Health Canada Santé Canada

1

Draft guidance document: Pre-market guidance for machine learning-enabled medical devices

Canada

Health Canada is the federal department responsible for helping the people of Canada maintain and improve their health. Health Canada is committed to improving the lives of all of Canada's people and to making this country's population among the healthiest in the world as measured by longevity, lifestyle and effective use of the public health care system.

Également disponible en français sous le titre :
Ébauche des lignes directrices préalables à la mise en marché pour les instruments médicaux fondés sur l'apprentissage machine

To obtain additional information, please contact:

Health Canada
Address Locator 0900C2
Ottawa, ON K1A 0K9
Tel.: 613-957-2991
Toll free: 1-866-225-0709
Fax: 613-941-5366
TTY: 1-800-465-7735
E-mail: publications-publications@hc-sc.gc.ca

Publication date: September 2023

Introduction

Artificial intelligence (AI) is a broad term for a category of algorithms and models that perform tasks and exhibit behaviours such as learning and making decisions and predictions. Machine learning (ML) is the subset of AI that allows ML training algorithms to establish ML models when applied to data, rather than models that are explicitly programmed.

Medical devices that use ML, in part or in whole, to achieve their intended medical purpose are known as machine learning-enabled medical devices (MLMD). "Medical purpose" refers to parts (a) through (e) of the "device" definition within the Food and Drugs Act (act). MLMD are subject to the act and associated Medical Devices Regulations (regulations).

In this guidance, "transparency" describes the degree to which appropriate and clear information about a device (that could impact risks and patient outcomes) is communicated to stakeholders. Transparency is an important aspect of the device's safety and effectiveness, and helps stakeholders make informed decisions.

This guidance introduces the concept of a predetermined change control plan (PCCP). A PCCP provides a mechanism for Health Canada to address cases where the regulatory pre-authorization of planned changes to ML systems is needed to address a known risk.

In the face of uncertainties and risks associated with ML and PCCPs, the ongoing safety and effectiveness of marketed MLMD can be strengthened by including terms and conditions (T&Cs) on medical device licences, as appropriate.

Health Canada has adopted the MLMD terms and definitions used by the International Medical Device Regulators Forum (IMDRF). Manufacturers are encouraged to review this document:

- Machine learning-enabled medical devices: Key terms and definitions (IMDRF N67 document)

In this guidance, "ML training algorithm" refers to the software procedure that establishes the parameters of an ML model by analyzing data. The "ML model" represents a mathematical construct that generates an inference or prediction based on new input data and is the result of an ML training algorithm learning from data. The "ML system" refers to an ML-enabled software that meets the definition of medical device as per Section 1 of the regulations, including ML models and the associated ML training algorithms.

Scope and application

This document provides guidance to manufacturers who are submitting a new or amendment application for Class II, III and IV MLMD under the regulations.

The information in this guidance relates to the ML system of an MLMD. It does not cover the non-ML information required in a medical device licence application.

Manufacturers should also consult other relevant guidance relating to medical devices, including the following:

- Guidance on supporting evidence to be provided for new and amended licence applications for Class III and Class IV devices, not including in vitro diagnostic devices (IVDDs)
- Class 3, in vitro diagnostic devices (IVD), new and amendment applications
- Class 4, in vitro diagnostic devices (IVD), new and amendment applications
- Software as a medical device (SaMD): Definition and classification
- Pre-market requirements for medical device cybersecurity – Summary
- Guidance for the interpretation of significant change of a medical device
- Guidance on clinical evidence requirements for medical devices

Policy objective

This guidance outlines supporting information to consider when manufacturers are demonstrating the safety and effectiveness of an MLMD:

- for the purposes of applying for or amending a Class II, III or IV medical device licence or
- at any other point in the device lifecycle

Policy statements

An MLMD can be standalone software that meets the definition of a medical device. It can also be a medical device that includes software that meets the definition of a medical device.

An MLMD can be an *in vitro* diagnostic device (IVDD) or a non-IVDD. The risk classification of an MLMD can range from Class I to Class IV.

Manufacturers should clearly state that the device uses ML in their cover letter for all Class II, III and IV applications for an MLMD. Furthermore, for MLMDs that have a PCCP, manufacturers should clearly state in their cover letter that their device includes a PCCP. Excluding such statements could delay the application process.

Manufacturers should include a justification for the proposed medical device classification applied to the MLMD. This justification should reference the classification rules outlined in Schedule 1 of the regulations.

Medical devices must meet the applicable requirements of sections 10 to 20 of the regulations. Manufacturers must ensure that objective evidence is available to support the intended use of the MLMD, the safety and effectiveness of the device and the associated claims.

An application must demonstrate that the MLMD (including the PCCP, as appropriate):

- meets, and will continue to meet, applicable safety and effectiveness requirements
- will maintain a high level of protection of health and safety and an acceptable level of risk when weighed against the benefits to the patient

Class II, III and Class IV applications must include the information listed in section 32 of the regulations. Additional information may be requested at any time during our review of an application (new or amendment) or after a device has been licensed.

Health Canada understands that manufacturers may use a variety of information, methodologies and evidence to demonstrate that their MLMD is safe and effective. Additionally, different intended uses or risk profiles may require different types or levels of evidence. As such, we have outlined information for consideration rather than prescribing the required information for all scenarios.

The guidance for implementation section of this document outlines the information to consider for an MLMD. If any of the information identified in this section is not available, manufacturers should offer a justification or provide alternative information, as applicable.

Data referred to or used by manufacturers should adequately represent the Canadian population and clinical practice. Any data used to develop the MLMD or demonstrate a device's safety and effectiveness should reflect the population for whom the device is intended. For example, this could include consideration of skin pigmentation, biological differences between sexes and other identity-based factors.

For those devices that are authorized with a PCCP, subsequent changes made according to the authorized PCCP do not require that you submit a medical device licence amendment application. PCCP-driven changes are subject to relevant post-market regulatory oversight.

For amendments to a device that are outside of an authorized PCCP, including changes to the PCCP itself, the regulations and relevant guidance documents should be consulted before implementation. It’s important to determine whether the change constitutes a significant change and requires an application for a medical device licence amendment.

A PCCP may be submitted with applications for a new medical device licence or a medical device licence amendment.

This guidance represents Health Canada’s current thinking. We will revise this guidance and adapt our policy approach as the technology matures and the regulatory oversight has been optimized.

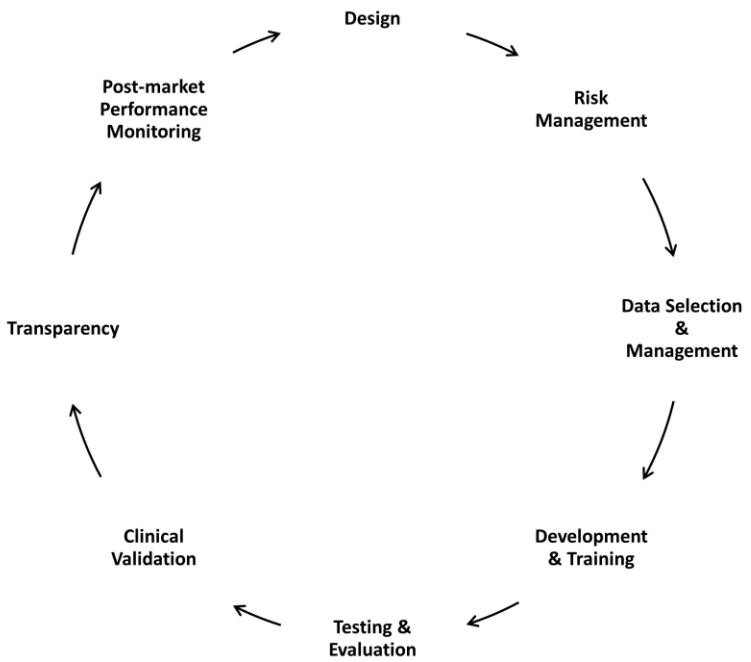
Guidance for implementation

Health Canada considers product lifecycle information to be essential in demonstrating the safety and effectiveness of an MLMD. From our perspective, the MLMD lifecycle includes the following components:

- design
- risk management
- data selection and management
- development and training
- testing and evaluation
- clinical validation
- transparency
- post-market performance monitoring

Error! Reference source not found. gives a visual overview of the content areas discussed in this document. H owever, the iterative components reflected in this circular lifecycle schematic are not mutually exclusive and may not occur in the order indicated.

Figure 1: MLMD product lifecycle



Alt text:

The MLMD product lifecycle is represented in a circle with 8 stages, illustrating an iterative process. The stages are:

- design
- risk management
- data selection and management
- development and training
- testing and evaluation
- clinical validation
- transparency
- post-market performance monitoring

Good machine learning practice

Good machine learning practice (GMLP) is important when designing, developing, evaluating, deploying and maintaining an MLMD. This helps to ensure safe, effective and high-quality medical devices.

The evidence provided with an application for an MLMD should include a description of how the manufacturer has adopted GMLP within the organization and implemented it throughout the product lifecycle. If applicable, this description should outline the quality practices implemented to ensure that the PCCP change description will be realized by following the PCCP change protocol.

Pre-determined change control plan: concept

A PCCP is the documentation intended to characterize a device and its bounds, the intended changes to the ML system, the protocol for change management and the change impacts. If included, a PCCP is considered part of the device design.

PCCPs should be risk-based and supported by evidence, take a total product lifecycle perspective and provide a high degree of transparency.

All modifications listed in a PCCP must ensure that the device continues to operate within its intended use. Changes listed in a PCCP should not include changes to the medical conditions, purposes or uses of an MLMD. Such changes require a medical device licence amendment application prior to implementation.

Appropriate changes to list in a PCCP include those where pre-authorization is necessary to address a known risk while upholding the benefits to the patient. An example of such a change would be the maintenance or improvement of performance to address the risk of ML performance degradation over time. This performance degradation can be due to changes to the environment, such as to the input data or the relationship between the input variables and the target variable.

The use of a PCCP allows timely and ongoing management of risks while retaining high regulatory standards to ensure device safety and effectiveness.

Sex and gender-based analysis plus

Sex and gender-based analysis plus (SGBA Plus or GBA Plus) is an analytical process used to assess how a product or initiative may affect diverse groups of people. This process can be incorporated into the risk management approach used across the lifecycle of the device.

Evidence demonstrates that biological, economic and social differences between diverse groups of people contribute to differences in health risks and outcomes, their use of health services and how they interact with the health system. Integrating SGBA Plus throughout the lifecycle of a medical device will lead to more equitable health outcomes for Canada's diverse population.

Over the lifecycle of the MLMD, manufacturers should apply SGBA Plus and consider the unique anatomical, physiological and identity characteristics of patients. This includes:

- taking into consideration sex and gender, racial and ethnic minorities, elderly and pediatric populations, and pregnant people
- collecting and analyzing disaggregated data on sub-populations in clinical studies, training data and test data, as appropriate

Design

Indications for use, intended use and contraindications

For any Class II, III or IV MLMD, the intended use or medical purpose should be made clear in the application. Provide all relevant information, including the following:

- the intended use and/or indications for use of the MLMD
- the medical purpose (for example, diagnosis, treatment, monitoring) and the intended conditions, diseases or disorders
- the intended patient population
- the intended user
- the intended use environment
- device function information, as applicable, including:
    - software inputs
    - software outputs
    - an explanation of how the software output fits into the healthcare workflow
    - the clinical degree of autonomy
        - the capacity to perform a clinical function with no or limited clinical user intervention
- contraindications
- all known limitations

Device description

Provide a detailed description of the MLMD, including any ML systems used to achieve an intended medical purpose. Consider including the following information in the description of the device or software:

- a statement that the device uses ML, which should also be included in the cover letter
- if applicable, a confirmation that the MLMD includes a PCCP, which should also be included in the cover letter
- a detailed description of the ML methods and ML training algorithms
    - ML methods such as supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning
    - ML training algorithm(s) such as convolutional neural network, logistic regression, language models or support vector machines
- a description of the ML system output, intended users, how the output is intended to be used within the healthcare workflow and the clinical degree of autonomy
    - the capacity to perform a clinical function with no or limited clinical user intervention
- an explanation of how the ML system works, the known factors influencing the output and the interpretation of the system behaviour, if available
    - for example, feature attributions to ML model predictions, how the outputs of the ML model are impacted by changing input properties, saliency maps
- descriptions of the following:
    - required device input parameters, input specifications and source(s) of device input(s)
    - all compatible medical devices, including software and hardware versions
    - hardware requirements (for example, CPU requirements, operating system)

Predetermined change control plan: content

A PCCP consists of the following 3 components:

1) Change description
2) Change protocol
3) Impact assessment

The detailed PCCP, if applicable to the device, should:

- be a standalone section in the submission, typically within either the ‘device description’ or ‘software’ section
- include references to any application information related to the PCCP that’s outside of the PCCP section, such as in the labelling or evidence used to demonstrate safety and effectiveness
- consider the information outlined in the following 3 sections

Change description

The change description is the documentation that characterizes the device and the proposed changes. It includes:

- a description of the initial baseline device design and performance as well as the design and performance envelope or limits over time:
  - such as performance specifications and associated performance thresholds, inputs, outputs and relevant technical specifications
- a list of specific changes to the MLMD that are proposed for pre-authorization that would otherwise be significant changes in the absence of an authorized PCCP
- with each change listed, a detailed description of the following:
  - motivation, rationale or trigger for the planned changes
    - for example, performance thresholds, scheduled time intervals, user feedback
  - cause or source of the changes to the device
    - for example, re-training with new or appended data
  - effect of the changes on the device
    - for example, modified performance, changes in device inputs or outputs
  - where the changes apply
    - for example, uniformly across all marketed devices, non-uniformly across marketed devices based on unique characteristics of a clinical site or patient
  - who will make the changes
    - for example, manufacturer, qualified clinical user, non-clinical user, patient, automatically by the software
  - planned frequency of changes
  - any anticipated modifications to the device description, labelling, user interface

Change protocol

The change protocol describes the set of policies and procedures that control how changes, as outlined in the change description, will be implemented and managed. The protocol ensures ongoing safety and effectiveness.

Aspects of the change protocol that may need to be part of the licence application include plans for ongoing:

- Data management
  - may include, for example, plans for collecting, annotating, curating, validating, determining reference standard or ground truth, quality assurance
- Risk management
  - may include, for example, plans for ongoing risk identification, monitoring and response
- Modification procedures
  - may include, for example, plans for re-training, learning techniques, update triggers, pre-update verification and validation methods, such as ML system performance validation and its impact on the performance of the MLMD if applicable
- Update procedures
  - may include, for example, version tracking and control such as traceability, ongoing documentation of the PCCP execution history, deployment plan, end-user communication plan, labelling update plan and user acceptance testing
- Monitoring

    - may include, for example, plans for post-update testing and performance monitoring, frequency of assessments and triggers for evaluation, statistical analysis plan, plans for device surveillance, complaint handling and reporting incidents
- Corrective actions
    - may include, for example, roll-back plans, backup and recovery procedures, retraining criteria and objectives, and customer communications

Each change in the change description should be clearly traceable to the relevant aspects of the change protocol (for example, through a traceability table).

Impact assessment

The impact assessment outlines the potential influence and implications of the changes listed in the PCCP. It should consider:

- the benefits and risks of implementing the PCCP and the risk controls in place
- how the change protocol will continue to ensure the ongoing safety and effectiveness of the device
- the collective impact of all proposed changes on the MLMD and the impacts on other elements of the clinical workflow, including on other medical devices

Risk management

Manufacturers should conduct the necessary risk management and consider providing descriptions of:

- the risks identified for the MLMD and the associated risk controls in place to eliminate or reduce those risks
- the technique used to perform the initial and ongoing risk assessment and the system used for risk level categorization and acceptability
- the results of the risk assessment

The following items, as applicable, should be considered in the risk analysis:

- erroneous outputs
    - such as false positive or false negative results, or incorrect information for use in diagnosis or treatment
- bias
    - note that SGBA Plus analysis may address some sources of unwanted bias
- overfitting
    - an issue that occurs when a model is fit to properties that are specific to the training examples (for example, random noise), resulting in a model that does not apply to the general problem it's meant to address
- underfitting

    - an issue that occurs when a model is not fit to all relevant properties of the population from the training examples, resulting in a model that does not apply to the general problem it's meant to address
- degradation of ML system performance
    - an issue that can occur due to shifts in population demographics or disease incidence, changes in clinical practice, changes in clinical disease presentation, changes in input format or quality
- automation bias
    - an issue that occurs when a user's conclusion is overly reliant on the device output while ignoring contrary data or conflicting human decisions
- alarm fatigue
    - an issue that occurs when a user is desensitized to alarms due to excessive exposure, which can result in missed alarms
- risks associated with using a PCCP
- impacts of a PCCP on risk management

When performing the risk management for an MLMD, consider referring to the current version of the following resource:

- ISO 14971, Medical devices - Application of risk management to medical devices

Data selection and management

When describing the selection and management of data for an MLMD, consider providing the following elements:

- descriptions of the training, tuning and test datasets used to develop and evaluate the ML system, such as:
    - sample sizes with and without the condition, clinical characteristics and demographic statistics
    - a comparison between the prevalence within the dataset and the intended population
    - methods and environments in which the data were collected
    - data collection devices
    - single versus multi-centre data, personalized data
    - justifications to support the dataset characteristics, for example, according to:
        - their relation to the intended use
        - statistical considerations
        - identity factors (such as sex, gender, race or age)
        - consideration of subgroups, such as vulnerable or under-represented populations
- data inclusion and exclusion criteria and a justification for removing any data
- descriptions of techniques used to address data imbalances (for example, specific sampling methods used to address a dataset that has low disease prevalence) and a justification
- a description of how data integrity was maintained during curation and how data quality and accuracy were ensured, including a description of any data augmentation practices
    - for example, geometric transformations intended to enhance the size and quality of datasets
- an explanation of how bias in the dataset was controlled during development

Development, training and tuning

Consider providing descriptions of the ML development, training and tuning approaches, including the following elements:

- a detailed description of the methods used to develop, train and tune the ML system and a justification to support these methods
- a characterization of the reference standard used in training and tuning, including:
    - the process and methodology used to define the reference standard
    - a justification to support the chosen reference standard
    - a description of the uncertainty and associated limitations
- a description of the inputs and parameters used to develop the ML system and any features extracted from the input data

Testing and evaluation

Consider including the following information on ML system performance testing as part of the performance/bench testing or software verification and validation:

- a description of the methods used to test or evaluate the ML system performance
- a characterization of the reference standard used in testing, including:
    - the process and methodology used to define the reference standard
    - a justification to support the chosen reference standard
    - a description of the uncertainty and associated limitations
- descriptions of the chosen performance metrics, acceptance criteria and operating point/threshold, with clinical and risk-based justifications
- evidence to demonstrate that the ML system performs as intended and meets expected performance requirements when integrated as part of the medical device system or software
- evidence to support the performance of the ML system for appropriate subgroups, including at the relevant intersections, for example according to:
    - identity factors (such as sex, gender, race, age)
    - vulnerable populations
    - under-represented populations
    - clinical status (such as diagnosis, stage, grade)
    - clinical features (such as tissue density, lesion type, co-occurrence of conditions)
- evidence to support inter-compatibility with all supported input and output devices
- robustness testing
    - for example, intentional testing with unexpected inputs
- estimate of the uncertainty of the outputs, with supporting evidence and a justification to support the method used to determine the uncertainty
- the ML software version that was tested, which should represent the appropriate release version
- an explanation of the software version numbering system and the identification and traceability of the ML system or model version

Clinical validation

In a medical device licence application for a Class III or IV MLMD, manufacturers should provide the appropriate clinical evidence, including clinical validation studies, to support the safe and effective clinical use of their device. This information should be available upon request for Class II MLMD.

For more information on clinical evidence requirements, consult:

- Guidance on clinical evidence requirements for medical devices
- Companion document: Examples of clinical evidence requirements for medical devices

The clinical evidence should support that the trained, tuned and tested ML system, and the MLMD with that ML system, is safe and effective and performs as intended in the intended population.

Examples of clinical evidence that can be used include:

- clinical validation studies, including descriptions of:
  - the type of study performed
  - the study design and statistical methods
  - the rationale for the study and methods, including the use of retrospective and/or prospective evaluations
  - a characterization of study participants and confirmation that the study population is independent of the data used for ML system development, training and tuning
  - the rationale for the study population, which may include:
    - the relation to the intended use
    - the representation across sex, gender, race, age and/or other identity factors
    - statistical considerations
  - study results
- relevant clinical data from published sources
- device-related investigations
  - for example, comparator device clinical data
- usability/human factors testing
- device-specific evaluations
- real-world evidence (RWE) and post-market clinical experience

The clinical evidence should accompany a justification to support the level of evidence. This justification should establish that the evidence is sufficient to demonstrate:

- the device is safe and effective for the intended population when used as described in the 'intended use' or 'indications for use' statement
- as appropriate, the impacts of the device on different sexes, genders and diverse populations, including racial and ethnic groups, and pediatric and older populations

Transparency

Transparency requirements should consider the various stakeholders involved in a patient's healthcare across the lifecycle of the device (for example, patients, users, healthcare providers and regulators).

Transparency should be considered throughout the device lifecycle, including in the:

- design of the device, including:
  - the ML system, software user interface, labelling and, if applicable, the PCCP
- medical device licence application
- device marketing
- device use

The following subsection outlines transparency considerations for MLMD labelling for the end-user.

Labelling

Manufacturers should provide copies of the directions for use or instructions for use for the device, including those pertaining to the ML system. Health Canada will review the labels against the requirements outlined in sections 21, 22 and 23 of the regulations.

The following ML system information should be considered for inclusion in MLMD labelling, as applicable:

- **Indications for use, intended use and contraindications** (refer to the section under Design)
- **instructions for the user**, such as:
  - how to use the ML system software to generate an output
  - how to interpret the software interface, including:
    - the ML system output and any information provided to help users interpret each output (for example, saliency maps and confidence scores)
  - how to perform calibrations, local validation and ongoing performance monitoring
- **device design information**, such as:
  - a statement that the device includes ML
  - how the ML system works, for example:
    - ML approaches
    - feature attributions to ML model predictions, factors influencing the output, if available
  - required device input parameters, input specifications and source(s) of device input(s)
  - compatible medical devices, including software and hardware versions
  - hardware and software requirements (for example, CPU requirements, operating system)
  - dataset characterizations of training and test datasets, such as:
    - data collection environment/method
    - determination of reference standard
    - sample sizes with and without the condition, clinical characteristics, demographic statistics
    - inclusion/exclusion criteria

- PCCP information, if applicable, such as:
  - a statement that the device includes a PCCP
  - the intended changes and expected update frequency
  - any requirements for the user to perform software updates
  - when a software update occurs and how it impacts the device performance, inputs, labelling or use (for example, how to obtain updated labelling or how improved performance will be communicated to them)
- **device performance information**, such as:
  - chosen performance metrics and acceptance criteria as well as the operating point/threshold
  - detailed results of the performance testing, including results for appropriate subgroups and the performance uncertainty (for example, confidence intervals)
  - summaries of clinical studies, if applicable, including detailed characterization of the study participants, methods and results
- **device limitation information**, such as:
  - data characterization limitations
  - limitations in the development techniques
  - limitations in the performance evaluation
  - known failure modes
  - applicable warnings or cautions related to the ML system

Product brochures, websites and marketing material with claims related to the ML system should also be provided, as these are also considered labelling.

Post-market performance monitoring

Manufacturers should consider including a description of the processes, surveillance plans and risk mitigations in place to ensure ongoing performance and inter-compatibility of the ML system.

This should consider the impact on ML system outputs or clinical workflows that could result from potential changes in the inputs to the ML model, changes to how the ML system outputs are handled by compatible products or any other relevant information. This may be addressed as part of the risk analysis and the PCCP, if applicable.

Terms and conditions

Terms and conditions (T&Cs) may be imposed on some medical device licences. This can help ensure that the device continues to meet the applicable safety and effectiveness requirements of the regulations after it's been approved.

As per section 36(2) of the regulations, the Minister may impose T&Cs requiring:

- tests to be performed on a device to ensure it continues to meet applicable safety and effectiveness requirements
- submission of the results and protocols of any tests performed

As per subsection 36(3) of the regulations, the Minister may amend T&Cs imposed on a medical device licence to take into account any new development with respect to the device.

The holder of a medical device shall comply with T&Cs of the licence as per subsection 36(4).

The level of risk, uncertainty and/or complexity of a specific situation will be considered when imposing or amending T&Cs, and when determining requirements for individual T&Cs.

Related links

- Software as a medical device (SaMD): Clinical evaluation
- Machine learning-enabled medical devices: Key terms and definitions
- Good machine learning practice for medical device development: Guiding principles
- What is Gender-based Analysis Plus