The 9th Data Science Roundtable Discussion

Utilization of Digital Biomarker (dBM) in Clinical Development

Note: The content of this discussion is a summary of the participants' exchange of opinions and does not represent the official views of their respective organizations.

Summary of Discussion Topics in Part 1 & Part 2

- Part 1: Availabilities of dBM

  Discuss drug development strategies including the use of dBM, challenges specific to dBM and their solutions, regulatory issues and regulatory considerations, and the involvement of biostatisticians in the use of dBM.

  – Topic 1: Advantages and Challenges of Using dBM
  – Topic 2: Development Strategies for Using dBM as the Primary Endpoint of Pivotal Studies
  – Topic 3: The Roles of Biostatisticians

- Part 2: Development and Validation of dBM as Assessment Measures

  Discuss points to consider when designing a study and interpreting the study results to evaluate the reliability, validity, and responsiveness and to set a clinical meaningful threshold of change, in contrast to PRO development, including statistical issues for each use of dBM.

  – Topic 1: Evaluation Methods and Criteria for Each Psychometric Property
  – Topic 2: Considerations and Challenges in Planning and Analyzing Studies Using dBM
  – Topic 3: Setting Clinically Meaningful Differences

Hypothetical Cases

Participants discussed based on these 3 topics.

[Hypothetical case 1]

- When the number of coughs per 24 hours is measured using a digital cough recording device (e.g., VitaloJAK) in patients with chronic cough, and the average number of coughs per hour is set as the primary endpoint of the pivotal study

[Hypothetical case 2]

- In patients with atopic dermatitis, when the scratching behavior during night sleep is measured using a wristwatch-type device (e.g., actigraphy), and the time of scratching per night is used as the primary endpoint of the pivotal study

[Hypothetical case 3]

- In the case of objectively and continuously collecting multiple physiological parameters such as skin potential, heart rate, and their variation using a wearable device, and applying algorithmic processing to these data in patients with endometriosis, the Pain Index, which is the dBM indicator corresponding to pain, is set as the primary endpoint of the pivotal study

Results of the Discussion in Part 1

The topics focused on by each group are different.

[Topic 1: Advantages and Challenges of Using dBM]

- Benefits, challenges, and countermeasures of using dBM in the disease area we are in charge of

[Advantages]

- Symptoms that were previously difficult to evaluate can be evaluated.
  - Example: The number of daily coughs, breathing during sleep, scratching time per night, etc.
  - Improving objectivity and biological relevance.
- It can be evaluated regularly and continuously.
- Easier to detect effects at an early stage.
- Creation of novel measurement concepts.
-

[Challenges]

- Missing data and compliance issues (intercurrent event)
  - Example: Not sure if the patient is sleeping or not.
- The algorithm for calculating the measurement can be unclear and black-boxed. It makes difficult to explain the validity of dBM.
- In drug development strategies, the issue is how and when to validate dBM.
  - The early phase is considered appropriate.

[Topic 1: Advantages and Challenges of Using dBM]

- Benefits, challenges, and countermeasures of using dBM in the disease area we are in charge of

[Challenges (continue)]

- The burden on patients increases.
  - Location restrictions
  - Behavioral change due to device placement
  - Simple operation is needed
- Selection bias may occur because only patients who are accepted to wear the device will be included.
- Difficulty in applying to external controlled studies
  - If the wearing of the device may change behavior, comparisons within the same study may be required.

[Solutions]

- Negotiate to receive information from dBM developers to ensure algorithm transparency.
- Implement measures to prevent missed measurements (e.g., device alerting if missing).
- Consider the mechanism of missing.
- Reduce the patient’s burden by minimizing the need for the participant's wearing time.
- Pilot in small groups to determine if long-term loading is possible

[Example]

- In development at Dischenne muscular dystrophy, there are examples where dBM has been adopted as a primary endpoint as an alternative to 6-minute walking distance.

[Topic 2: Development Strategies for Using dBM as the Primary Endpoint of Pivotal Studies]

- What dBM indicators can be the primary endpoint?
- What is necessary to use the dBM index as the primary endpoint?

[What is necessary to use the dBM index as the primary endpoint?]

- The requirements related to psychometric property necessary for using dBM differ for the types of endpoint.
  - Use as an exploratory endpoint.
    - We don't think that requirements are as strict as secondary endpoints.
    - First, set it as an exploratory endpoint in the early phase and collect data to confirm the property of dBM.
  - Use as a secondary endpoint.
    - As with the primary endpoint, although there are differences in the degree of requirements, it is necessary to confirm all the main properties of feasibility, responsiveness, content validity, criteria-related validity, and construct validity.
- The device and data processing methods should be adequately validated.
  - Assess whether there is a correlation with QOL.
- The extent of clinically significant improvement should be adequately assessed.
  - Example: Clarification of criteria such as how much a decrease in the number of coughs can be judged as "improvement“.
- Ensuring objectivity
- Confirmation of correlation with existing indices (convergent validity, discriminant validity)

[Topic 2: Development Strategies for Using dBM as the Primary Endpoint of Pivotal Studies]

- What dBM indicators can be the primary endpoint?
- What is necessary to use the dBM index as the primary endpoint?

[Activities required for promoting the utilization]

- Hold a consortium to promote information sharing through industry-government-academia collaboration.
  - Propose a new device.
- Cost reduction
  - However, at this point in time, whether or not the use of dBM is recommended is controversial.
- Widely communicate the benefits of biomarkers.
  - Improve the accuracy of evaluation.
  - Enables reduction in the number of participants and the study period, etc.
- Consider using it as part of co-primary endpoints.
- Early dialogue with regulators
  - When deciding to use the dBM indicator as the primary endpoint, it is advisable to consult with the regulator as soon as possible.
- Proactively managing large volumes of data
  - Example: There is a case where about 300,000 rows of data were generated per person in one day.
- Industry-government-academia cooperation
  - In the development of PRO, there is a track record of building a foundation led by academia.
  - In the case of dBM, tech companies may be the main focus, but it is difficult to complete clinical validation alone, so collaboration between industry, government, and academia is essential.

[Topic 3: The Roles of Biostatisticians]

- From what stage should biostatisticians be involved?
- What issues should biostatisticians be particularly involved in?

[From what stage should biostatisticians be involved?]

- Initial stage of considering the use of equipment

[What issues should biostatisticians be particularly involved in?]

- Handling of missing data
- Impact assessment on patient population
- Evaluation of sample size and power
- Providing input from the investigator's perspective
  - Check if the algorithm is not black-boxed.
  - (e.g. for cough monitors) Are there any doubts about the compression algorithm, etc.?
- Validation from the mathematical aspect
  - Consideration for setting cut-off values
  - If necessary, consider consulting with vendors.

Results of the Discussion in Part 2

The topics focused on by each group are different.

[Topic 1: Evaluation Methods and Criteria for Each Psychometric Property]

- What is the appropriate study design to evaluate each property?

- **Reliability**: Measurements are consistently obtained under the same conditions.
  - It is an important metric to evaluate in the early stages of development.
  - In the case of atopic dermatitis, evaluation in a phase 2 study is considered appropriate.
  - If a gold standard (e.g., true value based on video observation) exists, the reliability of dBM can be assessed by comparing it with that standard.
  - If no gold standard exists, the test-retest method (e.g., comparing measurements taken 1 and 2 weeks apart) is valid.
  - In a phase 2 study, evaluation during the screening period (i.e., treatment-free state) prior to treatment can be an option.

- **Responsiveness**: The ability to accurately detect changes when a patient's condition changes
  - If a gold standard exists, check if the dBM value changes in accordance with it.
  - Assess responsiveness by comparing treatment effects across multiple indicators, including dBM.
  - This can be evaluated in a phase 2 study.

- **Validity**: The extent to which the measure actually assesses what it is intended to measure
  - If it is difficult to evaluate using gold standards such as video observation, consider evaluating construct validity in addition to criterion-related validity.

- **Clinically meaningful differences**
  - Clinically meaningful differences should be clearly defined and evaluated when using dBM as a primary endpoint.

[Topic 1: Evaluation Methods and Criteria for Each Psychometric Property]

- What is the appropriate study design to evaluate each property?

- The use of dBMs for drug development may not necessarily require a special study design.

- How will the endpoints be defined?
  - Comparison of the between-group differences in the mean number of times
    - Interpretation may be difficult because individual patients generally have a different standard number of coughing.
  - Comparison of between-group differences in percentage of individual reduction (%) from baseline
    - It is possible to conduct evaluation considering individual differences.
  - Evaluation by Time Zone
    - Changes in symptoms may vary depending on the time of day, such as a decrease in coughing during the day or a decrease in coughing while sleeping that improves sleep.

    →It can also be worth considering evaluating the treatment effect by the time period such as "daytime" and "sleeping" rather than the number of coughs throughout the day.

- A study to evaluate the validity, etc. needs to be conducted in a population in which changes can be observed.
  - If almost all subjects have a condition where the number of coughs is zero, evaluation such as evaluation of responsiveness will be difficult, and therefore it is necessary to select an appropriate target population.

[Topic 1: Evaluation Methods and Criteria for Each Psychometric Property]

- What is the appropriate study design to evaluate each property?

pEP=primary endpoint、sEP=secondary endpoint

[What is the appropriate study design for dBM development (for use of dBM in pivotal studies of drugs)?]

- Studies to assess verification (accuracy), usability validation (operability), analytical validation (reliability, responsiveness), and clinical validation (content validity, criteria-related validity, and construct validity) are necessary.
- In addition to comparison with conventional indicators (e.g., pain VAS) at a specific time point, dBM can be evaluated as time series data.

[What is the appropriate design in a pivital sutdy for drug development using the dBM of Pain Index?]

- Take about 2 weeks as a baseline period (taking advantage of dBM's strength in being able to evaluate continuously).
- The use of the Pain Index as an allocation factor at randomization may be more accurate in avoiding bias than using pain VAS.
- The study duration and evaluation timepoints/timeframes should take into account the period during which the drug effect, as measured by dBM, is expected to manifest.
- What is clinically important is not pain at one time point, but pain x duration (AUC).
  - Patient voices and KOLs opinions should be considered for each disease.
- It may be difficult to immediately adopt it as a pEP. Instead, how about considering a gradual replacement over a much longer timeline than a single drug development project?
  - The pEP is the pain VAS. To minimize recall bias, patients record their pain VAS while referring to the dBM, such as Pain Index.
  - The pEP is defined as the pain VAS adjusted based on the dBM score—for example, if the dBM confirms that the pain exceeds the patient's previously reported "worst pain ever" on the VAS, the upper limit of the VAS scale may be corrected accordingly.
  - As experience accumulates with trials using "pEP: pain VAS, sEP: dBM," a consensus may emerge that dBM metrics (e.g., AUC of the Pain Index) are more appropriate. Based on that, the pEP could eventually be replaced by the dBM.

[Topic 2: Considerations and Challenges in Planning and Analyzing Studies Using dBM]

- What concerns and strategies should be considered when planning and analyzing clinical trials using dBM?

[Concerns]

- Occurrence of missing data
  - Narrowing down to only evaluable subjects at the time of analysis may lead to bias in the selection of target populations, raising concerns about its validity.
  - If the device is not installed, it is useful to send out an alert.

- Frequency of data transfer errors and equipment failures
  - Before starting any pivotal clinical trial, it is desirable to try it in additional trials or preliminary operations to understand the frequency of transfer errors and failures.

- Addressing device safety concerns
  - In case of safety issues, the response policy should be clearly stated in the protocol in advance.
    - Redness and itching caused by wearing a device in the case of atopic dermatitis

- Algorithm Disclosure
  - If a gold standard exists, there is an opinion that if the results obtained by the algorithm are consistent with that gold standard, that is sufficient and the disclosure of the algorithm is not necessary. On the other hand, there are opinions that it is not enough, and even if the results are consistent, it is necessary to check the content of the algorithm.
  - If it may be difficult to disclose the algorithm, it is necessary to consider this point from the device selection stage.

[Topic 2: Considerations and Challenges in Planning and Analyzing Studies Using dBM]

- What concerns and strategies should be considered when planning and analyzing clinical trials using dBM?

- The cough monitor counts the number of coughs after removing the cough-free period.
  - It is necessary to consider whether only the data after removal should be saved or whether the data for the entire measurement time should be stored.
  - Although raw data is enormous, it may contain important information, so it is basically desirable to store raw data.

- What if a non-compliance occurs during a study?
  - Since the risk is high, it is desirable to confirm the feasibility of implementation through pilot studies in advance.

- One of the main features of dBM is its ability to collect data continuously.
  - It is necessary to consider analysis methods that can appropriately handle measurement data at multiple points in time.

- It is possible that doctors may want to introduce the device for research purposes without ensuring that the measurement is appropriate.
  - Rather than paying the door because "validity is not guaranteed", it is important to establish a system that allows us to discuss together what kind of preparations are needed before the study starts.

[Topic 2: Considerations and Challenges in Planning and Analyzing Studies Using dBM]

- What concerns and strategies should be considered when planning and analyzing clinical trials using dBM?

- Missing data
  - There is a concern about missing data due to device disconnection → By considering the relationship between the wearing rate and the data acquisition rate in advance, etc., it is possible to define eligibility criteria (threshold) for case selection before the start of the study.
  - Handling of missing data → It is necessary to consider based on Estimand.

- Operation
  - Wearable device measurements can be used to make remote monitoring tools.

- Devices and algorithms
  - Concerns about the use of adopted devices for model changes, and risk of accuracy deterioration over time → Procurement of required quantities in advance and quality control
  - Risk that software updates in the device may alter the behavior of the raw data processing (algorithms) obtained from the sensor
    → Plan in advance, such as freezing updates or comparing behavior before and after updates.

- Generalizability
  - There is a concern about bias toward the subgroup of patients with successfully collected data—namely, those with high digital literacy and consistent device usage.
    → By comparing the background characteristics of patients with and without successful data collection, we aim to explore, by process of elimination, the limitations in the generalizability of the study results.

[Topic 3: Setting Clinically Meaningful Differences]

- What is the appropriate study design to set "clinically meaningful differences" and/or "clinically meaningful change thresholds"?

- Required data
  - Data that measures anchor indicators such as PRO and dBM at the same time are required.
  - It does not necessarily have to be an intervention trial, and observational studies may be sufficient.

- Clinically meaningful differences
  - It is useful to conduct a questionnaire for patients, such as "How many fewer coughs do you feel better?“.
  - Data on changes over time in the control treatment (i.e., standard of care) is obtained and evaluated by adding the expected superiority of the study treatment.

- If the new dBM indicator can be the gold standard,
  - For example, if there was previously no means to quantitatively measure the number of coughs, the new dBM indicator could serve as a gold standard in its own right.
  - In such cases, just checking whether the output is close to the "true value of the number of coughs" can be sufficient, just like the evaluation of diagnostics.