This may involve new inputs, algorithms or outputs, new intended target population, or novel intended use of the MLMD.

[2] IMDRF/SaMD WG/N41 (Edition 1): 2017 Software as a Medical Device (SaMD): Clinical Evaluation

When clinical experts interpret complex or ambiguous clinical data derived from a clinical trial against selected reference standard, there are bound to be disagreements in interpretation. These should be addressed among the clinical experts in a systematic and transparent manner. The areas of disagreement, the process of resolution (e.g., expert panel review, adjudication by senior experts, or data-driven consensus), and the rationale behind the final consensus should be documented.

To reduce unwanted bias in clinical study, it is advisable for clinical experts involved in determining the reference standards to be independent from the clinical investigator of the clinical study.

4. Primary endpoint and results analysis

The primary endpoint of a clinical trial for a medical device is the main outcome that the study is designed to evaluate. It is a critical measure that helps to determine the effectiveness and safety of the medical device being tested. The results analysis of the primary endpoint is essential in drawing conclusions about the device's performance and its potential impact on patient health.

The results of the primary endpoint are analysed to determine the device's effectiveness and safety. This analysis involves statistical methods to assess the significance of the observed outcomes and to compare them with the predefined acceptance criteria. This acceptance criteria for evaluating the results of clinical trials can be independently determined by the sponsor, who is required to justify and substantiate the establishment of these criteria.

A clinically meaningful primary endpoint can be based on the following performance indicators such as:

- Sensitivity
- Specificity
- Positive predictive value (PPV)
- Negative predictive value (NPV)
- Number needed to treat (NNT)
- Area Under the Curve (AUC)

These indicators are non-exclusive. They are often used together to assess the performance of diagnostic tests and predictive models when use in clinical settings.

Revision History

Guideline Version (Effective Date)	Revision
First Release (December 11, 2024)	R1.0