A Study of Reviews of Software as a Medical Device (SaMD) Utilizing AI

- Final Study Report -

Regulatory System Committee of the Japan Federation of Medical Devices Associations
Study Working Group related to Reviews of Software as a Medical Device Utilizing AI

TABLE OF CONTENTS

1. Introduction

1-1. Preface

Technological innovation in artificial intelligence (AI) has advanced dramatically, bringing innovation in various fields. Particularly in the medical field, its expected usefulness is recognized in promptly implementing diagnostic imaging and disease diagnosis, optimizing diagnosis support and treatment plans through the approval of medical devices applying deep learning, and promoting the development of medical devices utilizing AI. This report discusses issues and possibilities for future development while summarizing the current review results and perspectives concerning review requirements for Software as a Medical Device (SaMD) utilizing AI.

1-2. Background of activities

In October 2022, the “Study Working Group related to Reviews of Software as a Medical Device Utilizing AI” was launched as a working group under the umbrella of review subcommittees of the Legal Affairs Committee of the Japan Federation of Medical Devices Associations (JFMDA). The Working Group (WG) has continued discussions aimed at promoting transparency in the review process through discussions and examination with PMDA personnel in charge of reviews and standards as well as releasing the results on the website, etc., as needed based on the output from the “Study on Pharmaceutical Regulations for Software as a Medical Device Utilizing Advanced Technology Such As Artificial Intelligence” conducted by the Japan Agency for Medical Research and Development (hereinafter referred to as AMED), while considering the requirements of reviews of Software as a Medical Device (SaMD) utilizing AI (including medical devices and programs) and the current status of reviews.

Prior to each WG meeting, five sub-groups within the WG compiled opinions on issues concerning SaMD utilizing AI that were set in advance. Then, each group presented and shared opinions and proposals during WG meetings and exchanged opinions. In exchanging opinions, feedback was obtained from the PMDA personnel and AMED Regulatory Science (RS) research representatives participating in WG from their respective standpoints.

1-3. Items examined

Prior to the WG meeting, a questionnaire survey was conducted among participating members on the priority issues to be examined. The results are presented in Figure 1 (number of valid respondents = 17). The respondents exchanged opinions on the review points for medical devices using AI with the Office of Software as a Medical Device of the PMDA using the “Review Points for Computer-Aided Diagnosis Program to Support Interpretation of Medical Images*” available on the PMDA’s website, which deepened our understanding of points to note in reviews.

In addition, continued discussions were held on the (1) descriptions in the approval application form, (2) procedures when approved items are changed, and (3) AI eligible for certification.

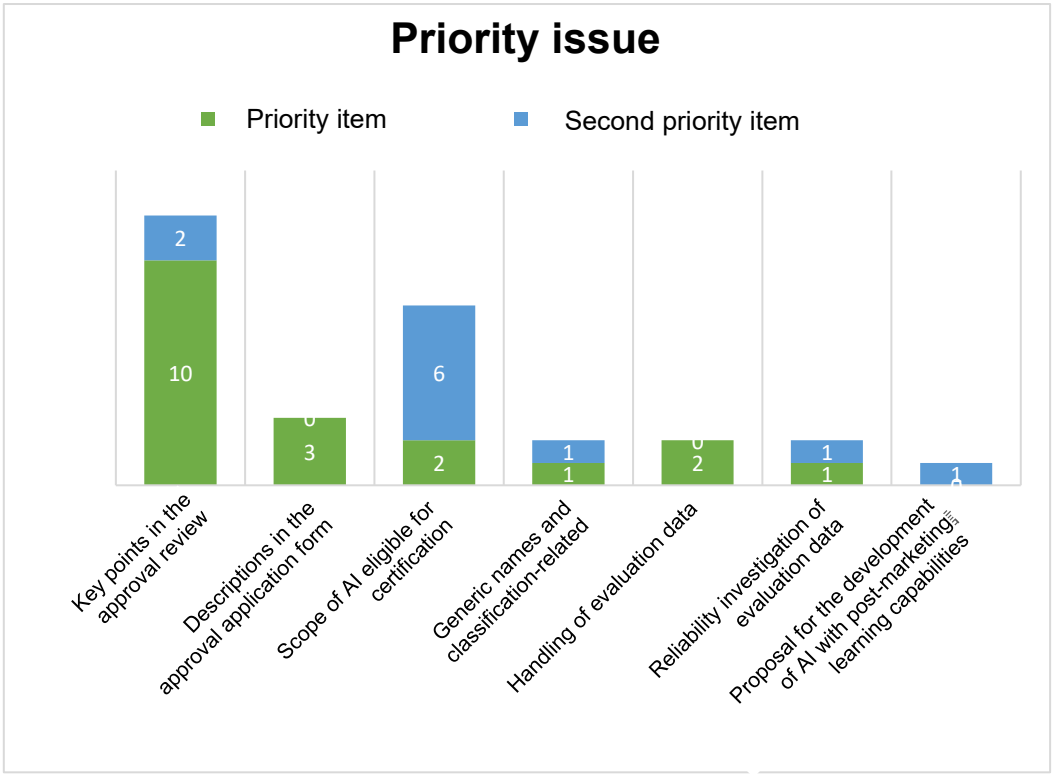

Fig. 1) Questionnaire survey on issues to be discussed at the WG (implement before the start of the WG)

(*Reference Link: https://www.pmda.go.jp/files/000251247.pdf)

2. Contents Examined and Results

2-1. Descriptions in the approval application form

Each group examined the descriptions in the approval application form as well as the entire approval application package and made specific proposals. For example, ideas were presented such as the source, volume, and facility of the learning data; how to create the correct labels; performance indicators; and the algorithm design basis. Some groups also proposed dividing the ideas into “items for which the description in STED is sufficient” and “items that should be specified in the application form.”

As a result of the discussion, the PMDA explained that to identify the AI itself, they have to request the description of learning data when the AI was developed and of the performance and characteristics resulting from verification tests in the application form based on the “concept of next-generation evaluation indices for AI-based diagnostic imaging support systems.” This is because the validity of performance related to the efficacy and safety shown in the attached

data was confirmed for the scope of the medical device including AI, as identified in the application form in the review.

It was decided to summarize the results of discussion at the WG on the description in the approval application form as a Q&A, which will be introduced as part of the FAQs* on PMDA's website for SaMD (already released in April 2025). The details are described below.

(*Reference link: https://www.pmda.go.jp/files/000264780.pdf)

Q: How should the principle of detection/diagnosis (algorithm) in diagnostic imaging support programs, etc., using machine learning be described in the application form?

A: In the approval application for a medical device, it is necessary to describe the information identifying the product for application in the application form and to prepare the evaluation, showing that the efficacy and safety are secured for the scope of the medical device, as identified in the application form as attached documents.

The "Notification concerning the Publication of the Guidance Materials concerning Application for Marketing Approval of Medical Device Software" (Administrative Notice dated March 31, 2016) states that applicants are required to understand the details of the design specifications for medical device software and to specify the types of input data and corresponding output data.

For descriptions in the approval application form for diagnostic imaging support programs, etc., using machine learning, refer to Attachment 4 Evaluation indices for AI-based diagnostic imaging support systems of the "Release of Evaluation Indices for Next-Generation Medical Devices" [PSEHB/MDED Notification No. 0523-2, dated May 23, 2019]. Because it is difficult to evaluate the performance of support systems subject to these evaluation indices based on only the principles (e.g., implemented detection/diagnostic algorithms), design specifications, etc., due to their characteristics (unlike usual medical devices), it is necessary to specify factors that affect their performance, scope of efficacy, limitations, etc.

If the processing process uses the principles of black boxes such as deep learning and it is difficult to show the detection/diagnostic algorithm at the time of approval application, it is necessary to show the detection/diagnostic network structure and program outline at the time of design and development. In addition, it is necessary to conduct AI training using appropriate learning data to have the performance required to achieve its purpose. Therefore, depending on the mechanism, specified performance, etc., of the support system to be evaluated, the contents of necessary items must be shown clearly with reference to the basic items shown in 6. (2) of these evaluation indices.

Regarding the descriptions in each column of the application form and attached documents for medical device software, refer to the related notifications, such as "8. Handling of Application for Marketing Approval" of "Handling of Medical Device Software" (PFSB/MDRMPE Notification No. 1121-33, PFSB/SD Notification No. 1121-1, and PFSB/CND Notification No. 1121-29, dated November 21, 2014) and "Examples of Marketing Authorization (Certification) Application Forms of Medical Device Software and Attached Data" (Administrative Notice dated February 10, 2015).

2-2. Post-marketing change control

The change procedure based on the existing notifications will be applied to post-marketing change control. However, with the expectation that there will be a scope that can be handled with a minor change notification, each group discussed what kind of action is required when a type of change was made based on the contents specified in the approval application form, and then the whole WG discussed it.

Various proposals were made by each group, including for classification and decision trees for the differentiation among partial changes, minor changes, and no procedure required. However, as a result of the discussion, it was concluded that no particular documentation is required at present, because a partial change approval application is made when the information specified in the approval application form is changed to ensure the efficacy and safety of AI described in the previous section.

The contents of the AMED RS study, “Study Contributing to Performance Evaluation During Post-Marketing Learning,” being conducted at the National Institute of Health Sciences were shared and discussed. Some of these discussions are shared below. Post-marketing learning was recognized as an important step for improving the performance of medical devices; however, as concerns about data consistency and bias have been raised, the importance of methodology for accurate evaluation was emphasized. Several groups discussed how to use learning data and standardization in the annotation process, showing interest in how to proceed with future actions. The research representative of AMED mentioned that despite the difficulty in completely preventing a decrease in generalization performance, it may be possible to maintain it by increasing the volume of learning data. Moreover, studies using data obtained in clinical practice are ongoing, showing the possibility that the performance may be assured by using these data for fine-tuning within medical institutions.

In addition, members participating in the WG commented that it is necessary to clarify how products with an automatic learning function are positioned legally after marketing, while touching on AI with post-marketing learning capabilities and on updates by each medical institution (including automatic learning). Currently, as post-marketing learning is led by manufacturers, it is difficult for medical institutions to conduct independent learning. Therefore, it is necessary to at least understand the concept of the quality management system and comply with it; specifically, it is necessary to organize the method of “Verification & Validation” and document each process at each medical institution. However, it is assumed that the number of institutions where it can be realized are limited in reality.

The PMDA shared that the WG meeting for next-generation evaluation indices for “AI-based diagnostic imaging support systems” discussed that if the performance of each device is likely to differ depending on the medical institution or the unit, it may not be approved as a single product in the first place. In addition, it was also mentioned that it may not be compatible with the current PMD Act.

2-3. Discussion on the concept of differentiation of approval/certification

An opinion was raised that although there are several certified products for medical devices utilizing AI, the judgment criteria for compliance with the certification standards have not been clarified. Therefore, based on information such as training for registered certification bodies published on the PMDA website* and responses* to inquiries from registered certification bodies* at the time of discussion, each group examined it and the whole WG discussed it.

As a result of the discussion, the following basic policy was shared: Compliance with the certification standards should be judged in the same manner as the existing medical devices, i.e., whether the function and performance are equivalent to those of existing products, regardless of the presence or absence of the function and performance using AI. Opinions were shared regarding requirements for compliance with certification standards, including on the (1) risk level of intended use and function/performance contributed by AI (no change of risk); (2) evaluation of substantial equivalence (from the viewpoint of input/output, comparison with products for which the company obtained approval/certification, and non-provision of new medical care such as diagnosis and treatment); and (3) limitation of the accessory and not main functions. A framework for reviewing the possibility of certification for an automated diagnostic function using AI depending on the degree of physician/user involvement was also proposed (e.g., CADe, brain segmentation, and puncture support).

(*Reference Link: Training for registered certification bodies, contents of consultation from registered certification bodies, and responses)

3. Points Requiring Further Discussion

3-1. To what extent is the concept of “product identification” necessary?

Discussion about what information on AI should be included in the approval application form as approval items and post-marketing change control led to the realization that the concept of specifying products in the approval application form may not be appropriate, especially from the viewpoint of “AI.” There are various types of medical devices; for medical devices that also require frequent improvements other than AI, it is assumed that product identification may not be appropriate. This chapter will be described particularly based on the following points, on the premise of SaMD utilizing AI.

○ Refining the description on performance:

The approval application form for medical devices includes descriptions such as “Performance and safety specification” based on the results of the performance evaluation test. Many commented that specifications, etc., related to performance in the approval application

form should be described with a certain scope, and performance improvement within the scope (e.g., quality improvement such as response to outliers and improvement of performance to enhance the generalizability of machine learning) should be made possible with procedures other than the partial change application (requiring no procedure or within the scope not requiring review, such as a minor change notification). This opinion is based on the idea that it is important for the marketing authorization holder to conduct design verification tests and validation tests according to the procedures specified in the in-house QMS to continuously provide easier-to-use medical devices; this way, they can ensure that the launched products continue to fall within the specified performance scope in their responsibilities.

The PMDA provided a supplemental explanation, wherein if the product is identified with the description of performance within a certain scope, it is necessary to show that efficacy and safety within the scope are secured at the time of approval application, and each product needs to be reviewed, because the concept varies depending on product characteristics and how it is specified in design specifications.

○ Description of “learning data”:

Since the “learning data,” which are AI input information, are one of the elements to define the product, a certain level of description is required in the approval application form. Therefore, at present, if additional learning takes place after approval, the regulatory procedure (application for partial change approval or notification under the Improvement Design within Approval for Timely Evaluation and Notice [commonly called IDATEN]) is necessary. Similar to the above, it may become easier to respond to customer needs by ensuring that additional learning for securing AI performance is within the scope not requiring a review by making descriptions of learning data flexible in order to ensure the performance is within a certain scope. In periodic additional learning, use of the IDATEN system described below is expected, but many commented that the review should be unnecessary if additional learning is aimed at ensuring a certain scope of performance.

The PMDA provided a supplemental explanation that it is necessary to show the efficacy and safety of the entire scope of application at the time of the approval application by identifying the product with a flexible description.

3-2. Examination of effective utilization of the Improvement Design within Approval for Timely Evaluation and Notice (commonly called IDATEN)

Each group examined advantages of the IDATEN system and areas for improvement, and the whole WG discussed them. When the WG checked internally, a total of four companies have used this system.

The advantages raised for this system were high predictability of launch timing (e.g., the time to product release can be shortened if the condition is met and the shipment date can be set in advance), the ability to make multiple changes in a step-by-step manner, and reduction of burden on applicants through no reliability investigation. In particular, it was mentioned that IDATEN is effective when similar changes are made multiple times and when the scope of improvement can be identified prior to the test.

Areas for improvement were related to complexity, wherein the confirmed “change plan” must be changed when a change is made. In addition, there was an opinion that the usual partial change approval application is faster because a prior face-to-face consultation is essential. In addition, several members commented that it is difficult to understand how its use differs from that of the two-step approval. Regarding this point, the PMDA commented that “The concept of two-step approval and IDATEN are different systems. The two-step approval involves obtaining first-step approval for the intended use, which is limited to the scope shown in the performance evaluation test, etc., as well as for a change to the original intended use by obtaining further evidence for the second-step approval. On the other hand, IDATEN is not a system assuming changes to the intended use based on clinical data, and other changes are assumed such as performance improvement and addition of the product lineup.”

When using IDATEN for SaMD utilizing AI, performance improvement through the addition of learning data is assumed. However, many members agreed that it may not be suitable to devices employing highly novel technologies such as AI. As for the reasons, some commented that it is easier to make the usual partial change application after the test, because the current IDATEN system requires a change plan assuming the state after the change from the beginning as well as submission of the draft description of the planned approval form. However, the development side wants to include various ideas for improvement to respond to the customer’s needs as much as possible until the deadline; it may also be unable to obtain test results as planned because not enough experience has been accumulated for the technology.

The PMDA stated that “In the IDATEN application, we recognize that details of the state after the change and complete draft description of the planned approval form are not necessary in all cases. Although it depends on the level of detail required, there may be acceptable cases if the concept of the description after the change can be shared. Therefore, the PMDA should be consulted for individual cases when using IDATEN.”

In addition, due to the few actual case examples, there was an opinion that it may be necessary to share case examples to improve understanding of which items and changes are suitable for this IDATEN system. The list of change plan confirmation products* has been shared on the PMDA website since May 2025.

(*Reference Link: List of change plan confirmation products, information on approval of software as a medical device, etc.)

3-3. Other issues related to devices utilizing AI

Although the WG could not examine them, other various issues were raised concerning medical devices utilizing AI. They are listed below from the viewpoint of expectations for future discussion.

1) Difficulty in collecting information on learning and evaluation data (from the viewpoint of cost and protection of personal information)

Multiple regulations must be considered in collecting learning and evaluation data, such as the Personal Information Protection Law and ethical guidelines; thus, the hurdles are high. To prevent bias in evaluations, evaluation data need to be isolated from developers. Therefore, a proposal was made to implement measures, for example, to allow companies to utilize data pooled at academic societies, along with existing information, for development.

2) Handling of AI (adaptive type) that automatically learns after marketing in clinical practice

For AI that automatically learns after launch in medical practice, it is necessary to organize the concept of division of responsibilities between medical institutions/doctors and marketing authorization holders as regards who is responsible for automatically learned contents. It is also necessary to examine the relationship between the Medical Practitioners Act and approval under the Pharmaceutical and Medical Device Act, ways to identify products, and methods for quality assurance/performance evaluation in medical practice. Further, it is necessary to clarify the relationship with the Personal Information Protection Act for the data at the time of automatic learning.

3) AI-related standards

From the viewpoint of international harmonization, it is necessary to refer to related standards; however, in the stage where various standards are under review, it is necessary to pay close attention to the trends in future investigations.

4) Handling of products using generative AI

Such products may be technically realized in the near future; therefore, a review of regulations will be necessary based on how they are used.

5) Concept of the scope requiring no partial change application/minor change notification/change procedure for each change

Regardless of the use or non-use of AI, the concept of the basic change procedure is the same as that for existing medical devices and is based on the degree of impact on efficacy and safety. Although a collection of specific case examples will be useful, it may take time to be able to compile them concretely because there are only a few results and cases at present.

6) Concept of products requiring approval review and products eligible for certification based on certification standards

As described in Section 2-3, the Certification Administration WG of JFMDA has continuously examined the handling of accessory functions, substantial equivalence, and automatic diagnostic functions discussed in this report.

7) CADe certification standardization

Approval review is required at present, but there was an opinion that CADe can be standardized if it can demonstrate equivalence to existing products in terms of the target disease and modality equivalence. Among products currently requiring approval review, a proposal was made regarding, for example, whether certification standard/approval standard can be made for lesion detecting AI (CADe) in radiological imaging and for the second/concurrent reader types.

For CADe, the PMDA compiled the points to consider in approval reviews in the “Review Points for Computer-Aided Diagnosis Program to Support Interpretation of Medical Images,” which is available on the PMDA website. The PMDA expressed its view that it is possible to establish certification and approval standards if it is possible to clarify the clinical positioning (it is used for whom, for what purpose, etc., in clinical settings) and specify the requirements common to the target disease and modality based on the validity, etc., of the evaluation package and evaluation test for multiple approved products.

4. Conclusion

Proactive participation of the people concerned such as PMDA personnel and comments from various viewpoints from industry participants clarified the concepts of the review of SaMD utilizing AI in the current review system. At the same time, the industry’s expectations for SaMD “utilizing AI,” which has advanced one step from the existing SaMD, as well as difficulties and concerns in handling them, were identified.

On the other hand, this WG is intended to promote transparency of reviews within the framework of the current review system, and it is considered necessary to review the expectations for programs utilizing AI in general and the ideal state of the system assuming its introduction to medical care beyond the framework of the current review system.

In the future, focusing on the above [points requiring further discussion] and [issues raised other than those related to review], it is necessary to consult with the persons concerned regarding how to review and proceed with the review.

End of document