

デジタルバイオマーカーを臨床試験の主要評価項目として利用するために必要なバリデーションの評価について

令和 7 年 6 月

日本製薬工業協会 医薬品評価委員会

データサイエンス部会

2024 年度 継続タスクフォース 10

1	はじめに	5
2	dBm のバリデーションにおいて考慮すべき事項	6
2.1	デジタル臨床評価の関連性 (Relevance of Digital Clinical Measure)	6
2.2	ベリフィケーション (Verification)	8
2.3	分析バリデーション (Analytics Validation)	9
2.3.1	適切な参照標準の選択と利用	9
2.3.2	分析アルゴリズムが健康の概念や健康の特徴を正しく測定, 検出, 予測できるか	9
2.3.3	変化を検出する能力	10
2.3.4	アルゴリズムによるバイアスの説明	11
2.4	臨床バリデーション (Clinical Validation)	11
2.4.1	センサー由来のエンドポイントが意図した意味のある健康の特徴を測定できること	12
2.4.2	測定における臨床的に意味のある変化	12
2.4.3	健康の特徴を正しく推定するために必要な最小限のデータ量, 欠損データの許容範囲	13
3	dBm のバリデーション事例	13
3.1	背景情報	14
3.1.1	DMD の患者層	14
3.1.2	ペインポイント	14
3.1.3	ActiMyo はどのような機器か[18]	15
3.1.4	ActiMyo のベリフィケーション[19]	15
3.1.5	DMD における SV95C[19]	15
3.2	主要評価項目として利用する際に必要な追加のデータ	16
3.3	定性的エビデンス (内容的妥当性)	17
3.3.1	患者の健康にとって重要なこと (MAH) の調査	17
3.3.2	興味のある評価対象 (COI) の調査	17
3.3.3	測定対象のアウトカムの選定	18
3.3.4	デジタルバイオマーカーを用いた自宅測定の意義の確認	18
3.4	定量的エビデンス	19
3.4.1	正確性 : Accuracy	19
3.4.1.1	方法	19
3.4.1.2	結果	19
3.4.2	再現性 : Repeatability	19
3.4.2.1	再テスト信頼性 : Test-retest reliability	19
3.4.2.1.1	方法	19
3.4.2.1.2	結果	20
3.4.2.2	頑健性 : Robustness	20

3.4.2.2.1	方法	20
3.4.2.2.2	結果	20
3.4.3	構成概念妥当性 : Construct validity	20
3.4.3.1	既知グループ妥当性 : Known-group validity	20
3.4.3.1.1	方法	20
3.4.3.1.2	結果	21
3.4.3.2	収束的妥当性 : Convergent validity	23
3.4.3.2.1	方法	23
3.4.3.2.2	結果	23
3.4.4	反応性 : Responsiveness	24
3.4.4.1	疾患の自然経過における変化を検出できる能力 : Ability to detect change	24
3.4.4.1.1	方法	24
3.4.4.1.2	結果	25
3.4.4.2	状態を改善させる治療による変化を検出できる能力	26
3.4.4.2.1	方法	26
3.4.4.2.2	結果	26
3.4.5	意味のある変化の閾値 : meaningful change threshold	27
3.4.5.1	分布に基づく方法	28
3.4.5.1.1	方法	28
3.4.5.1.2	結果	29
3.4.5.1.2.1	SV95c	29
3.4.5.1.2.2	6MWD, NSAA 及び 4SC	30
3.4.5.2	アンカーに基づく方法 (患者内での閾値の設定)	30
3.4.5.2.1	方法	30
3.4.5.2.1.1	PODCI 及び CGI-C	30
3.4.5.2.1.2	6MWD 及び NSAA	31
3.4.5.2.2	結果	31
3.4.5.2.2.1	PODCI 及び CGI-C	31
3.4.5.2.2.2	6MWD 及び NSAA	32
3.4.5.2.3	SV95C の MCT の設定のまとめ	33
4	おわりに	33
5	参考文献	34

略語・用語

略語	略していない表現（英語）	略していない表現（日本語）
4SC	4-Stair Climb test	4 段階昇降テスト
6MWD	6 Minute Walking Distance	6 分間の歩行距離
6MWT	6 Minute Walking Test	6 分間歩行試験
BioMeTs	Biometric Monitoring Technologies	バイオメトリックモニタリングテクノロジー
CHMP	Committee for Medicinal Products for Human Use	欧州医薬品委員会
CGI-C	Clinical Global Impression of Change	－
COA	Clinical Outcome Assessments	臨床アウトカム評価
COI	Concept of Interest	（MAH のうち）興味のある評価対象
CTTI	Clinical Trials Transformation Initiative	－
DHT	Digital Health Technology	デジタルヘルス技術
dBm	Digital Biomarker	デジタルバイオマーカー
DiMe	Digital Medicine Society	－
EMA	European Medicines Agency	欧州医薬品庁
FDA	Food and Drug Administration	米国食品医薬品局
ICC	Intraclass Correlation	級内相関係数
MAH	Meaningful Aspect of Health	患者の健康にとって重要なこと
MCT	Meaningful Change Threshold	意味のある変化の閾値
MDC	Minimal Detectable Change	最小の検出可能な変化
NMD	NeuroMuscular Disorders	神経筋疾患
NSAA	North Star Ambulatory Assessment	ノース・スター歩行能力評価
PFDD	Patient Focused Drug Development	患者を中心とした医薬品開発
PODCI	Pediatrics Outcomes Data Collection Instrument	－
PPMD	Parent Project Muscular Dystrophy	－
QOL	Quality of Life	生活の質
SD	Standard Deviation	標準偏差
SEM	Standard Error of Measurement	標準測定誤差
SV95C	Stride Velocity 95th Centile	－

1 はじめに

本報告書におけるデジタルバイオマーカー (Digital Biomarker : dBM) は、「臨床試験で用いられる各種デジタルヘルス技術 (Digital Health Technology : DHT) から得られる客観的・定量的に収集・測定した生体データ (生理学的データや行動のデータ)」とする。本タスクフォースでは dBM を臨床開発の主要評価項目として活用するためにはどのようなエビデンスが必要になるか事例検討を行った。

dBM が、臨床開発にて価値をもたらすためには、その測定が意図された用途に適合し、対象となる治験参加者が負担なく使用できるといった条件を理解することが重要である。例えば、パーキンソン病の総合的な評価尺度である Unified Parkinson's Disease Rating Scale (UPDRS) の代わりに開発されていたアプリは、疾患全体の症状を説明するものでないとした事例がある[1]。一方で、ウェアラブルデバイスを用いて収集したデータから構築した評価項目である Stride Velocity 95th Centile (SV95C) は、EMA から主要評価項目として利用することが承認されている。dBM から構築した評価項目が意図した臨床試験の目的に適合していることを保証するためには、臨床概念と dBM の適切な選定及び頑健でフェーズに適したバリデーションが必要になる。

日本製薬工業協会 医薬品評価委員会 臨床評価部会の 2021 年、2022 年度タスクフォースでは、「実際に製薬企業が dBM を利用した臨床試験を実施するには何が必要か」という視点で、dBM 活用の意義を再確認し、どのような準備が必要で、様々な課題にどのように対処すべきか、先行事例や各種ガイドラインを中心に調査し、まとめている[1]。2022 年度タスクフォース報告書では、医薬品開発における dBM の活用意義と実務的手順について紹介されている。

本報告書は、統計担当者、データマネジメント担当者が主に活動するデータサイエンス部会としての特色を活かし、ベリフィケーション、バリデーションといったエビデンスの創出にフォーカスし、SV95C が主要評価項目として認められるまでに実際に行われた試験の目的、実施試験の内容を解説することにより、実務担当者が開発計画を立てる際に有用な情報を提供することを目的に作成した。2022 年度の報告書[1]も事前に目を通していただくことで、本報告書の理解が進むものと考えます。

2 章では、デジタル医療分野の専門家を中心に 2019 年に米国で設立された非営利団体である Digital Medicine Society (DiMe) のウェブサイトで公開されているチェックリスト[2]に基づき、dBM を用いたピボタル試験の主要評価項目として開発するために必要なエビデンスについて解説する。なお、ここで紹介している情報のほか、Clinical Trials Transformation Initiative (CTTI) *1が示している、新たなエンドポイント開発のフローチャート[3]も参考にできる。3 章では、EMA から公開されている SV95C に関する意見書に基づき、主要評価項目として規制当局に認められるバリデーション評価について具体例を紹介する。DiMe のチェックリストとは異なるまとめ方をしているため、本 TF で考えた対応関係を 2 章に記載した。

本資料が、臨床試験の企画立案を行う開発担当者、統計担当者、ウェアラブルデバイスの開発者、またデータ加工するためのアルゴリズム開発担当者といった方々が、dBM を開発す

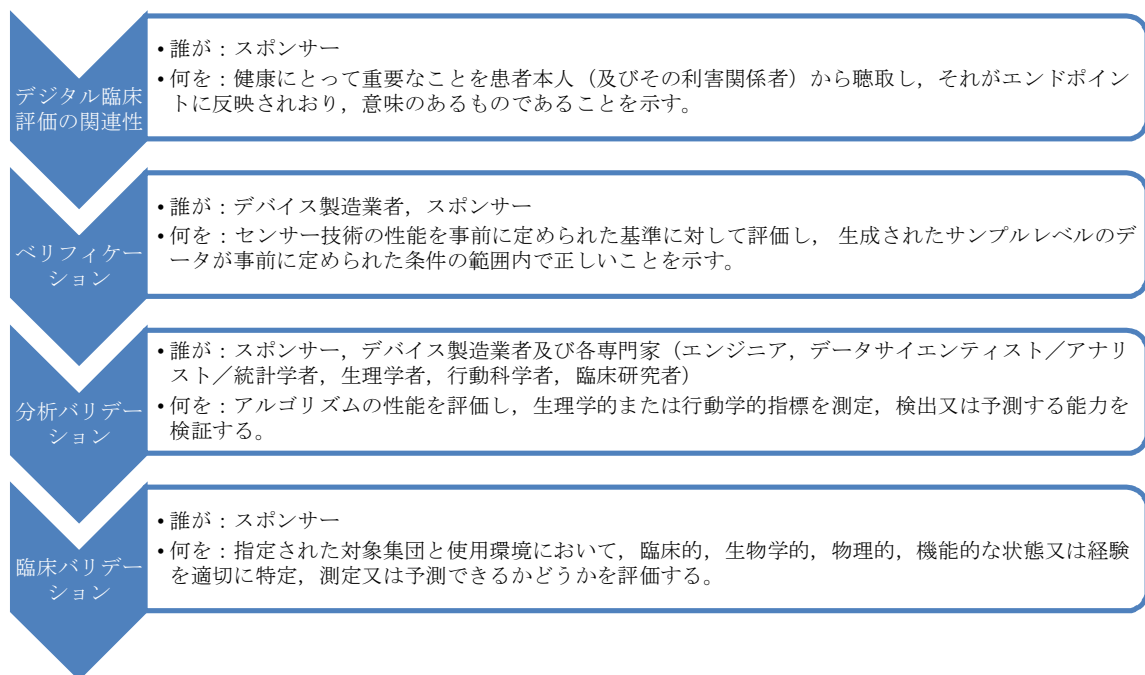
*1CTTI:2007 年にデューク大学と FDA によって設立された官民パートナーシップ組織

る際の一助となれば幸いである。

2 dBM のバリデーションにおいて考慮すべき事項

デジタル臨床評価をピボタル試験の評価項目として利用するためには、その主要評価項目が試験の目的に合致していること、臨床的に適切であること、DHT によってデータが適切に取得されること、技術が目的に適合していることを事前に確認し、スポンサーはバリデーションを実施する必要がある。ただし、臨床試験のニーズや試験の背景は様々であるため、どのようなバリデーションが必要かを一概に決めることはできない。

ここでは、DiMe の The Playbook Measurement ウェブサイトで公開されているチェックリスト[4]をもとに、ピボタル試験の主要評価項目に dBM を利用する際に考慮すべき事項を①デジタル臨床評価の関連性 (Relevance of Digital Clinical Measure)、②ベリフィケーション (Verification)、③分析バリデーション (Analytical Validation)、④臨床バリデーション (Clinical Validation) の 4 段階に分類し、2.1 から 2.4 で紹介する (図 1 参照)。DiMe の The Playbook では、本チェックリストを利用し、スポンサーは、context of use (どのような状況で、どのような目的で利用するか) に応じてデジタル臨床評価の使用を裏付けるこれらの証拠を収集し文書化することを推奨している。



[5] [6]より一部改変して引用

図 1 ピボタル試験の主要評価項目に dBM を利用する際に考慮すべき事項

2.1 デジタル臨床評価の関連性 (Relevance of Digital Clinical Measure)

dBM の選択と開発をするには、まずエンドポイントが疾患や障害を有する患者にとって重

要な変化を示すことを明確に説明する必要がある[4][7]。The Playbook では、デジタル臨床評価の関連性プロセスで以下をチェック項目に挙げている。

- ✓ 臨床的な意義・関連性 (Clinical meaningfulness and/or relevance of measure)

論文[4][7]では dBM の選択と開発にあたり 4 段階の枠組みを推奨している。デジタル臨床評価が患者にとって重要なことと関連性があることを示すには、この 4 段階の枠組みを特定していくことが有効であろうと記載されている。

- ✓ 患者の健康にとって重要なこと (MAH; Meaningful Aspect of Health)
- ✓ MAH のうち興味のある評価対象 (COI; Concept of Interest)
- ✓ 測定されるアウトカム
- ✓ エンドポイント

日本製薬工業協会 医薬品評価委員会 臨床評価部会の成果物[8]に枠組みと歩行機能に関する例が記載されているので具体的な事例として再掲する。

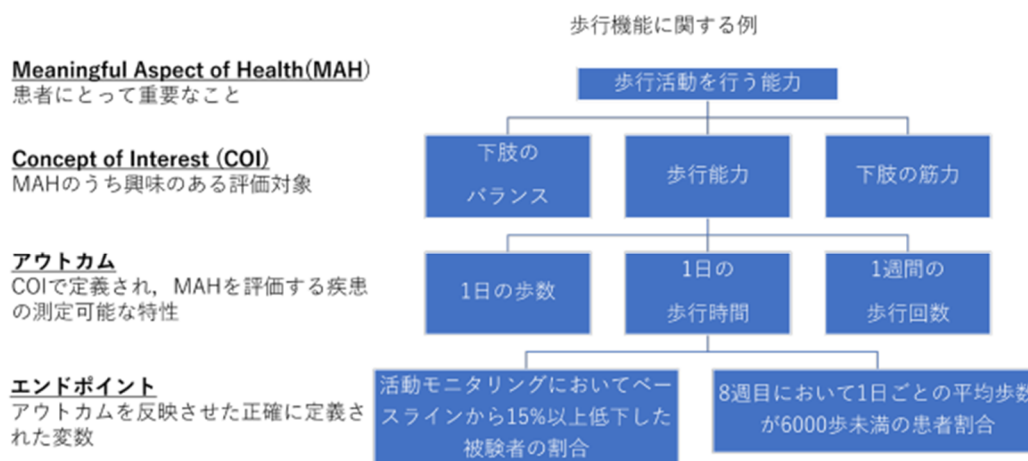


図 2-2 治療的介入の意義からエンドポイント設定までの 4 レベルの枠組みと歩行機能に関する例

FDA ガイダンスでは、DHT によって得られたデータからエンドポイントを開発することと、他の手段で収集されたエンドポイントを開発することは同じ原則を適用できるとしている[9]。さらに、FDA は、医薬品開発や規制上の意思決定において患者経験データ及びその他の関連情報をどのように収集し提出するのかに対して複数のガイダンスを提供している。これら、患者中心の医薬品開発 (PFDD) ガイダンス文書[10][11][12]は、dBM を使用して MAH や COI を評価するためのエビデンスを提供するのに役立つであろう。CTTI は、DHT が関心のあるアウトカムを捉えるための最適な手段であるかを判断するにあたり、患者、介護者、医療従事者、医療保険者等を含む多様なステークホルダーと協働・対話することを推奨している。また、その際に活用できる質問例も提供している[3]。3 章以降に dBM のバリデーション事例として SV95C をとりあげているが、この事例では、患者やその他関係者へのアンケート調査の内容が示されている。(3.3 定性的エビデンス (内容的妥当性))

次章以降で、dBM の評価の詳細なステップを紹介する。

2.2 ベリフィケーション (Verification)

DHT のデータ取得技術の適切性を確認するために、製造業者によるベリフィケーションが必要になるが、The Playbook では、ベリフィケーションプロセスで以下をチェック項目に挙げている。

- ✓ 事前規定した基準に対するセンサー技術の性能試験に関すること (Performance testing of the sensor technology against pre-specified criteria)
- ✓ 事前規定した条件の範囲内で測定データが正しいこと (Sample level data is correct within the limits of the pre-specified conditions)

文献[5]では DHT のベリフィケーションプロセスで何を評価すべきか、何をエビデンスとして提供すべきか以下のように提案している。

このプロセスでは DHT センサーが一連の設計仕様を満たしていることを示し、(A) センサーがアナログデータを適切に取得していること、(B) 取得したデータを変更するファームウェアが適切な出力データを生成していることを保証する。

DHT は、加速度、電圧、静電容量、光などをサンプリングするセンサーを含むが、そのセンサー技術が、グラウンドトゥールース参照標準 (実際の直接測定した値) と比較した場合に、時間の経過とともに一貫して (センサー内比較)、複数のセンサー間で均一に (センサー間比較)、許容される精度と正確性をもってデータを取得していることを示す。

表 1 ベリフィケーションで提供すべきエビデンス

ベリフィケーション することで解決した こと	この DHT は、受け入れ可能な品質のサンプルレベルのデータを生成するのに十分な性能を有しているか
期待される文書	DHT に関するエビデンス。十分な技術的性能特性 (様々なテクノロジーや環境条件での正確性、信頼性、精度、時間経過に伴う一貫性及び均一性) のデータ 統合ハードウェアの性能仕様 出力データの仕様 ソフトウェアシステムテストの概要 ベリフィケーションテストの制限事項 例: ベリフィケーション中にテストされなかった特定の既知アイテム

なお、文献[5]で留意すべきこととして、性能要件が使用ケースによって異なる可能性があることがあげられている。たとえば、心拍数の精度は、高リスク患者で心房細動を検出するために重要だが、健康な若いアスリートの長期的な安静時心拍数モニタリングではそれほど重要ではない。したがって、ベリフィケーションプロセスには、適切な基準を特定するために意図される使用条件 (context of use: どのような状況で、どのような目的で使用されるかの明確な定義や説明) を考慮する必要がある。

ベリフィケーションのエビデンスは DHT 製造元が提供すべきと文献[5]にも記載があるが非公開となることがある。エビデンスが開示されない場合、その旨を明示し、スポンサーは分析バリデーションに焦点をあてることになる。

なお、SV95C の事例では、用いられたウェアラブルデバイスである ActiMyo®が各種基準に適合していることが示されている (3.1.4 ActiMyo のベリフィケーション)。

2.3 分析バリデーション (Analytics Validation)

The Playbook では、分析バリデーションのプロセスで以下をチェック項目に挙げている。

- ✓ 適切な参照標準の選択と利用 (Selection and use of an appropriate reference standard)
- ✓ 分析アルゴリズムが健康の概念や健康の特徴を正しく測定、検出、予測できるか (The algorithm is correctly measuring, detecting, or predicting a health concept or aspect of health)
- ✓ 変化を検出する能力 (Ability to detect change)
- ✓ アルゴリズムによるバイアスの説明 (Description of algorithmic bias)

分析バリデーションには、生理学的及び行動学的測定基準を生成するためのバイオメトリックモニタリングテクノロジー (Biometric Monitoring Technologies : BioMeTs) の評価が含まれる。これには処理されたデータの評価が含まれ、被験者を使ったテストが必要である[13]。

2.3.1 適切な参照標準の選択と利用

分析バリデーションの過程では、アルゴリズムによって生成された測定値は適切な参照標準に対して評価されなければならない。参照標準とは、アルゴリズムによって生成された測定値がどういった意味をもつか又は正確であるかなどを判断するために用いる他の測定方法であり、文献調査や臨床的コンセンサスをもとに、選択した参照標準がゴールドスタンダードであることを説明する必要がある[4]。例えば、入眠／覚醒の評価ではポリソムノグラフィが参照標準となるため、新たな dBM の開発ではポリソムノグラフィに対して分析バリデーションを示すことになる。他にも、酸素飽和度は動脈血検体、心拍変動は心電計、歩行力学などのバイオメカニクスはモーションキャプチャシステムを参照標準とすることが考えられる。1 つの測定基準に対して複数の参照標準が存在する可能性があり、すべての参照標準がセンサーに基づいているわけではない。人間の観察者が手動でカウントする方がより実用的である場合は、手動によるカウントを参照標準とする。しかし、手動による測定は一般的に最もユーザーエラーを招きやすいため、参照標準として最良の選択というわけではない。

なお、以下に例示するとおり、すべての参照標準が同じ品質であるとは限らないため、参照標準の分析バリデーションについて単一の精度の閾値設定は推奨しない。

1. すべての参照標準が完全に客観的であるとは限らない。
2. 光学式モーションキャプチャシステムのような表向き客観的な参照標準でも、操作者によるバイアスが最終的な測定値のばらつきを増加させることがある。
3. ゴールドスタンダードとなる参照標準が明確に定義されていない場合もある。

最も重要なことは、選択した参照標準がどのような測定を行い、どのように測定値が出力されているかを理解することである。

低品質な前時代の測定法を参照標準に設定すると、誤差をもたらす可能性がある。専門組織などによって文書化された、厳密で定量的な参照標準を選択すべきである[5]。

SV95C の事例では、理学療法士が測定した 6MWD を参照標準として評価されている。(3.4.1 正確性 : Accuracy)

2.3.2 分析アルゴリズムが健康の概念や健康の特徴を正しく測定、検出、予測できるか

The Playbook には、GCP の要求事項に従って実施された分析バリデーション試験のみが受

入れ可能であり、以下に示すように、そのシステムのすべてのアルゴリズム出力のエビデンスを文書に記載することと述べられている[4]。

1. 出力される測定値の説明
2. 測定値の算出方法の概要（可能であれば具体的な詳細を含める）
3. 測定値の検証のためにどの参照標準が選択されたか
4. 算出された測定値と参照標準との直接比較の結果（統計解析方法を含む）
5. 前述の直接比較試験で使用された被験者集団、実験条件、プロトコルの記述
6. 分析バリデーションが、管理された実験室環境で実施されたか又は使用される場所／方法に類似した実環境で実施されたかについての詳細な考察
7. 独立したサンプルで検証されていること

分析バリデーションは、アルゴリズムの性能や表現型又は健康状態の有無を測定、検出又は予測するアルゴリズムの能力に焦点を当て、ヒト被験者を対象として評価しなければならない。分析バリデーションは、センサー技術の開発を担当するエンジニアリングチーム、データ科学者／分析者／統計学者、生理学者又は行動科学者、そしてヒト被験者のテストを担当してデータを取得しアルゴリズムを導き出す臨床チームのコラボレーションが理想的である。このような多分野にまたがるチームは、すべて一つの組織内にある場合もあれば、技術メーカーと分析会社、学術組織、医療製品メーカーとの間に分かれている場合もある。市販の技術を使用することを選択した新医薬品のスポンサーは、通常、自ら分析バリデーションを実施する必要がある[5]。

市販の技術メーカーは、可能な限り幅広い消費者層に製品を販売するために、様々な被験者集団に幅広く適用できる汎用的なアルゴリズムの開発に重点を置くことが多い。新医薬品のスポンサーが規制当局の承認を得るため、新医薬品の安全性や有効性を評価する際に BioMeTs を使用しようとする場合、スポンサーは必然的に関心のある患者集団に正確に的を絞った適用範囲の狭い特殊なアルゴリズムの開発に重点を置く。スポンサーは一般に、臨床試験集団を通じて、アルゴリズムを開発するための特定の健康状態にある患者の大規模なデータセットにアクセスできる。スポンサーは、臨床試験（初期及び後期の両方）において前向きな探索的指標として BioMeTs を含めることができ、収集されたデータをアルゴリズムの開発に用いることができる。その結果、スポンサーは臨床試験中に収集された他のデータをアルゴリズムの代用参照標準として使用することができる。スポンサーは、アルゴリズムの分析バリデーションを詳細に文書化すべきであり、これらの結果を規制機関に提出することが求められる[5]。

SV95C の事例でも、分析バリデーション結果を文書化し規制機関に提出したことが示されている。(3.4.1 正確性 : Accuracy)

2.3.3 変化を検出する能力

The Playbook には、使用目的の範囲内で、個々の患者や異なる患者間において意味のある変化を検出するためのエビデンスを提供することと示されている[4]。なお本報告書において、変化を検出する能力 (Ability to detect change) は、後に示す臨床バリデーションで評価すべき内容と重複すると本 TF は考えるが、本報告書では The Playbook のフレームワークに基づき、

分析バリデーションのセクションにそのまま記載している。また本報告書では、反応性とは個人内の変化に対する感度のことであり、反応性と変化を検出する能力を同じ意味で用いている。

測定概念に関して変化が認められた個人又は集団（臨床試験における患者と同様）において、それをセンサ由来のエンドポイントがスコアの差として経時的に同定することができるというエビデンスを示すべきである。個々の患者における変化を検出する能力の評価においては、対象とする患者集団において予想される測定範囲における改善と悪化の大きさを記述し、信頼性をもって検出できるというエビデンスを示す。これは、測定に関心のある概念に変化を生じさせることが知られている介入を伴う対照研究（比較試験）によって、最も厳密に実証される。患者の状態が関心のある概念に関して変化したことを示す他の信頼できる指標を用いた同時測定で評価することにより、真の変化を同定し、比較反応性を理解することができる。対象概念の増減に対する感度と、対象患者集団に予想される全範囲にわたる変化に対する感度に関するエビデンスを示すべきである[7]。

患者間の差を検出する能力と患者内の変化を検出する能力が等しいと仮定することはできない。各差を検出する感度の評価は、対象患者集団における探索的試験（例えば第Ⅱ相試験）の一部として行われるかもしれないし、観察研究で得られるかもしれない。対象患者集団又は類似の（代表的な）患者集団について、その技術及び提案された評価項目が以前に評価されていれば、当該エンドポイントの感度に関するデータが査読付き文献で得られるかもしれない。同等とは言えないが、患者間の差を検出する感度が高い測定法は、多くの場合、患者内の変化を検出する反応性も高い。信頼性の低い測定法、例えば再テスト信頼性が低い測定法では、変化の検出に反応しにくいことが多い[7]。

SV95C の事例では、疾患の進行による陰性変化、治療介入による陽性変化といった切り口でエンドポイントの変化に関するデータが示されている（3.4.4 反応性：Responsiveness）。

2.3.4 アルゴリズムによるバイアスの説明

The Playbook では、性能又はバイアスは、生態学的に妥当な条件及び対象集団において評価されなければならない、分析バリデーションの試験集団は多様性／代表性がなければならないと示されている[4]。

アルゴリズムは人間が作り出したものであり、完璧なものではない。BioMeTs を倫理的に評価するためには、アルゴリズムのバイアスにも注意を払う必要がある。アルゴリズムにおけるバイアスは、代表性がない、あるいは不完全なデータ、欠陥のある情報への依存から生じる可能性がある。例えば、皮膚がんを検出するような技術において、有色人種の患者では正確に診断できる確率が低くなるような場合がある[14]。

2.4 臨床バリデーション (Clinical Validation)

The Playbook では、臨床バリデーションのプロセスで以下をチェック項目に挙げている。

- ✓ センサー由来のエンドポイントが意図した意味のある健康の特徴を測定できること
(The endpoint derived from the processed sensor data measures the intended meaningful aspect of health)
- ✓ 測定における臨床的に意味のある変化 (Clinically meaningful change in the measure)

- ✓ 健康の特徴を正しく推定するために必要な最小限のデータ量、欠損データの許容範囲
(Minimum data set required / acceptable limits of missing data)

臨床バリデーションは、エンドポイントに使用される処理されたセンサーデータが、人の機能の意図した意味のある側面を適切に識別、測定あるいは予測できること証明するプロセスである[7]。

2.4.1 センサー由来のエンドポイントが意図した意味のある健康の特徴を測定できること

The Playbook では、エンドポイントは他の臨床指標と関連して期待される推移を示すか、臨床指標と期待どおりの相関があることを示す証拠が必要となる、と示されている[4]。

センサー由来のエンドポイントが意図された MAH と十分に強い関連性を有し、意図された有益性の主張に対して十分に包括的であることを適切に示すための証拠をすべて記載するが、このような内容的妥当性の証拠は、提案された使用状況に関連したものでなければならない、とされている[7]。

測定対象として選択された概念が既に確立され、受け入れられているような状況においては、広範囲にわたる新たな証拠は必要なく、既存の文献やその他の証拠によってそのエンドポイントを受け入れることを正当化できる場合もある。例えば、喘息や COPD の肺機能検査を携帯型スパイロメーターで行う場合などである。しかし、新しいセンサーに基づくエンドポイントが従来認められていた評価方法とは異なる方法を用いる場合、新たなエンドポイントと確立された評価方法とが本質的に同じ情報を提供していることを示す根拠が必要となる[7]。

これまでの評価方法は、治療効果を正しく反映するものとして受け入れられてきたが、モバイルセンサーを用いた測定では、より高頻度、より長期間にわたり、患者の真の状態をより良く、より正確に測定できる利点があるかもしれない。さらには、高感度な測定を低コスト低負担で実施できる可能性も考えられる[7]。

構成概念妥当性とは、処理されたセンサー由来のデータが、COI を正確に測定しているのかということである。例えば、評価期間中に患者が歩いた歩数のカウントを提供することを意図している場合、デバイスが提供する値（及びファームウェアや他の関連ソフトウェアが変換する値）が正確に検出してカウントされた歩数であることを示すというものである。構成概念妥当性の証拠は、多くの場合、並行妥当性のアプローチによって示される。つまり、目的の臨床変数を評価するものとして既に受け入れられている他の評価をセンサー由来のエンドポイントと同時に測定し、通常、相関分析又は回帰分析を通じて、適切に類似していることを確認するというものである。研究者は、センサー由来のエンドポイントが比較対象となる方法とどの程度密接に関連すべきかを特定し、構成概念妥当性を実証するための許容基準を定義すべきである[7]。

SV95C の事例では、対象患者と対照集団とのスコア差や、SV95C と既存のスコアが同様の推移で観察されることを示している（3.4.3 構成概念妥当性：Construct validity）。

2.4.2 測定における臨床的に意味のある変化

The Playbook では、重症度が異なることが知られている患者グループを区別できる証拠、

重症度の経時的変化を検出できるという証拠、変化がないと予想される場合に安定しているという証拠を示すことと示されている[4]。

センサー技術から得られる測定値をエンドポイントとして使用するためには、それを明確に解釈する必要がある。しかし、センサー由来のエンドポイントにおける統計的に有意な改善に臨床的意義があるかどうかについて、解釈する基準がなければそれを判断することはできない。臨床的に意味のある患者内変化の閾値は、例えば記述法や回帰法を用いて、アンカー（ある指標における臨床的变化を解釈するために用いる別の指標）に基づく方法で得られたデータから推定される[7]。

既存の臨床尺度をデジタル化する場合、臨床バリデーションはほとんど完了している。臨床尺度が新しいものであるか、新しい環境で測定される場合、より包括的な臨床バリデーションが必要である[15]。

SV95C の事例では、有意な疾患進行がないと仮定される期間において SV95C に変化がないこと、分布に基づく方法及びアンカーに基づく方法により臨床的に意味のある変化を示している(3.4.2.1 再テスト信頼性: *Test-retest reliability*, 3.4.5 意味のある変化の閾値: *meaningful change threshold*)。

2.4.3 健康の特徴を正しく推定するために必要な最小限のデータ量、欠損データの許容範囲

The Playbook には、健康の側面を正しく推定するために必要な最小限のデータ量／概念は事前に定義されるべきであると示されている[4]。

10 万人以上の参加者に手首装着型加速度計を 7 日間装着し、得られたデータから身体活動量を収集した UK Biobank Study の結果が報告されている。この試験では、7 日間完全に装着していた約 3 万人の参加者を対象とした欠損データシミュレーションを行っている。そこでは、完全に装着した場合の測定値と装着時間が短かった場合の測定値を比較し、本来の測定値の 10% 以内の値を得るために 72 時間の装着時間が必要であるとされている[16]。

どのようなアルゴリズムによって測定値が得られるかによるところも大きいですが、必要最小限のデータセットや欠損データについても配慮が必要である。

SV95C の事例では、患者内変動が許容範囲に収まるために必要な機器装着時間を検討している(3.4.2.2 頑健性: *Robustness*)。

3 dBM のバリデーション事例

欧州医薬品庁(EMA)は、医薬品の研究開発において有望であることが示された革新的な開発手法の適格性に関する意見や、新規の方法論に対するサポートレターを公開している[17]。そこで公開されている SV95C に関する意見書では、当初は SV95C と身体機能の長期間にわたる相関関係などのデータが不足していると判断されたため、SV95C はデュシェンヌ型筋ジストロフィー(Duchenne Muscular Dystrophy: DMD)患者を対象とした検証的又は探索的試験の主要評価項目ではなく副次評価項目として用いることが認められていた[18]。その後、いくつかのデータを追加することで、SV95C は従来の指標に代わる外来 DMD 患者を対象とした優越性試験における主要評価項目として用いることも適格と判断された[19]。

3 章ではこの意見書の内容を紹介する。DMD, ActiMyo, SV95C に関する情報を 3.1 に、副次評価項目ではなく主要評価項目として認められるために必要な事項を 3.2 に、SV95C において実際に検証された事項を 3.3 及び 3.4 に記載した。

なお、ActiMyo や SV95C については、日本製薬工業協会 医薬品評価委員会 臨床評価部会の成果物にも記載されているので参照していただきたい[1] [8]。

3.1 背景情報

3.1.1 DMD の患者層

DMD は、進行性の遺伝性疾患で、主に男児に発症し、5000 人に 1 人が罹患している[20]。DMD 児の歩行は進行性の筋力低下や関節拘縮の影響で歩行速度が遅く、ステップ長が短縮する。同じ速度で歩く場合でも、短い歩幅を補うため歩調が速くなる傾向がある。現在、以下のアウトカム指標が用いられている。

- 6 分間歩行試験（6 Minute Walk Distance : 6MWD）は、患者が最大限の努力（走ることは許されない）で、25 m 間隔の 2 つのコーンの間を往復して歩いた距離を測定する試験である。神経筋疾患患者にとっては非常に負担が大きい指標であり、その日の患者のモチベーションや集中力に大きく依存する。評価者には十分なトレーニングが求められ、すべての評価者が標準化されたスクリプトを使用する。ただし、例えば評価者が試験中に患者を励ます方法など、一部に主観的な要素が残る。
- ノース・スター歩行能力評価（North Star Ambulatory Assessment : NSAA）には、17 種類の機能活動が含まれる（10 m の歩行/走行、座位から立位への立ち上がり、片足立ち、ステップ台の昇降、仰臥位から座位への移行、床からの立ち上がり、ジャンプなど）。患者は 3 段階評価で採点される[21]。これは主観的な測定方法であり、臨床的に意味のある変化は非常に大きい（リニアスケールで 8~9 ポイント）[22]。
- 4 段階昇降テスト（4-Stair Climb test : 4SC）は、患者が 4 段の階段を上るのに必要な最短時間を測定するもので、単に筋力だけでなく、運動の実行能力（モータープラクシス）も評価する。この測定は非常に迅速で約 2 秒程度で実施されることもあり、患者や理学療法士の反応時間を超えてしまうことがある。このようなノイズ（誤差）が生じるため、臨床試験ではより多くの被験者を入れるか、試験期間を延ばす必要が出てくる。

3.1.2 ペインポイント

DMD 患者では、歩行パラメータの変化は臨床管理の指針となり、介入試験の主要評価項目となり得る。したがって、身体機能の変化が臨床的に意味があるかどうかを判断することが重要である。しかし、上記のすべての指標は、一定のパフォーマンスを下回る患者や歩行できない患者に十分に適用できない。また、評価のために患者は専門機関まで移動する必要があるが、これは患者や家族にとってストレスや支障となる。ゆえに病院から離れた場所でも、歩行能力を失いつつある患者にも適応でき、継続的かつ受動的にモニター可能な革新的アプローチが求められている。DMD やその他神経筋疾患における疾患の進行や治療効果を測定することは、すべての臨床開発計画における課題である。新薬の有効性を検証することの難しさ[23]は、この課題の深刻さと、この患者集団に対する新たな信頼性の高い評価指標の開

発の緊急性を強く浮き彫りにしている。

3.1.3 ActiMyo はどのような機器か[18]

CE マークに適合したクラス I 医療機器である ActiMyo®は、2つのバッテリー駆動携帯型センシングデバイスとドッキングステーションから構成されるウェアラブルデバイスであり、磁気慣性技術によって動きと軌跡を連続的に記録し分析することができる。歩行者の場合、1つのセンシングデバイスを手首に、もう1つを足首の近くに装着する。データは各センシングデバイスの内部メモリに格納され、夜間にドッキングステーションで充電する際、ドッキングステーションに転送される。

ActiMyo によって、足首に取り付けたウェアラブルデバイスから得られるデータから、ストライドの軌跡が再構築される。デバイス装着者が歩くと、この軌跡から装着者の歩行能力が5つの異なる変数によって連続的に定量化される。このうち、デバイス装着期間中に得られたすべてのストライドにおけるストライド速度（ストライドの距離をその継続時間で割ったもの、単位は m/s）の上位 5%に相当する値である SV95C は、装着者がとった最も早い歩行を反映しているので、装着者の最大の歩行能力を示す良い指標である。

3.1.4 ActiMyo のベリフィケーション[19]

デバイスは、すべての歩行速度におけるストライド（低速から高速、方向転換のストライド）を検出できる必要がある。下肢が踵を中心に回転しているという原則にのっとり、足首の加速度と角運動速度を結びつけるモデルに基づいて、ストライドの開始から終了までを判別している。ストライドとストライド速度は、 1σ (68% CI) で誤差が 2.5%以下で正確に測定できるものとする。

また、CE マーク適合のための、セキュリティ (EN 60601-1:2007 professional healthcare, EN 60601-1-11:2015 at home)、電磁環境両立性 (IEC 60601-1-2:2014 professional healthcare, IEC 60601-1-2:2014 at home)、生体適合性 (ISO 10993-1:2009)、有用性 (IEC 60601-16:2010, IEC 62366-1:2015) の試験に合格している。なお、ソフトウェア開発は EC 62304 に準拠している。

3.1.5 DMD における SV95C[19]

DMD のような近位筋力低下を特徴とする進行性の神経筋疾患 (NeuroMuscular Disorders : NMD) における疾患進行と治療への反応性を測定することは、すべての臨床開発計画における課題である。DMD における潜在的に有効な治療アプローチの数は増えており、1年といった長期にわたる臨床試験において、臨床的に意味のある治療効果を示す有用な臨床アウトカム評価 (COA) の需要はこれまで以上に高まっている。この目的において、SV95C は他のゴールドスタンダードである COA (例: 6MWD) やその他の機能スケールと比較して、特に有用である。それは、6MWD のような従来の COA には制御された試験環境下で得られた限られた時点のデータしか反映されないのに対して、SV95C には、制御されていない現実の日常環境下で被験者が移動する際の最大歩行速度を一定期間継続的に収集したデータが反映されているからである。また、SV95C は、検査時や主観評価時における患者のモチベーションや疲労レベルに依存しない。このように、SV95C は、標準化された他の測定法で認められるようなバイアスに影響されにくく、他の測定法によって捉えられた結果よりも患者の実際の運

動機能をより表した、より臨床的に適切な評価である。したがって、SV95C は、DMD 患者の歩行能力の維持、改善又は低下における治療の有効性を示すことを目的とした臨床試験や、疾患の経過を特徴付けることを目的とした自然史研究において、より適切な評価であると言える。

歩行能力を失いつつある DMD 患者では、最大歩行速度は時間の経過とともに減少するため、これが安定化又は改善されれば、疾患状態の改善又は進展抑制を示すことになる。したがって、SV95C は、治療前と治療後の比較（被験者内比較）又は治療患者と未治療患者の比較（被験者間比較）のいずれかによって、治験薬によって誘発される歩行速度の変化を評価するために使用することができる。SV95C は、疾患の自然経過における歩行速度のベースラインからの時間的变化を評価するためにも使用することができるため、より広範な測定戦略の一部として使用することができる。

3.2 主要評価項目として利用する際に必要な追加のデータ

2.3 章、2.4 章ではそれぞれ分析バリデーション、臨床バリデーションで評価すべき特性やプロセスについて説明した。一般に分析バリデーションでは正確性や変化を検出する能力、臨床バリデーションでは内容的妥当性、構成概念妥当性、信頼性、感度（または、個人内の変化に対する感度である反応性）、意味のある変化の閾値などを評価する必要がある。なお、2.3.3 章で述べたとおり、分析バリデーションにおける変化を検出する能力は、臨床バリデーションにおける感度（または、個人内の変化に対する感度である反応性）の評価の過程で、評価可能な特性と本 TF は考える。

副次評価項目としての利用を目的としたガイダンスに記載されているとおり、当初は SV95C と既存の臨床的に関連する COA との相関に基づき、SV95C は正確性及び信頼性が高く、変化に対する感度が高く、臨床的に関連することが確認されたため、ピボタル試験の副次評価項目として利用することが認められた。しかしながら、ピボタル試験の主要評価項目として利用するには評価が十分ではなく、追加データが必要と判断された。欧州医薬品委員会（Committee for Medicinal Products for Human Use : CHMP）からのフィードバックは以下のとおりであった。

主要評価項目としての利用を検討する場合、患者数を増加し、長期的な追跡調査によって得られるより頑健なデータが有益である可能性がある。SV95C と参照機能テストとの長期的な相関関係を強化し、基準となるデータを拡充し、MCID の臨床的関連性の正当性をさらに支持することが推奨される

主要評価項目としての利用を目的としたガイダンスに記載されているとおり、このフィードバックを踏まえて、定性的エビデンスである内容的妥当性（3.3 参照）、定量的エビデンスである再テスト信頼性（3.4.2.1 参照）、構成概念妥当性（3.4.3 参照）、反応性（3.4.4 参照）及び意味のある変化の閾値（MCT）（3.4.5 参照）に関する追加的なデータ収集や解析を実施した。正確性（3.4.1 参照）及び頑健性（3.4.2.2 参照）については、副次評価項目としての利用を目的とした評価で十分と判断し、追加的なデータ収集は実施しなかった。定量的エビデンスの評価に用いた各試験の症例数は、主要評価項目としての利用を目的としたガイダンス

[19]の「Table 1 : Source of Data for the DMD and Control Populations」を参照いただきたい。

3.3 章及び 3.4 章では、主要評価項目として利用する際に必要なエビデンスを中心に説明する。

3.3 定性的エビデンス（内容的妥当性）

臨床試験のための Patient-Reported Outcome (PRO) 使用ガイダンス[24]では、内容的妥当性は「尺度に含まれる項目が測ろうとしている興味のある対象領域の内容を反映しているかどうかに関する性質」と述べられており、妥当性の中でも最重要とされる事項である。本報告書の対象は PRO ではなく dBM であるが、両者には多くの共通項があり、相互に参考になることが多い。

SV95C の主要評価項目としての利用を目的としたガイダンスでは、DMD 患者が重要視する健康上の側面は何か、それはどのような身体症状として現れるか、どのようなアウトカムに落とし込むべきかという調査が行われている。また、前述のガイダンス[24]において「PRO 尺度においては、作成にあたって対象者から十分に意見を聴取し反映させているかどうか」が特に重要であると述べているとおり、当事者（患者、介護者及び医療従事者）に対する意見聴取が肝心であり、調査手法としてはヒアリングが中心となる。本件ではアンケート調査、学会での意見聴取及びパブリックコメントが活用されている。

3.3.1 患者の健康にとって重要なこと（MAH）の調査

- ・患者及び介護者に対するアンケート調査

オンライン調査を用いて NMD 患者及び介護者のアンケート調査を行った。患者及び介護者 549 名から回答を得た。そのうち 92 名は DMD 患者及びその家族からの回答であった。アンケート結果から、歩行を含む身体活動の維持は DMD 患者／介護者の観点から重要な要素であることが確認された。具体的には、患者が最初に気づく症状を尋ねたところ、最も多かったものは歩行機能障害であった（DMD 患者全体の 65.9%）。また、臨床試験に対して期待することを質問したところ、最も多くの患者は進行の抑制を挙げており（DMD 患者全体の 45.7%）、具体的にどの症状の進行が重要かを尋ねたところ、歩行可能な患者の過半数は歩行機能を挙げた（歩行可能な DMD 患者の 54.8%）。さらに、改善を望む症状は何かを尋ねたところ、最も多かった回答は筋力（DMD 患者全体の 29.3%）であり、次が歩行機能（DMD 患者全体の 17.3%）であった。

以上のように、DMD 患者にとって運動機能障害が最も重要な症状であることが示唆された。

3.3.2 興味のある評価対象（COI）の調査

- ・患者及び介護者に対するアンケート調査

前述のアンケート調査において、DMD が身体活動の具体的にどの要素にもっとも制限を与えるかを調査したところ、歩行可能及び歩行不能のいずれの DMD 患者でも、歩行が最も多く挙げられた（DMD 患者全体の 29.4%）。さらに、身体活動関連症状の中でどれが重要かというアンケートにおいて、最も重要（Very important）と回答された上位 3 つの要素は転倒

のリスク、自己移動、歩行機能であり、全てが歩行機能と関連していた。([19] Figure 3)

以上のように、DMD 患者にとって運動機能障害の中でも、特に歩行機能の維持が重要であることが示唆された。

3.3.3 測定対象のアウトカムの選定

・患者及び介護者に対するアンケート調査

前述のアンケート調査において、歩行中の最高速度の変化が歩行改善を代表しているか具体的に尋ねたところ、全体では、回答した DMD 患者の 57.8%が同意し、この同意は歩行可能及び歩行不能の DMD 患者の両方に対して認められた。DMD を含む NMD の外来患者 119 名からの回答のみを考慮すると、患者の 84%が歩行中の最高速度の変化は歩行の改善を示すことに同意した。

・医療従事者に対する調査

オンライン調査で医療従事者に対して質問を行ったところ、ほとんどの回答者は最大歩行速度を測定して患者の歩行能力の変化(低下又は改善)を評価することの価値を認めた。([19] Figure 2)

以上の調査結果から、患者の歩行機能を評価する上で、ストライド速度の上位 95%を指標とすることは妥当であるという定性的な結論が得られた。

3.3.4 デジタルバイオマーカーを用いた自宅測定の意義の確認

・ Parent Project Muscular Dystrophy (PPMD) 年次総会において患者とその家族に対してアンケート調査を行った。回答した家族 156 名のうち、11%が臨床試験の参加において長距離の移動が必要なことを課題に感じていることがわかった。

・ SV95C の副次評価項目としての利用を目的とした EMA ガイダンス策定時に得られたパブリックコメントから、下記のとおり現状の 6MWT などの既存の評価指標に対する課題が認められた。

- ① 評価日の患者の状態(意欲や集中力など)や小児の成長及び知的成熟に大きく影響を受ける
- ② 患者が来院する必要があるため患者とその家族の負担が大きい
- ③ 測定に時間がかかり、疲労を誘発するため患者への負担が大きい
- ④ 測定に時間がかかり、スペースを必要とするなど病院に対する負担も大きい

・ 複数の医師から、実生活における機能評価は重要であるとのレターが得られた。

- ① SV95C は自宅での患者の自発的な最大速度を表し、患者の実際の最高の成績をより良く表している。(Lena Szabo, Head of Neurology ward in the Second Department of Pediatrics in Hungary)
- ② COVID-19 のパンデミックの経験がいくつかの臨床試験の過程を混乱させ、日常生活の中でデータを収集する必要性が明らかになった。(Laurent Servais, Professor of

以上のように、患者、介護者及び医療従事者全てが、デジタルデバイスを用いて在宅で症状を評価することの価値を認めた。

3.4 定量的エビデンス

本章では定量的エビデンスについて説明する。上述のとおり、正確性と頑健性を除く定量的エビデンスについては、主に主要評価項目として利用する際に必要なエビデンスについて説明する。正確性と頑健性については、副次評価項目として利用する際に評価したエビデンスについて説明する。

3.4.1 正確性 : Accuracy

正確性は、dBM が実際の健康状態や疾患の指標をどれだけ正しく反映しているかを表す。dBM の測定値が実際の値に近いほど正確性が高いことを意味する。

3.4.1.1 方法

DMD 患者 23 名 (31 回のテスト結果を分析) を対象に、6MWT 中に理学療法士が測定した 6 分間の歩行距離と、ウェアラブルデバイスで計算した距離を比較した。

また、モーションキャプチャールームで、8 名の健康な被験者が 3 つの定義された軌道に基づいて、3 つの異なる歩行リズム (遅い, 通常, 速い) で歩行したデータを収集した。

3.4.1.2 結果

理学療法士が測定した距離とウェアラブルセンサーで計算された距離が概ね一致していることが示された (6MWT の各 25 m の廊下の端でコーンを回る距離を調整した後の差と 95% 信頼区間幅は、 0.75 ± 8.9 m であった)。

また、モーションキャプチャールームでの評価では、システムにより 98.7% の歩幅を検出し、速い歩行速度では、システムにより歩幅速度を高精度で測定することができた (平均の差 0.01 cm/s, 平方根平均の差 1.02 cm/s)。

3.4.2 再現性 : Repeatability

3.4.2.1 再テスト信頼性 : Test-retest reliability

再テスト信頼性は、同じ測定条件で測定を行った際に、測定間の結果の一致性を評価するものである。今回は連続量であるため、評価には級内相関係数 (Intraclass Correlation : ICC) を用いる。ICC には複数のタイプがあるが、本検討では 2 元配置ランダム効果モデルによる ICC を用いて、数値の一致の程度を評価する。

3.4.2.1.1 方法

各測定回において 50 時間以上測定した DMD 患者 52 名の測定データを用いた。1 ヶ月間の測定を連続して 2 回記録し、ICC を計算することにより評価した。また、測定間の一致性を視覚的に評価するために Bland-Altman plot も図示した。

3.4.2.1.2 結果

SV95C の再テスト信頼性を評価した結果、ICC とその 95%信頼区間は 0.970 (0.947, 0.983) となり、2 ヶ月間に有意な疾患進行がないと仮定した下で、良好な再テスト信頼性が確認された。さらに、Bland-Altman plot を図示した結果、均一な分布となり、良好な再テスト信頼性が示唆された ([19] Figure 6B)。

3.4.2.2 頑健性 : Robustness

頑健性とは、無効な入力が存在する場合やストレスの多い環境下においても、システムが機能し続ける程度を表す。

3.4.2.2.1 方法

制御されていない設定下で、180 時間以上 ActiMyo を装着した DMD 患者 28 名を対象に患者内変動を評価し、記録期間の閾値を検討した。

また、45 名及び 10 名の DMD 患者を対象に SV95C の変動に対する記録時点（午前 vs. 午後、平日 vs. 週末など）の影響を評価した。

3.4.2.2.2 結果

DMD 患者を対象とした患者内変動の評価では、180 時間に渡り評価した SV95C の変動は小さい (4.41%) ことが示された。また 50 時間以上の記録データでは、SV95C の患者内変動は 6.38% となり、6MWD（患者内変動は約 10%）と比較しても、許容範囲内であることが示された。一方、50 時間未満の記録データでは、患者内変動は指数関数的に増大することが示唆された。

45 名及び 10 名の DMD 患者を対象に、それぞれ SV95C の変動に対する記録時点（午前 vs. 午後、平日 vs. 週末）の影響を評価した。45 名の患者では、午前（平均：1.564 m/s、標準誤差：0.384 m/s）と午後（平均：1.600 m/s、標準誤差：0.387 m/s）では有意差は認められなかった（差の平均：0.036 m/s）。一方、10 名の患者では、平日と週末では有意差が認められた（差の平均：-7.34 m/s）。このように、平日と週末で活動に大きな相違があることから、バイアスを制御するため、毎日データを収集することが重要である。

3.4.3 構成概念妥当性 : Construct validity

3.4.3.1 既知グループ妥当性 : Known-group validity

既知グループ妥当性は構成概念妥当性の一部であり、関連性の構成概念の点で異なることが知られているグループ間、すなわち、先験的に仮定された臨床的に異なるグループ間の識別能力である。

3.4.3.1.1 方法

SV95C の臨床的妥当性を確認するため、5 歳から 14 歳の DMD 患者 125 名を、筋疾患のない対照被験者 66 名と比較検討した。さらに、年齢とともにパフォーマンスが悪化することが予想されることから、年齢層別（DMD 集団では 5~7 歳、対照被験者集団では 6~7 歳、両集

団では 8～14 歳) の比較も行った。比較は独立標本マン・ホイットニーU 検定 (2 標本の場合) 又は独立標本クラスカル・ワリス検定 (3 標本以上の場合) で行った。

3.4.3.1.2 結果

全体として, SV95C は DMD 患者と対照被験者を識別することができ, DMD 患者 (1.563 m/s) では対照被験者 (2.713 m/s) と比較して SV95C スコアの中央値が低かった ($p < 0.001$) (表 2 参照)。

表 2 DMD 患者集団と対照被験者集団における SV95C の比較

SV95C (m/s)	例数	平均	中央値	SD	最小値	最大値	P 値*
DMD	125	1.571	1.563	0.3818	0.700	2.500	< 0.001
CTRL	66	2.621	2.713	0.4578	1.500	3.600	

CTRL : 対照群, DMD : デュシェンヌ型筋ジストロフィー, SD : 標準偏差, SV95C : 歩行速度の 95 パーセンタイル

*独立標本マン・ホイットニーU 検定

[19] Table 26 (一部改変)

既存の COA でも同様の結果が得られ, DMD 集団における 6MWD スコアの中央値は対照被験者集団と比較して有意に低く (中央値 : 389.0 m vs. 605.0 m, $p < 0.001$) (表 3 参照), 4SC では有意に高かった (中央値 : 3.40 秒 vs. 1.27 秒, $p < 0.001$) (表 4 参照)。

表 3 DMD 患者集団と対照被験者集団における 6MWD の比較

6MWD (m)	例数	平均	中央値	SD	最小値	最大値	P 値*
DMD	109	389.4	389.0	75.6	25.0	512.0	< 0.001
CTRL	63	606.7	605.0	63.5	464.0	761.0	

6MWD : 6 分間の歩行距離, CTRL : 対照群, DMD : デュシェンヌ型筋ジストロフィー, SD : 標準偏差,

*独立標本マン・ホイットニーU 検定

[19] Table 27 (一部改変)

表 4 DMD 患者集団と対照被験者集団における 4SC の比較

4SC (s)	例数	平均	中央値	SD	最小値	最大値	P 値*
DMD	109	3.86	3.40	1.63	1.29	8.70	< 0.001
CTRL	3	1.32	1.27	0.17	1.19	1.51	

4SC : 4 段階昇降テスト, CTRL : 対照群, DMD : デュシェンヌ型筋ジストロフィー, SD : 標準偏差,

*独立標本マン・ホイットニーU 検定

[19] Table 28 (一部改変)

さらに, 年齢層別の比較では, 年少層 (DMD 群 5～7 歳, 対照群 6～7 歳), 年長層 (8～14 歳) とともに, SV95C と 6MWD は対照群と比較して DMD 群で統計的に有意に低かった ($p < 0.001$) (表 5 参照)。

表 5 SV95C, 6MWD における年齢層の影響 (DMD 患者集団と対照被験者集団の比較)

年齢層	例数	平均	中央値	SD	最小値	最大値	P 値*
年齢 (y)							
DMD [5 - 7]	57	6.4	6.6	0.8	5.0	7.9	0.015
CTRL [6 - 7]	17	7.0	7.0	0.65	6.0	8.0	
DMD [8 - 14]	68	9.5	9.0	1.47	8.0	14.0	0.001
CTRL [8 - 14]	49	10.5	10.2	1.7	8.0	14.2	
SV95C (m/s)							
DMD [5 - 7]	57	1.723	1.680	0.346	1.009	2.426	< 0.001
CTRL [6 - 7]	17	2.627	2.673	0.351	1.478	3.074	
DMD [8 - 14]	68	1.444	1.386	0.366	0.723	2.470	< 0.001
CTRL [8 - 14]	49	2.619	2.742	0.493	1.474	3.556	
6MWD (m)							
DMD [5 - 7]	47	401.1	407.0	60.3	247.0	507.0	< 0.001
CTRL [6 - 7]	16	549.5	564.5	45.9	464.0	602.0	
DMD [8 - 14]	62	380.5	373.5	84.9	25.0	512.0	< 0.001
CTRL [8 - 14]	47	626.2	625.0	56.7	489.0	761.0	

6MWD : 6 分間の歩行距離, CTRL : 対照群, DMD : デュシェンヌ型筋ジストロフィー, SD : 標準偏差, SV95C : 歩行速度の 95 パーセンタイル

*独立標本マン・ホイットニーU 検定

[19] Table 30 (一部改変)

DMD 患者集団及び対照被験者集団それぞれの集団における年齢層間の比較において, SV95C 及び 4SC では, 年少層の DMD 患者 (5~7 歳) と年長層の DMD 患者 (8~14 歳) の間に統計的に有意な差が認められた (それぞれ $p < 0.001$ 及び $p = 0.005$) (表 6 参照)。年長層の DMD 患者は年少層と比較して歩幅が小さく, 4 段上がるのに時間がかかったことを示しており (SV95C の中央値 : 1.39 m/s vs. 1.68 m/s, 4SC の中央値 : 3.75 秒 vs. 3.06 秒), このことは, 自然経過の知見から予想されることであった。DMD 患者においては, 6MWD と NSAA には年齢層間で統計的な有意差は認められず, 6MWD と NSAA は年少層と年長層の差を特徴付ける感度が低いことが示唆された。対照群の年齢層間で比較すると, 6MWD において統計的有意差が認められた (6 分間の歩行距離の中央値は, 年長層 (625.0 m) で年少層 (564.5 m) より高かった ($p < 0.001$)). 全体として, 小児では成長と疾患の進行が交絡因子であり, SV95C は 6MWD や NSAA よりも疾患によって誘発される差異を検出する感度が高いと考えられる (表 6 参照)。

表 6 SV95C, 6MWD, NSAA, 4SC における年齢層の影響 (DMD 患者集団及び対照被験者集団それぞれの集団内における比較)

年齢層	例数	平均	中央値	SD	最小値	最大値	P 値*
年齢 (y)							
DMD [5 - 7]	57	6.4	6.6	0.8	5.0	7.9	< 0.001
DMD [8 - 14]	68	9.5	9.0	1.5	8.0	14.0	
CTRL [6 - 7]	17	7.0	7.0	0.6	6.0	8.0	< 0.001
CTRL [8 - 14]	49	10.6	10.2	1.7	8.0	14.2	
SV95C (m/s)							
DMD [5 - 7]	57	1.723	1.680	0.3459	1.009	2.426	< 0.001
DMD [8 - 14]	68	1.444	1.386	0.3656	0.723	2.470	
CTRL [6 - 7]	17	2.627	2.673	0.3511	1.478	3.074	0.514
CTRL [8 - 14]	49	2.619	2.742	0.4926	1.474	3.556	
6MWD (m)							
DMD [5 - 7]	47	401.1	407.0	60.3	247.0	507.0	0.264
DMD [8 - 14]	62	380.5	373.5	84.9	25.0	512.0	
CTRL [6 - 7]	16	549.5	564.5	45.9	464.0	602.0	< 0.001
CTRL [8 - 14]	47	626.2	625.0	56.7	489.0	761.0	
NSAA (#)							
DMD [5 - 7]	47	23.96	23.00	4.800	10	33	0.107
DMD [8 - 14]	62	21.97	22.00	7.126	2	33	
4SC (s)							
DMD [5 - 7]	47	3.38	3.06	1.42	1.29	8.57	0.005
DMD [8 - 14]	62	4.21	3.75	1.70	1.70	8.70	

4SC : 4 段階昇降テスト, 6MWD : 6 分間の歩行距離, CTRL : 対照群, DMD : デュシェンヌ型筋ジストロフィー, NSAA : ノース・スター歩行能力評価, SD : 標準偏差, SV95C : 歩行速度の 95 パーセンタイル
*独立標本マン・ホイットニーU 検定

[19] Table 31 (一部改変)

3.4.3.2 収束的妥当性 : Convergent validity

収束的妥当性とは, ある尺度のスコアが, 同じ構成要素を測定する他の COA のスコアとどの程度関連しているかを示すものである。

3.4.3.2.1 方法

SV95C の収束的妥当性は, 107 名の DMD 患者を対象に, SV95C と既存の COA (6MWD, 4SC, NSAA) との相関を検討することによって評価した。

SV95C の収束的妥当性は, 合計 107 名の患者のデータを用いて, SV95C と既存の COA 指標 (6MWD, NSAA, 4SC) との相関を評価した。評価は, ベースライン時の観測値に加え, 3 ヶ月, 6 ヶ月, 9 ヶ月及び 12 ヶ月経過後の観測値及びベースラインからの変化についても実施された。

3.4.3.2.2 結果

DMD 患者において, ベースラインの SV95C と 6MWD, NSAA 及び 4SC とのスピアマンの相関係数はそれぞれ 0.657, 0.644 及び -0.634, ピアソンの相関係数はそれぞれ 0.678, 0.676 及び -0.622 であり, いずれも統計的に有意な相関がみられた (すべて $p < 0.001$)。特筆すべき点として, SV95C と 6MWD の評価において, SV95C が 1.5 m/s 未満の患者層では良好な相関がみられたのに対し, 1.5 m/s 超の患者層では相関が弱くなる傾向がみられた ([19] Figure 8)。6MWD には患者は可能な限り早く歩行する (ただし, 走らない) ことを指示されることに起

因する天井効果がみられたが、SV95C ではそのような効果は存在しないため、これが患者層によって異なる相関がみられた一因となったと考えられる。このことは、SV95C は、6MWD と比較して、歩行機能をより良く識別できる指標である可能性を示唆するものと考えられる（表 7 参照）。

概して、3 ヶ月、6 ヶ月、9 ヶ月及び 12 ヶ月経過後の観測値についても、ベースライン時と同様に良好で統計的に有意な相関がみられた。しかしながら、ベースラインからの変化量については、有意な相関はみられなかった。これは、時間経過に伴う変化やその変動及び歩行能力の低下を検出する能力が指標によって異なることに起因した可能性が考えられる（3.4.4 参照）。

表 7 ベースラインにおける SV95C と他の臨床アウトカム指標の相関

ベースライン	相関係数（ノンパラメトリック）	SV95C	6MWD	NSAA	4SC	相関係数（パラメトリック）
SV95C			0.678 <0.001 107	0.676 <0.001 107	-0.622 <0.001 107	ピアソン p 値（両側） 例数
6MWD	スピアマン p 値（両側） 例数	0.657 <0.001 107		0.746 <0.001 107	-0.531 <0.001 107	ピアソン p 値（両側） 例数
NSAA	スピアマン p 値（両側） 例数	0.644 <0.001 107	0.674 <0.001 107		-0.641 <0.001 107	ピアソン p 値（両側） 例数
4SC	スピアマン p 値（両側） 例数	-0.634 <0.001 107	-0.518 <0.001 107	-0.639 <0.001 107		

4SC：4 段階昇降テスト，6MWD：6 分間の歩行距離，BL：ベースライン，NSAA：ノース・スター歩行能力評価，SV95C：歩行速度の 95 パーセンタイル

[19] Table 33（一部改変）

3.4.4 反応性：Responsiveness

反応性とは、変化が存在する場合、その変化を検出する能力である。

3.4.4.1 疾患の自然経過における変化を検出できる能力：Ability to detect change

3.4.4.1.1 方法

SV95C の疾患の自然経過における変化を検出できる能力は、安定した副腎皮質ステロイドによる治療を受けている患者又は少なくとも 6 ヶ月前から副腎皮質ステロイドを開始した患者の 3 ヶ月（81 名）、6 ヶ月（59 名）、9 ヶ月（39 名）及び 12 ヶ月（28 名）における SV95C の経時的変化を用いて評価した。変化なしを帰無仮説とする Wilcoxon 符号付順位検定を用いて各時点における変化を評価し、平均値を標準偏差で除して標準化反応平均（standardized response mean：SRM）を算出した。さらに、参考として、SRM を検出するために必要なサン

プルサイズを $n = 2 \times \Phi^2 / \text{SRM}^2$ で算出した。ここで、第一種の過誤 α 及び第二種の過誤 β の値に応じて、 Φ の値は 2.802 ($\alpha=5\%$, $\beta=20\%$), 3.242 ($\alpha=5\%$, $\beta=10\%$) 及び 3.605 ($\alpha=5\%$, $\beta=5\%$) となる。また、年齢層で層別した患者についても同様の解析を行った。

3.4.4.1.2 結果

安定した副腎皮質ステロイドによる治療を受けている患者又は少なくとも 6 ヶ月前から副腎皮質ステロイドを開始した患者において、SV95C に経時的な低下がみられ、ベースラインからの変化はいずれの時点でも統計的に有意であった ($p < 0.001$)。SRM にも経時的な増加がみられた (3 ヶ月後の 0.45 から 12 ヶ月後の 1.03)。これらの結果は、SV95C は 12 ヶ月間の観察又は 3 ヶ月時点であっても疾患の進行を検出するのに十分な感度を有している可能性を示唆していると考えられる (表 8, [19] Figure 15 参照)。

SRM から推定されたサンプルサイズに基づくと、3 ヶ月の観察での SV95C の減少を検出するためには相対的に多数の被験者が必要であるが、6 ヶ月の時点では半数近くに減少し、現在の臨床試験で必要とされるサンプルサイズよりも小さくなる可能性がある。

表 8 安定した副腎皮質ステロイド治療を受けている DMD 患者集団における SV95C の経時変化

SV95C 変化	3 ヶ月後	6 ヶ月後	9 ヶ月後	12 ヶ月後
例数	81	59	39	28
中央値 (m/s)	-0.044	-0.067	-0.110	-0.204
平均 (m/s)	-0.051	-0.080	-0.119	-0.241
SD (m/s)	0.113	0.134	0.186	0.234
P 値*	< 0.001	< 0.001	< 0.001	< 0.001
SRM**	0.45	0.60	0.64	1.03
サンプルサイズの推定 $\alpha = 5\%$ $\beta = 20\%$	78	44	38	15
サンプルサイズの推定 $\alpha = 5\%$ $\beta = 10\%$	104	58	51	20
サンプルサイズの推定 $\alpha = 5\%$ $\beta = 5\%$	128	72	63	25

DMD : デュシェンヌ型筋ジストロフィー, SD : 標準偏差, SRM : 標準化反応平均, SV95C : 歩行速度の 95 パーセンタイル

*1 標本の Wilcoxon 符号順位検定

**SRM は | 平均 | / SD により算出

サンプルサイズは $2 \times \Phi^2 / \text{SRM}^2$ により算出。(α , β) = (5%, 20%), (5%, 10%), (5%, 5%) に対応する Φ はそれぞれ 2.802, 3.242, 3.605 となる。

[19] Table 42 (一部改変)

年齢層別に解析した結果、各年齢層において、3 ヶ月時点から 12 ヶ月時点までのベースラインからの変化の中央値は継続的に低下しており、最大速度の喪失が時間とともに進行していることが示唆された。8 歳から 14 歳の患者では、5 歳から 7 歳の患者と比較して、より大きな減少がみられた (8 歳から 14 歳の患者では、ベースラインからの変化の中央値は 3 ヶ月時点で -0.044 m/s, 12 ヶ月時点で -0.210 m/s, 5 歳から 7 歳の患者では、ベースラインからの変化の中央値は 3 ヶ月時点で -0.023 m/s, 12 ヶ月時点で -0.197 m/s であった) ([19] Table 43 参照)。

さらに、SV95C の反応性は 12 ヶ月間継続して追跡調査された 17 名の患者でもみられた。症例数は少ないものの、3 ヶ月時点から 12 ヶ月時点にかけて、ベースラインからの変化の中央値が継続的に低下した（各時点のベースラインからの変化の中央値はそれぞれ -0.043, -0.067, -0.157, -0.197 であった）（[19] Table 56 参照）。

3.4.4.2 状態を改善させる治療による変化を検出できる能力

3.4.4.2.1 方法

SV95C の状態を改善させる治療による変化を検出できる能力は、副腎皮質ステロイドによる治療を開始した 11 名の DMD 患者の SV95C の経時的推移を用いて評価された。評価方法は、3.4.4.1.1 に記載した方法と同様である。

3.4.4.2.2 結果

副腎皮質ステロイドの投与を開始した 11 名の DMD 患者において、SV95C の改善を示す変化に対する感度を評価した。少ない被験者数であるにも関わらず、3 ヶ月の時点で SV95C に有意な正の変化がみられた（ $p=0.003$ ）。その後 6 ヶ月後まで SV95C 変化スコアの中央値に上昇がみられた。9 ヶ月及び 12 ヶ月でもさらなる改善がみられた（変化は統計的に有意ではなかったが、例数が少ないことによる可能性がある）。統計的に有意な変化がみられた 3 ヶ月及び 6 ヶ月後では、いずれも SRM は 1 を超えており、SV95C は治療による改善の効果を検出するための感度が良好であると考えられる（表 9 参照）。

表 9 副腎皮質ステロイド治療を開始した DMD 患者集団における SV95C の経時変化

SV95C 変化	3 ヶ月後	6 ヶ月後	9 ヶ月後	12 ヶ月後
例数	11	7	7	5
中央値 (m/s)	0.090	0.211	0.218	0.307
平均 (m/s)	0.135	0.247	0.208	0.266
SD (m/s)	0.128	0.172	0.233	0.275
P 値*	0.003	0.018	0.063	0.080
SRM**	1.05	1.44		
サンプルサイズの推定 $\alpha = 5\% \beta = 20\%$	14	8		
サンプルサイズの推定 $\alpha = 5\% \beta = 10\%$	19	10		
サンプルサイズの推定 $\alpha = 5\% \beta = 5\%$	23	13		

DMD : デュシェンヌ型筋ジストロフィー, SD : 標準偏差, SRM : 標準化反応平均, SV95C : 歩行速度の 95 パーセンタイル

*1 標本の Wilcoxon 符号順位検定

**SRM は $|\text{平均}| / \text{SD}$ により算出

サンプルサイズは $2 \times \Phi^2 / \text{SRM}^2$ により算出。 $(\alpha, \beta) = (5\%, 20\%), (5\%, 10\%), (5\%, 5\%)$ に対応する Φ はそれぞれ 2.802, 3.242, 3.605 となる。

[19] Table 64 (一部改変)

全体として、SV95C は早い時期に DMD 患者の身体機能の変化を検出することができ、既存の COA と比較して少ない被験者数で変化を検出できると考えられる。

3.4.5 意味のある変化の閾値 : meaningful change threshold

意味のある変化の閾値とは、関心のある評価項目でのスコアの変化が患者にとって意味があるかどうかを示す値である。当該評価項目での変化の大きさは、患者が重要と認識している程度を反映する必要がある。また、意味のある変化の閾値には、治療群間での意味のある変化の差の閾値と、患者内での意味のある変化の閾値があり、両者は異なる閾値であることから、一般に値も異なる。両方とも重要ではあるもの、FDA や EMA は患者内での意味のある変化の閾値に重点を置いている。また、意味のある変化の閾値の導出は、薬事承認や labeling claim（添付文書に記載する効能・効果の範囲）を裏付けるために、評価項目を開発したり検証したりする際に重要となる。さらに、意味のある変化の閾値は単独で開発すべきではなく、研究デザイン、MAH、COI、アンカーの尺度及びデジタルエンドポイントの開発自体と並行して検討すべきである[25]。

意味のある変化の閾値を設定する方法には主に 3 つの方法がある。アンカーに基づく方法 (anchor-based approach) [26]、分布に基づく方法 (distribution-based approach)、定性的な方法 (qualitative-based approach) である。アンカーに基づく方法では、アンカーとして機能する外部の評価項目を用いて、その評価項目での変化と、関心のある評価項目での意味のある変化の閾値を関連付けて解釈する。デジタルエンドポイントにより評価した COI と、外部の基準として選択したアンカーの評価項目との関連性に基づき、意味のある変化を達成した患者を定義する方法である。アンカーとしては、シンプルで解釈し易い評価項目を用いることが望ましいが、何よりも関心のある評価項目と十分に関連したアンカーを選択することが重要である。さらに、1 つではなく複数の独立したアンカーを選択することで、より適切で信頼性の高い閾値の設定が可能になると考えられる。次に、分布に基づく方法では、患者の声（主観）を直接取り込むのではなく、その代わりに測定値の統計的な分布に基づいて閾値を設定する。本手法は、アンカーに基づく方法により導出した閾値に対して追加的な説明を加えることを目的に、一般に補助的に用いられる。しかしながら、適切なアンカーが見つからない場合には、主に分布に基づく方法を用いて閾値を設定するケースも考えられる。本方法には、標準測定誤差 (SEM) を用いる方法や標準偏差 (SD) を用いる方法などがある。最後に定性的な方法は、患者のインプットにより、患者にとって重要なスコアの変化を構成するものを決定することで、意味のある変化の閾値を理解するために補足的に用いられる。

これらの 3 つの方法の利点・欠点を表 10 に示す。なお、実際には、1 つではなく複数の方法を適用することで、より適切で信頼性の高い閾値を設定することが推奨される。

表 10 意味のある変化の閾値を設定する各方法の利点と欠点

方法		内容
アンカーに基づく方法	利点	患者の経験と関連している。 治療群間及び患者内での閾値を導出する方法がある。 規制当局に推奨されている。
	欠点	様々なデジタルエンドポイントがある中で、COIに一致するアンカーを特定することは一般に困難であるため、アンカーの選択が問題となる可能性がある。 「外部のPROに基づくアンカーにおける日常生活での特異的な変化」に対するデジタルエンドポイントでの特異的な変化（例：歩行速度の改善（メートル／秒））の定義が抽象的である。 測定時点が多い強縦断データ（intensive longitudinal data）に対応した適切な解析手法が必要となる場合がある。
分布に基づく方法	利点	簡単に適用できる方法である。 アンカーに基づく方法により導出した閾値に補足的な説明を加えることができる。 治療群間及び患者内での変化に対する方法がある。
	欠点	患者からのインプットを必要としないため、臨床的に意味のある変化又は差であるかどうか疑わしい場合がある。 閾値の推定値は、ばらつきや選択した時点に依存するため、用いるサンプル間で異なる。 特に主要評価項目としての利用に際しては、規制当局に受け入れられない。 デジタルエンドポイントによって測定したパラメータに対するヒト固有の変動のため、推定値が増大する可能性がある。
定性的な方法	利点	患者の声に基づく方法である。 デジタルエンドポイントが一般的に意味があることを確認するのに役立つ。 アンカーに基づく方法を適用する際に、アンカーの解釈に役立つ。 複数の方法を用いる場合、試験データと併合できる。
	欠点	単一の値に設定することは非常に困難である。 中等度のあるいは活発な身体活動など、特定のデジタルエンドポイントに関連付けたり、議論することが患者にとって困難である。 治療間における閾値の設定が困難である。

[25]の Table 4 を一部改変

SV95C では、副次評価項目として利用する際には、まずは分布に基づく方法のみを用いて閾値を設定していたものの、主要評価項目として利用する際には、EMA からの推奨に基づき、分布に基づく方法に加えて、アンカーに基づく方法を用いて意味のある閾値の設定を行った。以下では、実際に用いた両方法の詳細について説明する。

3.4.5.1 分布に基づく方法

3.4.5.1.1 方法

SEM とは測定誤差に起因する変化量を指し、以下の公式により算出可能である。

$$SEM = SD \times \text{SQR} (1 - ICC)$$

なお、SD は標準偏差、SQR は平方根、ICC は級内相関係数とする。ICC は再テスト信頼性の評価で用いた 2 元配置ランダム効果モデルにより算出可能である（3.4.2.1 参照）。本試験における ICC の算出には、各評価期間で 50 時間以上のデータを有する 103 名の患者データを用いた。

最小の検出可能な変化 (Minimal Detectable Change : MDC) とは測定により検出可能な最小の変化量 (測定誤差よりも統計的に大きい最小の変化量) を指し、顕著な能力の変化を表す指標である。MDC は信頼係数と SEM を用いて、以下の公式により算出可能である。

$$\text{MDC} = \text{z-score} \times \text{SEM} \times \text{SQR}(2)$$

例えば、信頼係数 95%, 90% 及び 80% に対応する z-score はそれぞれ 1.960, 1.645 及び 1.282 となる。また MDC は一般に測定値の $0.2 \times \text{SD}$, $0.5 \times \text{SD}$ 及び $0.8 \times \text{SD}$ の大きさも考慮して設定する。

3.4.5.1.2 結果

3.4.5.1.2.1 SV95c

DMD 患者集団 (年齢層 : 5 歳から 14 歳) 及び対照被験者集団 (年齢層 : 6 歳から 14 歳) での SEM はそれぞれ 0.070 m/s 及び 0.156 m/s となり、対照被験者集団での変動が大きかった。対照被験者集団は疾患による制限を受けておらず、様々な活動を行う可能性が高いことから、スライド速度での大きな変動に関連している可能性が示唆された。信頼係数 80% の SV95C の MDC はそれぞれ 0.127 m/s (DMD 患者集団) 及び 0.282 m/s (対照被験者集団) となった。年少層 (DMD 患者集団 : 5 歳から 7 歳, 対照被験者集団 : 6 歳から 7 歳) 及び年長層 (DMD 患者集団及び対照被験者集団 : 8 歳から 14 歳) でも同様の結果が認められた (表 11)。DMD 患者集団の SV95C の $0.5 \times \text{SD}$ は 0.191 m/s となった (表 12)。これらの MDC や SD の結果から、DMD 患者において、約 0.1 m/s から約 0.2 m/s の変化は測定誤差を超えることが示唆された。

表 11 DMD 患者集団における SV95C の MCID 及び MDC

	DMD			CTRL		
	[5 - 14]	[5 - 7]	[8 - 14]	[6 - 14]	[6 - 7]	[8 - 14]
例数	103	43	60	55	17	38
ICC*	0.962	0.956	0.957	0.851	0.814	0.862
95% CI	[0.943 - 0.974]	[0.918 - 0.976]	[0.928 - 0.974]	[0.744 - 0.913]	[0.500 - 0.932]	[0.735 - 0.928]
SV95C-RP 平均 (m/s)	1.467	1.609	1.365	2.383	2.266	2.435
SV95C-RP SD	0.360	0.338	0.342	0.404	0.321	0.429
SEM** (m/s)	0.070	0.071	0.071	0.156	0.139	0.159
RP に対する SEM (%)	4.780	4.412	5.193	6.539	6.113	6.545
MDC80% (m/s)	0.127	0.129	0.129	0.282	0.251	0.289
MDC90% (m/s)	0.163	0.165	0.165	0.362	0.322	0.371
MDC95% (m/s)	0.194	0.197	0.197	0.432	0.384	0.442

DMD : デュシェンヌ型筋ジストロフィー, CI : 信頼区間, CTRL : 対照群, ICC : クラス内相関, MCID : 臨床的に重要な最小差, MDC : 最小の検出可能な変化, RP : 記録期間, SD : 標準偏差, SEM : 測定標準誤差, SV95C : 歩行速度の 95 パーセンタイル

*ICC は 2 元配置ランダム効果モデルにより算出

**SEM は $\text{SD} \times \text{SQR}(1 - \text{ICC})$ により算出

***MDC は $\text{z-score} \times \text{SEM} \times \text{SQR}(2)$ により算出。信頼係数 95%, 90% 及び 80% に対応する z-score はそれぞれ 1.960, 1.645 及び 1.282 となる。

[19] Table 65 (一部改変)

表 12 DMD 患者集団における SV95C の閾値

ベースライン	例数	平均	中央値	SD	0.2 SD	0.5 SD	0.8 SD
SV95C (m/s)	125	1.571	1.563	0.382	0.076	0.191	0.305

SD : 標準偏差, SV95C : 歩行速度の 95 パーセンタイル

[19] Table 66 (一部改変)

3.4.5.1.2.2 6MWD, NSAA 及び 4SC

信頼係数 80% の 6MWD, NSAA 及び 4SC の MDC はそれぞれ 36.3 m, 2.78, 1.310 s と報告されている[27]。0.5×SD はそれぞれ 37.9 m, 3.2 及び 0.79 s となった。また, SV95C のベースラインに対する SEM (4.78%) は 6MWD (5.14%) と同程度であり, NSAA (6.73%) 及び 4SC (18.72%) よりも低かったことから, SV95C の測定誤差による変動は小さいことが示唆された (表 13)。

表 13 DMD 患者集団における 6MWD, NSAA 及び 4SC の閾値

ベースライン	例数	平均	中央値	SD	MDC 80% *	SEM	ベースラインに対する SEM (%)	0.2 SD	0.5 SD	0.8 SD
6MWD (m)	107	390.0	389	75.73	36.3	20.0	5.14	15.1	37.9	60.6
NSAA	107	22.9	23	6.33	2.78	1.53	6.73	1.3	3.2	5.1
4SC (s)	107	3.82	3.4	1.578	1.310	0.72	18.72	0.32	0.79	1.26

4SC : 4 段階登りテスト, 6MWD : 6 分間歩行距離, BL : ベースライン, DMD : デュシェンヌ型筋ジストロフィー, MDC : 最小の検出可能な変化, NSAA : ノーススター歩行評価, SD : 標準偏差, SEM : 測定標準誤差

*6MWD, NSAA 及び 4SC の MDC は cTAP-MDA-2021 ポスターに公表済み (<https://ctap-duchenne.org/wp-content/uploads/2021/03/cTAP-MDA-2021-poster-MDC-analyses-for-functional-outcomes-FINAL.pdf>)。SEM は公表済みの MDC80%に基づき算出

[19] Table 67 (一部改変)

3.4.5.2 アンカーに基づく方法 (患者内での閾値の設定)

3.4.5.2.1 方法

3.4.5.2.1.1 PODCI 及び CGI-C

Pediatrics Outcomes Data Collection Instrument (PODCI) 尺度及び Clinical Global Impression of Change (CGI-C) 尺度を用いて患者内 MCT を評価した。第 48 週に各尺度のデータを有する DMD 患者数はそれぞれ 15 名及び 12 名であった。

CGI-C 尺度は臨床医が評価する単一項目の尺度であり, 第 48 週 (二重盲検期の終了時) の全般的な患者印象度のベースラインからの変化を評価する。回答は 7 つのカテゴリ (「非常に改善」「かなり改善」「わずかに改善」「変化なし」「わずかに悪化」「かなり悪化」「非常に悪化」(それぞれ 1 から 7 のスコアに対応) からなる。意味のある変化を評価する際には「改善」「変化なし」「悪化」の 3 カテゴリに分類し, 同一期間中に評価した SV95C での変化の平均をカテゴリ別に算出した。

PODCI 尺度は筋骨格疾患に対する身体機能及び健康関連 QOL を測定するための尺度であり, 成人患者及び小児患者に対して利用可能である。回答は両親が記入するか, 青少年患者

の場合は自己報告も可能である。PODCI 尺度は 83 から 86 の質問と、5 つのコア尺度（「上肢と身体機能」「移動と基本的モビリティ」「スポーツと身体機能」「痛み／快適さ」「幸福」と 1 つの全般的機能尺度で構成される。各 PODCI サブ尺度のスコアの取りうる範囲は 0 から 100 であり、スコアが低いほど健康関連 QOL が低いことを示す。「移動と基本的モビリティ」は、臨床試験の副次評価項目で用いられていることから、変化のアンカーとして利用した。変化を表すスコアの閾値は 10% に設定した。同一期間中に評価した SV95C の変化の平均は、変化を表すカテゴリ別に算出した。

3.4.5.2.1.2 6MWD 及び NSAA

CT-B 試験で実施された参照機能テストの 48 週間の変化に基づき、患者内 MCT を評価した。その際には、6MWD (13 名) 及び NSAA (12 名) の信頼係数 80% の MDC (それぞれ 36.3 m 及び 2.78) を用いた。

3.4.5.2.2 結果

3.4.5.2.2.1 PODCI 及び CGI-C

CT-B 試験に登録された DMD 患者の両親又は臨床医がそれぞれ記入した PODCI 尺度及び CGI-C 尺度を用いて SV95C の MCT を評価した。なお、CT-B 試験は治験薬の有効性が認められなかったため、早期に中止した。CGI-C 尺度は第 48 週又は試験終了時の診察時に評価し、第 48 週に評価した CGI-C 尺度のみを解析対象とした (12 名)。PODCI 尺度は第 48 週に評価した (15 名)。

第 48 週の SV95C と CGI-C 尺度の間に強い相関が認められた (スピアマンの相関係数 $\rho = -0.816$, P 値 = 0.001)。また、CGI-C 尺度の悪化が大きいほど SV95C は低くなる傾向が認められた。第 48 週の SV95C のベースラインからの変化と CGI-C の間に相関は認められなかったが、臨床医が「変化なし」か「悪化」と報告した患者の多くは、SV95C について信頼係数 80% の MDC (-0.127 m/s) 以上に (負の方向に) 変化した。

同様に、第 48 週の SV95C と PODCI サブスコア (PODCI のサブドメインである「移動と基本的な移動性」に対応するサブスコア) の間に相関が認められた (スピアマンの相関係数 $\rho = 0.611$, P 値 = 0.015)。一方、第 48 週の SV95C のベースラインからの変化と PODCI サブスコアの間に相関は認められなかった。しかしながら、1 名を除いて、ベースラインと比較して、第 48 週に「移動と基本的な移動性」のスコアが低いと報告した親の子供は、SV95C の減少が認められた。

CGI-C と PODCI のカテゴリに基づき、各グループ内の意味のある変化を評価する解析結果を表 14 に示す。全体的に、CGI-C に基づき「改善」と分類された (臨床医の回答が「非常に改善した」、「かなり改善した」又は「わずかに改善された」であった) 患者は、48 週間にわたる SV95C の変化 (減少) の中央値が最も小さかった (-0.025 m/s)。PODCI に基づき「改善」($\geq 10\%$ の増加と定義) に分類された患者はいなかった。一方で、CGI-C に基づき「悪化」と分類された (臨床医の回答が「わずかに悪化」、「かなり悪化」又は「非常に悪化」であっ

た) 患者や、PODCI に基づき「悪化」と分類された患者 (≥10%の減少と定義) では、最も減少が大きかった (それぞれ-0.280 m/s 及び-0.245 m/s)。

表 14 第 48 週の SV95C の患者内での意味のある変化

	例数	48 週間にわたる SV95C の変化の平均	48 週間にわたる SV95C の変化の中央値
CGI-C			
改善	4 (わずかに改善 4 名)	-0.175 m/s	-0.025 m/s
変化なし	5	-0.302 m/s	-0.210 m/s
悪化	3 (かなり悪化 3 名)	-0.370 m/s	-0.280 m/s
PODCI			
改善 (≥10%の増加)	0	N/A	N/A
変化なし	9	-0.221 m/s	-0.100 m/s
悪化 (≥10%の減少)	6	-0.293 m/s	-0.245 m/s

CGI-C : Clinical Global Impression of Change 尺度, N/A : 該当せず, PODCI : Pediatrics Outcomes Data Collection Instrument 尺度, SV95C : 歩行速度の 95 パーセンタイル

[19] Table 68 (一部改変)

治験薬の有効性が認められなかったにもかかわらず、48 週後に大多数は「わずかに改善」又は「変化なし」と報告し、わずか 3 名が「悪化」と報告した。PODCI では患者／親による改善は報告されなかったが、臨床医によって報告された PODCI 又は CGI-C の負の変化は、子供の臨床状態の重要な変化 (例えば、走る能力の喪失や階段の上り下りができなくなること) に関連している可能性が高く、つまりは病気の経過の特定の節目に相当し、意味のある変化を特定するために使用できると考えられる。「変化なし」又は「悪化」と報告した患者では、48 週後の SV95C の変化の中央値は-0.100 から-0.280 m/s の間であった。それにもかかわらず、「変化なし」と「悪化」の SV95C の変化の中央値の差は、PODCI では-0.145 m/s, CGI-C では-0.070 m/s であった。この結果と、悪化した患者の中には全体的な中央値の変化よりも小さな変化を示す患者がいることを考慮して、アンカーに基づく MCT は約-0.1 m/s とすることで、意味のある変化を示すと考えられる。

3. 4. 5. 2. 2. 2 6MWD 及び NSAA

SV95C の MCT は、アンカーに基づく方法を用いて、CT-B 試験で実施された参照機能テストである 6MWD (13 名) と NSAA (12 名) の 48 週間の変化に基づき評価した。NSAA の 80%の信頼水準の MDC (2.78) と 6MWD (36.3 m) に基づき、患者内の意味のある変化を評価する解析結果を表 15 に示す。48 週後の NSAA の変化の中央値 (-2 から-3 ポイント) に基づき「悪化」と分類された患者では、SV95C の変化の中央値は-0.115 m/s であった (4 名)。同様に、6MWD の変化の中央値 (-30 から-40 m) に基づき、「悪化」と分類された患者では、SV95C の変化の中央値は-0.130 m/s であった (3 名)。このように、本方法でも約 0.1 m/s の MCT を支持する結果であった。

表 15 参照機能テストに基づく第 48 週の SV95C の患者内の意味のある変化

	例数	48 週間にわたる SV95C の変化の平均	48 週間にわたる SV95C の変化の中央値
NSAA			
改善 [2; 3]	1		
変化なし (-2; 2)	7	-0.246 m/s	-0.090 m/s
悪化 [-3; -2]	4	-0.228 m/s	-0.115 m/s
6MWD			
改善 [30; 40]	1		
変化なし (-30; 30)	9	-0.140 m/s	-0.090 m/s
悪化 [-40; -30]	3	-0.287 m/s	-0.130 m/s

6MWD：6 分間歩行距離，NSAA：ノーススター歩行評価，SV95C：歩行速度の 95 パーセンタイル

[19] Table 69（一部改変）

3. 4. 5. 2. 3 SV95C の MCT の設定のまとめ

分布に基づく方法及びアンカーに基づく方法による解析結果から、DMD 患者の変化が 0.07 m/s で評価した測定誤差を超えるためには、少なくとも約-0.10 m/s のスコアの変化が必要であり、-0.10 から-0.20 m/s のスコアの変化には意味があることが示唆された。

これらの閾値は、未治療患者では 6 ヶ月、9 ヶ月及び 12 ヶ月に達成し（ベースラインからの変化量の中央値はそれぞれ-0.07、-0.11 及び-0.20 m/s）、ステロイドを開始した患者では、早ければ 3 ヶ月で達成した（SV95C の変化量の中央値は 3 ヶ月及び 6 ヶ月でそれぞれ 0.09 及び 0.211 m/s）。9 ヶ月及び 12 ヶ月に認められた減少は、6MWD 及び NSAA のベースラインからの変化量で確認された（6MWD の変化量の中央値は 9 ヶ月及び 12 ヶ月でそれぞれ-36.3 m 及び-31.5 m，NSAA の変化量の中央値は 9 ヶ月及び 12 ヶ月でそれぞれ-1 及び-2）。ステロイドによる短期的な（早ければ治療開始後 3 ヶ月の）効果が DMD 患者の全体的な臨床改善につながることから、当該時点の SV95C の変化量の中央値である 0.09 m/s は、測定誤差を超えており、MCT となりうることが示唆された。

以上より、MCT を 0.1 m/s に設定することは妥当と考えられる。この値は測定誤差よりも大きく、（PRO 及び参照機能テストに基づき）悪化したと考えられる患者におけるアンカーに基づくスコアの変化と、ステロイドを開始した後に改善した患者のスコアの変化とも一致している。

さらに、ストライド速度の 0.1 m/s の変化は、6 分間歩行距離の 36 m の変化に相当し、これは DMD 患者の疾患経過における 6MWD の意味のある変化の閾値と一致する。最後に、軽度な DMD 患者（6MWD < 350 m かつ SV95C < 1.50 m/s）では、線形回帰モデルを当てはめた結果（ $R^2 > 0.5$ ）、SV95C の 0.1 m/s の変化は 6MWD 及び NSAA 合計スコアではそれぞれ 28 m 及び 2.4 の変化に相当する。これらは臨床的に意味のある変化と考えられる。

4 おわりに

日本製薬工業協会 医薬品評価委員会 データサイエンス部会では、2023 年度、2024 年度に「dBM の開発」というテーマで活動を行い、臨床試験のエンドポイントとして、特にピボタ

ル試験の主要評価項目に dBM を活用するためにはどのようなエビデンスが必要になるか、という観点での事例検討をまとめ本成果物を作成した。

SV95C については、ピボタル試験の主要評価項目として用いられている、いわゆる成功事例である。他疾患での dBM の事例については、CTTI が発行した文書「Developing Novel Endpoints Generated by Digital Health Technology for Use in Clinical Trials」[28]の参照先にパーキンソン病、心不全、糖尿病、デュシェンヌ型筋ジストロフィーの事例が紹介されているので、参考にしてほしい。

2 章では、dBM のバリデーションにおいて考慮すべき事項として、dBM を用いたピボタル試験の評価項目として利用するために必要になるバリデーションについて、DiMe が公開しているチェックリストを引用して、デジタル臨床評価の関連性、ベリフィケーション、分析バリデーション、臨床バリデーションの 4 つのバリデーションステップが必要であることを説明した。これらは科学的、さらに規制上の受容性を確保するための基盤であり、検討には多くの時間とリソースを要する。そのため、早期から規制当局との対話を重ね、戦略的かつ段階的に進めることが重要である。

3 章では EMA からピボタル試験の主要評価項目として認められている SV95C を題材に、どのようなバリデーションが実施されているかを確認し、主要評価項目として受け入れられるために必要なエビデンスの具体例を内容を明らかにした。定性的エビデンス（内容的妥当性）や定量的エビデンス（正確性、再現性、構成概念妥当性、反応性、意味のある変化の閾値）について綿密な検証が行われており、実践的な知見が得られた。

2 章、3 章を通じて、読者の皆様にはピボタル試験の主要評価項目として用いることができる dBM のエビデンスについて理解を深めていただければ幸いである。

成果物を作成する段階で、dBM に関わるデバイスメーカーであるソニーネットワークコミュニケーションズ株式会社、dBM 開発プラットフォーム及び得られたデータを解析するアルゴリズムを開発する株式会社テックドクターと dBM 開発における課題を意見交換した。本 TF では、販売されている／用意されたウェアラブルデバイス等の機器に関するベリフィケーションは機器開発メーカーが示した上で提供されるものと考えていた。一方、議論の中で、加速度センサーが含まれるもので、加速度センサーの正確性の確認が製薬メーカー側で実施している事例を共有頂いた。アルゴリズム開発者からは、解析アルゴリズムを作成していく上での考え方、TF メンバーからは臨床試験での評価指標に関する考え方について双方の経験、知見を共有することができた大変有意義な機会となった。dBM の開発は、センサー技術やアルゴリズムの性能評価、臨床的有用性の検証など、多岐にわたる専門性が求められる。本報告書で説明しているとおり dBM の開発には多くのエビデンスの蓄積と長期的な取り組みが必要になることや、機器メーカーやアルゴリズムを開発するベンダーとの責任範囲を明確にして開発を進めることの重要性などをあらためて認識できた。さらに、規制当局対応でも関係者が連携して準備を進める必要があると確認する良い機会となった。意見交換会に参加いただいた皆様に感謝申し上げる。

5 参考文献

- [1] 日本製薬工業協会 医薬品評価委員会 臨床評価部会「医薬品開発におけるデジタルバ

- イオマーカー (dBM) の利活用と要件」 (2022 年 4 月)
[\[https://www.jpma.or.jp/information/evaluation/results/allotment/digital_vaio_maker_202204.html\]](https://www.jpma.or.jp/information/evaluation/results/allotment/digital_vaio_maker_202204.html) (Accessed Mar 2025)
- [2] Digital Medicine Society (DiMe) The Playbook Measurement
[\[https://playbook.dimesociety.org/dossiers/measurement/\]](https://playbook.dimesociety.org/dossiers/measurement/) (Accessed Mar2025)
- [3] CTTI Flowchart of Steps for Novel Endpoint Development [\[https://ctti-clinicaltrials.org/wp-content/uploads/2021/06/CTTI_Novel_Endpoints_Flowchart.pdf\]](https://ctti-clinicaltrials.org/wp-content/uploads/2021/06/CTTI_Novel_Endpoints_Flowchart.pdf) (Accessed Mar 2025)
- [4] Digital Medicine Society (DiMe) The Playbook Measurement, Pivotal-Primary Endpoint Checklist [\[https://drive.google.com/file/d/1k3BHeO57RUjY8xLCnfIJWxE4kaV6qzLR/view\]](https://drive.google.com/file/d/1k3BHeO57RUjY8xLCnfIJWxE4kaV6qzLR/view) (Accessed Mar 2025)
- [5] Manta C, Patrick-Lake B, Goldsack JC. Digital Measures That Matter to Patients: A Framework to Guide the Selection and Development of Digital Measures of Health. Digit Biomark. 2020 Sep 15;4(3):69-77
- [6] Jennifer C. Goldsack, Andrea Coravos, Jessie P. Bakker, Brinnae Bent, Ariel V. Dowling, Cheryl Fitzer-Attas, Alan Godfrey, Job G. Godino, Ninad Gujar, Elena Izmailova, Christine Manta, Barry Peterson, Benjamin Vandendriessche, William A. Wood, Ke Will Wang & Jessilyn Dunn. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs) Digital Medicine 2020;3(55)
- [7] Walton MK, Cappelleri JC, Byrom B, Goldsack JC, Eremenco S, Harris D, Potero E, Patel N, Flood E, Daumer M. Considerations for development of an evidence dossier to support the use of mobile sensor technology for clinical outcome assessments in clinical trials. Contemp Clin Trials. 2020;91:105962. doi: 10.1016/j.cct.2020.105962.
- [8] 日本製薬工業協会 医薬品評価委員会 臨床評価部会 「医薬品開発におけるデジタルバイオマーカー(dBM)の活用意義と実務的手順」 (2023 年 6 月)
[\[https://www.jpma.or.jp/information/evaluation/results/allotment/CL_202306_TF1.html\]](https://www.jpma.or.jp/information/evaluation/results/allotment/CL_202306_TF1.html) (Accessed Mar 2025)
- [9] FDA Digital Health Technologies for Remote Data Acquisition in Clinical Investigations [\[https://www.fda.gov/media/155022/download\]](https://www.fda.gov/media/155022/download) (Accessed Mar 2025)
- [10] FDA patient-focused drug development guidance series for enhancing the incorporation of the patient's voice in medical product development and regulatory decision making
[\[https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical\]](https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical) (Accessed Mar 2025)
- [11] FDA Patient-Focused Drug Development: Selecting, Developing, or Modifying Fit-for-Purpose Clinical Outcome Assessments [\[https://www.fda.gov/media/159500/download\]](https://www.fda.gov/media/159500/download) (Accessed Mar 2025)
- [12] 日本製薬工業協会 医薬品評価委員会 データサイエンス部会 「Patient-Focused Drug Development への患者経験情報の活用-患者の声に基づいた意思決定のために-」 (2025 年 4 月) [\[https://www.jpma.or.jp/information/evaluation/results/allotment/eo4se30000009bew-att/DS_202504_PFDD.pdf\]](https://www.jpma.or.jp/information/evaluation/results/allotment/eo4se30000009bew-att/DS_202504_PFDD.pdf) (Accessed Mar 2025)

- [13] Witt D, Kellogg R, Snyder M, Dunn J. Windows into human health through wearables data analytics. *Curr Opin Biomed Eng.* 2019;9:28-46. doi:10.1016/j.cobme.2019.01.001
- [14] Digital Medicine Society (DiMe) The Playbook, Micro-Playbook Bias
[https://docs.google.com/presentation/d/1krXccDkcyl941UxflPMaYrqJrmLyvqSPc39MpJVm49Y/edit?slide=id.gc7b34943d3_0_139#slide=id.gc7b34943d3_0_139] (Accessed Mar 2025)
- [15] Digital Medicine Society (DiMe) The Playbook, Micro-Playbook Change
[https://docs.google.com/presentation/d/19o_669XgoLGGAfSwkSGy38mX9Rw6OKnbcve5ZbKOvA/edit?slide=id.gcaebf49665_0_208#slide=id.gcaebf49665_0_208] (Accessed Mar 2025)
- [16] Doherty A, Jackson D, Hammerla N, Plötz T, Olivier P, Granat MH, White T, van Hees VT, Trenell MI, Owen CG, et al. Large scale population assessment of physical activity using wrist worn accelerometers: the UK Biobank study. *PLoS One.* 2017;12(2):e0169649.
- [17] EMA Opinions and letters of support on the qualification of novel methodologies for medicine development [<https://www.ema.europa.eu/en/human-regulatory-overview/research-development/scientific-advice-protocol-assistance/opinions-letters-support-qualification-novel-methodologies-medicine-development>] (Accessed Mar 2025)
- [18] Qualification opinion on stride velocity 95th centile as a secondary endpoint in Duchenne Muscular Dystrophy measured by a valid and suitable wearable device
[https://www.ema.europa.eu/en/documents/scientific-guideline/qualification-opinion-stride-velocity-95th-centile-secondary-endpoint-duchenne-muscular-dystrophy-measured-valid-and-suitable-wearable-device_en.pdf] (Accessed Mar 2025)
- [19] Qualification Opinion for Stride velocity 95th centile as primary endpoint in studies in ambulatory Duchenne Muscular Dystrophy studies
[https://www.ema.europa.eu/en/documents/scientific-guideline/qualification-opinion-stride-velocity-95th-centile-primary-endpoint-studies-ambulatory-duchenne-muscular-dystrophy-studies_en.pdf] (Accessed Mar 2025)
- [20] Mah JK, Korngut L, Dykeman J, Day L, Pringsheim T, Jette N. A systematic review and metaanalysis on the epidemiology of Duchenne and Becker muscular dystrophy. *Neuromuscul Disord.* 2014;24(6):482-91.
- [21] Scott E, Eagle M, Mayhew A, Freeman J, Main M, Sheehan J, et al. Development of a functional assessment scale for ambulatory boys with Duchenne muscular dystrophy. *Physiother Res Int.* 2012;17(2):101-9.
- [22] Mayhew AG, Cano SJ, Scott E, Eagle M, Bushby K, Manzur A, et al. Detecting meaningful change using the North Star Ambulatory Assessment in Duchenne muscular dystrophy. *Dev Med Child Neurol.* 2013;55(11):1046-52.
- [23] Bushby K, Finkel R, Wong B, Barohn R, Campbell C, Comi GP, et al. Ataluren treatment of patients with nonsense mutation dystrophinopathy. *Muscle Nerve.* 2014;50(4):477-87.
- [24] 臨床試験のための Patient-Reported Outcome (PRO) 使用ガイダンス
[<https://www.lifescience.co.jp/pro/article02.html>] (Accessed Mar 2025)
- [25] Mc Carthy M, Burrows K, Griffiths P, Black PM, Demanuele C, Karlsson N, Buenconsejo J, Patel N, Chen WH, Cappelleri JC. From Meaningful Outcomes to Meaningful Change Thresholds: A

- Path to Progress for Establishing Digital Endpoints. *Ther Innov Regul Sci* 2023;57(4):629-45.
- [26] FDA draft guidance for industry, FDA staff, and other stakeholders Patient-Focused Drug Development: Incorporating Clinical Outcome Assessments Into Endpoints for Regulatory Decision-Making (April 2023). [<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-focused-drug-development-incorporating-clinical-outcome-assessments-endpoints-regulatory>] (Accessed Mar 2025)
- [27] Muntoni F, Signorovitch J, Sajeev G, Done N, Yao Z, Goemans N, McDonald C, Mercuri E, Niks E, Wong B, Servais L, Straub V, de Groot I, Tian C, Manzur A, Vandenborne K, Dieye I, Lane H, Ward SJ. Minimal Detectable Changes in Functional Measures in Duchenne Muscular Dystrophy: A Study of Multiple Centers, Networks and Trial Arms. Presented at the Muscular Dystrophy Association Virtual Clinical & Scientific Conference, March 15th –18th, 2021.
- [28] CTTI Developing Novel Endpoints Generated by Digital Health Technology for Use in Clinical Trials [<https://ctti-clinicaltrials.org/wp-content/uploads/2022/03/CTTI-Digital-Health-Trials-Novel-Endpoint-Acceptance-Recommendations.pdf>] (Accessed Mar 2025)

データサイエンス部会 2024 年度継続タスクフォース 10

タスクフォース推進委員

- | | |
|------------|-------|
| ○ MSD 株式会社 | 柴田 康晴 |
| 中外製薬株式会社 | 蔵多 雅子 |

タスクフォースメンバー

- | | |
|--------------------|-----------------------------|
| ○ 株式会社三和化学研究所 | 福岡 純司 |
| ○ 日本イーライリリー株式会社 | 前河 智子 |
| ○ ファイザーR&D 合同会社 | 吉田 瑞樹 |
| 田辺三菱製薬株式会社 | 中 誠則（2023 年 4 月～2024 年 4 月） |
| ノボ ノルディスク ファーマ株式会社 | 西田 朋由 |

監修

担当副部長 土屋 悟 住友ファーマ株式会社

○：本報告書の作成に中心的に関与したメンバー

以上の資料作成にあたり，本資料の査読を実施頂いた査読担当者の諸氏に感謝いたします。