論文とETIC 19

AIリハビリは、誰にとって使いやすいのか

本人と支援者の評価差を、点数と語りから読む



Daiki Morimoto / Unknown Lab
ORCID: 0009-0009-6444-750X
Version 1.0 / 公開日 2026-11-11
DOI: 10.5281/zenodo.24122283
未査読 / 専門職としての著者確認済み
独立専門レビュー未実施 / AI支援執筆
ライセンス: CC BY 4.0

30秒で読む

脳卒中後の自宅リハビリを支えるAI搭載プラットフォームの試作版を、欧州4か国の71人が試した。患者向け画面（RGS）を45分操作した後の使いやすさの点数（SUS、0〜100）は、脳卒中経験者本人が平均58.9、支援者が75.6で、本人のほうが16.7点低かった（95%信頼区間−27.7〜−5.7）。同じ画面を触っても、利用する当人のほうが「複雑で扱いにくい」と答えたことになる。

ここから言えるのは、使いやすさを設計に戻すには、本人が実際に操作したうえでの評価と、利用を支える条件（人の関与、時間、費用）を直接調べる必要がある、という点までである。差が生じた原因、自宅で使い続けられるか、機能回復に役立つかは、この研究では決まらない。

退院後の空白と、使う人の受け止め方

原著が出発点に置くのは、退院後にリハビリの量が落ち込む「治療の空白」である。専門職の不足、待機期間、専門施設までの距離は欧州内でも地域差が大きく、リハビリを自宅へ延ばすデジタル技術に期待が集まっている。AIには、運動の出来を見て課題を調整する役割が見込まれている。

著者らは、こうした技術が定着するかどうかは、アルゴリズムの精度に加えて、使う人の受容、信頼、デジタル技能、臨床の仕事の流れとの相性に左右されると述べる。複数の部品からなるAIリハビリシステムについて、開発の早い段階で関係者の受け止め方を調べた実証研究は乏しい、というのが原著の現状認識である。

本研究が加えたものは二つある。一つは、本人、支援者、医療専門職という三つの立場を、4か国で同じ手順により調べたこと。もう一つは、点数と語りをテーマごとに並べて照合したことである。背景には著者らのグループによる系統的レビューがある。原著の記述によれば、デジタルリハビリ技術の分野でSUSの構成概念妥当性には十分な証拠があった一方、内容妥当性（項目がこの分野で測るべき内容を覆っているか）は評価されておらず、質的データと併せて解釈することが勧められた。本研究はその勧めを実行した形になる。このレビュー自体は本稿では読んでいない。

参加した71人と選ばれ方

研究はスペイン、イタリア、フランス、ルーマニアの臨床施設で行われた。参加者は71人で、脳卒中経験者本人27人、支援者17人、医療専門職27人である。本稿でいう支援者は原著のcaregiversにあたる。募集基準はformal / informalの双方を対象とし、家族だけに限らないが、その内訳は報告されていない。リハビリを処方・監督する医療専門職は、別の群として募集された。

フォーカスグループ（少人数の集団面接）は17回開かれた。当初計画は10回で、フランスとイタリアで追加開催ができたため増えた。ルーマニアは運営上の制約から本人だけを募集したので、支援者と専門職の結果は残り3か国に由来する。

本人は平均67歳（33〜86歳）、発症後1〜360か月で、回答のあった26人のうち慢性期（6か月超）が20人、亜急性期が6人だった。支援者は平均62歳（20〜81歳）、専門職は平均32歳（21〜56歳）で全員が大学卒、臨床経験は平均6.7年（標準偏差7.4）である。デジタル技能を「とても低い・基本」と自己評価した人は本人13人（1人無回答）、支援者6人、専門職4人だった。年齢、学歴、技能が群によって大きく違うことは、後の数値の読み方に関わる。

選ばれ方には注意が要る。本人は、各施設の臨床家が外来の経過観察や入院中に該当者を見つけて募集した。目的に沿って選ぶ抽出である。重い認知機能低下（MoCA 19点未満に相当）、集団討論が難しい言語障害、重い視覚障害、麻痺側上肢の強い痙縮がある人は除かれた。判断は施設スタッフが診療歴と日常の診察から下しており、検査を一律に実施してはいない。体験時にRGSを操作できない人も実務上除かれた。辞退者や選外の人の記録は系統的に残されておらず、どんな人が参加できなかったのかは分からない。著者ら自身、デジタル機器への準備が整った人に偏った可能性を認めている。募集とデータ収集の暦上の期間は、本文には見当たらなかった。

表1 三つの群と、それぞれが触れたもの

群	人数（FG回数）	国	体験	SUSで評価した対象
脳卒中経験者本人	27人（6回）	4か国	RGSapp・RGSclinicを45分操作	RGS
支援者（caregivers）	17人（5回）	スペイン・イタリア・フランス	本人群と同じ	RGS
医療専門職	27人（6回）	スペイン・イタリア・フランス	RGSの操作に加え、MIMSの実演15〜20分	MIMSとRGSを合わせた全体

出典: 原著2.2節、2.6節、表1から作成。FG回数は表1の内訳を数えた。読み方: 同じ対象を評価したのは上の2群だけである。専門職の体験は、方法の節では「45分の操作と実演」、結果と限界の節では「15〜20分の実演後に評価」と書かれている。

試作システムの中身と、1回のセッション

評価の対象は、欧州プロジェクトAISNが開発中のプラットフォームである。患者側の入口がRGS（Rehabilitation Gaming System）で、運動と認知の課題をゲーム形式で行う。RGSは課題の出来を監視し、経験則とベイズ的な計算によって速度や動かす範囲などの難易度を自動で調整する。臨床家側の入口がMIMSというウェブ上の管理画面で、患者情報の管理、運動メニューの処方、実施状況と成績の確認に使う。

表2 部品ごとの扱い

部品	役割	この研究での扱い
RGSapp	スマートフォン・タブレットの訓練アプリ	操作した
RGSclinic	タッチ画面と光学センサーを備えた施設用端末	操作した
RGSweb	ブラウザとウェブカメラによる自宅訓練	紹介のみ
RGSwear	スマートウォッチによる麻痺側上肢の見守り	紹介のみ
MIMS	臨床家の管理画面（処方は手動）	専門職に対話形式で実演
運動推薦CDSS・回復予測モデル	計画中のAIモジュール	実演と模型画面のみ。使用なし

出典: 原著2.4〜2.6節から作成。読み方:「AI搭載」と呼ばれるシステムのうち、参加者が体験したAI的な処理はRGSの難易度調整である。推薦と回復予測は、意見を聞くための提示にとどまる。

セッションは臨床施設内で行われ、全体で最長2.5時間だった。前半が操作体験で、進行役は最初だけ手助けし、その後は観察に回ってつまずきを見た。15分の休憩をはさみ、後半は台本に沿った集団討論を行い、討論の最後にSUSへ回答した。途中でやめた人はいない。これは1回の手続きを全員が終えたという意味で、自宅で何週間も使い続けた記録ではない。

SUSという物差し

SUS（System Usability Scale）は、Brookeが手早く使える尺度として提案した10項目の質問紙である[2]。各項目に5段階で同意の程度を答え、回答を項目ごとに0〜4点へ直して合計し、2.5を掛けて0〜100の得点にする。Brookeの出発点は、使いやすさは製品単独では決まらず、誰が、どんな課題を、どんな環境で行うかを定めて初めて評価できる、という立場にある。SUSが返すのは使った人の主観的な全体評価で、個々の項目を独立した尺度として読むことは想定されていない。

原著は総得点をSauroとLewisの慣用の等級（A+〜F）に当てはめ、68点を平均（C）としている。68は臨床上の合格線ではない。また、Brookeは使用の直後、振り返りの討論より前に答えてもらうことを想定していた。本研究の回答は集団討論の後である。この順序が点数を動かしたかどうかは原著で検討されておらず、分からない。これは読むときの留意点としてUnknown Labが挙げるものである。

同じ画面への点数は、本人のほうが低かった

表3 SUS総得点と群間比較

群（評価対象）	平均±標準偏差	等級
本人 27人（RGS）	58.9±16.5	D
支援者 17人（RGS）	75.6±13.3	B
専門職 27人（MIMSとRGS）	72.7±12.8	B−

比較	平均差	95%信頼区間	p値
本人−支援者（同じ対象）	−16.7	−27.7〜−5.7	0.002
本人−専門職（対象が異なる）	−13.8	−23.5〜−4.1	0.003
支援者−専門職（対象が異なる）	2.9	−7.1〜12.8	0.76

出典: 原著3.3節と表2から作成。3群の差はWelchの分散分析でF[2, 41.0] = 8.07、p = 0.001、η² = 0.21。1項目が無回答だった2人は残り9項目から按分されている。読み方: 同じ評価対象についての群間差を示すのは比較の1行目である。全体平均68.1±16.0（C）は異なる対象への評価を混ぜた値なので、システムの出来を代表する数字としては使えない。

本人と支援者の差は16.7点で、信頼区間は約6点から28点まで幅がある。本人27人のうち13人は51.7未満、つまり最下位の等級Fに入った。

項目別（1〜5の同意度）では、本人は支援者より「複雑だ」（2.93対1.65）、「扱いにくい」（2.85対1.76）、「一貫していない」（2.63対1.76）に強く同意し、「頻繁に使いたい」（3.41対4.47）と「自信を持って使えた」（3.56対4.24）は低かった。一方、使いやすさ（3.63対4.06）、覚えやすさ（3.59対3.76）、使う前に学ぶ必要（2.78対2.82）には差が見られなかった。技術的な手助けの必要は本人3.37、支援者2.59で、有意ではない（p = 0.10）。項目別の比較は語りと照合するための補助であり、SUS本来の読み方からも総得点ほどの重みは置けない。

支援者の点数の意味を確認しておく。支援者が答えたのは、自分で同じ課題を操作した感想である。本人の体験を代わりに見積もった値とは読めない。本人と支援者を親子や夫婦の組として対応させた解析も報告されていない。この差が示すのは、同じ画面でも操作する人によって評価が大きく違ったという観察までで、家族が本人の苦労を見誤るという話へは広げられない。

デジタル技能との関係は意外な形で出た。全員を混ぜると、技能の自己評価が高い人、学歴が高い人、若い人ほど点数が高かった（技能 $p = 0.008$、学歴 $p = 0.012$、年齢 $\tau = -0.20$）。ただし専門職は若く、全員が大学卒で技能も高いため、著者らはこの関連を群の構成による交絡と解釈している。本人群の中だけで見ると、技能（$F[2, 6.0] = 0.62$、$p = 0.567$）、学歴（$p = 0.772$）、年齢（$\tau = 0.01$、$p = 0.950$）のいずれも点数と関連しなかった。支援者では年齢が高いほど点数が低く（$\tau = -0.49$、$p = 0.008$）、専門職では技能が中程度の人が高い人より15.0点低かった。

本人群で関連が検出されなかったことは、技能が無関係であることや、高齢でも支障がないことの証明にはならない。27人では小さな関連を見逃しうるし、操作できない人は参加していない。量的解析はすべて探索的で、多重比較の補正はなく、著者らはp値を記述的な目安として扱っている。

語りから取り出された六つのテーマ

討論は録音され、各国の言語のまま文字に起こされた。各施設でその言語を母語とする研究者1人がNVivoでコードを付け、暫定コード、要約、代表的な発言を英訳して共通のコード表を作り、施設をまたぐ合意会議で解釈の食い違いを解消した。独立した2人が別々にコードを付ける手順は踏んでいない。テーマごとに何人が語ったかも報告されていないので、引用の数から多数意見を割り出すことはできない。

表4 六つのテーマ

テーマ	語られた内容
1 操作しやすさと臨床的な有用性	専門職と支援者は「直感的」と評した。本人の語りは、活動的でいるための運動の価値に向いた
2 個別化とフィードバック	自分の状態に合った課題と、その場での手応えが求められた。専門職は、画面の時間が日常の活動を押しのける懸念も述べた
3 技術的な信頼性と信頼	本人は、止まったときや通信が切れたときに一人で対処できるかを案じた。支援者はデータの安全と、目で見て試せる説明を、専門職は自動出力への頼りすぎを挙げた
4 デジタル技能の格差	高齢の人や端末に不慣れな人は、自宅で使うのに継続的な手助けが要ると見込まれた
5 人の仲介と治療関係	3群とも完全自動のモデルを退け、療法士による励ましと安心を求めた
6 仕事の流れと経済・制度	専門職は遠隔での見守りを評価しつつ時間を最大の制約に挙げ、支援者は費用と償還を問うた

出典: 原著3.4節から作成（要約は筆者）。読み方: 並びは原著の番号順で、重要度や語った人数の順ではない。

印象に残る発言をいくつか挙げる（いずれも拙訳）。フランスの本人は「クラッシュしたら、どうすればいいのか」と問い、イタリアの支援者は、こうした話を聞くと「人との接触をなくすためのものだと思ってしまう」と語った。スペインの専門職は「最大の制約は時間だ。患者に会い、データを集め、プラットフォームに手をかける時間」と述べ、フランスの専門職は「ゲームを日常生活の活動へ翻訳してほしい」と求めた。

テーマ3の語りは、起こるかもしれない不具合への予想である。セッション中に実際に起きた不具合は系統的に記録されておらず、故障の頻度を示すデータはない。テーマ5も参加者の希望の記述であり、将来のAIが人の役割を代われないことの証明ではない。

点数と語りを並べると見えるもの

この研究の読みどころは、テーマごとに語りと数値を横に並べ、同じ方向を指す（一致）、逆を指す（食い違い）、互いに別の側面を補う（補完）のどれにあたるかを判定した統合表にある。

表5 語りと点数の照合（要約）

テーマ	語り	対応する数値	判定
1 操作しやすさ	支援者・専門職は「直感的」	本人58.9、支援者75.6。差は主に複雑さと扱いにくさの項目	食い違い
2 個別化	難易度の調整を歓迎、画面時間への懸念	SUSに該当項目なし	補完
3 信頼性	本人が不具合と一人での対処を案じる	自信3.56対4.24、不一貫2.63対1.76	一致
4 デジタル技能	大きな障壁として語られる	本人群内では技能・学歴・年齢と関連なし	食い違い
5 人の仲介	全群が療法士の関与を求める	手助けの必要は本人3.37、支援者2.59、専門職2.30	本人は一致、支援者は補完
6 時間と費用	時間、費用、償還が障壁	SUSに該当項目なし（専門職72.7）	補完

出典: 原著表3と3.5節から作成。読み方: 判定は原著の著者による。語りが点数の原因であることを示す表ではない。

読み筋は三つある。第一に、周囲の人が語った「直感的」という印象は、本人の点数には表れなかった。第二に、技能不足は討論で強く心配されたのに、本人の点数とは結びつかなかった。著者らは、技能への懸念は自宅で一人で使う場面の予想に向いており、見守られた体験の直後に付けた点数とは向き先が違った可能性を挙げ、より大きな標本での検証を求めている。第三に、人の関与、臨床家の時間、償還という障壁は、SUSの質問が扱わない領域から出てきた。支援者はRGSを高く評価しながら、臨床家の

関与をなお求めていた。点数だけを集めていれば、この三つはどれも見えなかった。

著者らはこれらのパターンを記述的なものと位置づけ、この設計ではテーマと点数の因果関係は言えないと明記している。

著者の結論と、言える範囲

著者らの結論は慎重である。現時点では、支援なしに本人が自宅で自立して使うことは現実的に見えない。3群とも、AIによる運動の提供と臨床家の見守りを組み合わせるハイブリッド型を支持した。得られたのは関係者の認識と早期の使いやすさの手がかりで、臨床成果ではない。結果は画面と操作の改良、高齢の利用者向けの簡略な練習モジュール、推薦機能と予測モデルの組み込み方に反映され、ハイブリッド型は今後の多施設ランダム化比較試験で検証されるという。

因果について整理する。役割は無作為に割り当てられたものではなく、本人と支援者は年齢、学歴、麻痺の有無、国の構成（ルーマニアは本人のみ）など多くの点で異なる。差が身体機能によるのか、疲労によるのか、機器への慣れによるのかは切り分けられない。体験は見守りのある1回きりで、難易度調整の有無を比べた実験でもない。機能回復、生活の変化、継続利用、費用、臨床家の労働時間はいずれも測られていない。

限界

表6 示したこと、示していないこと、次に要る比較

論点	示したこと	示していないこと	次に要る比較
本人と支援者の差	同じRGSで16.7点の差	差の原因、組になった本人と支援者の違い	本人と支援者を組にした評価
専門職の点数	管理画面を含む全体への第一印象	本人・支援者との優劣	対象と体験時間をそろえた評価
デジタル技能	本人群内で関連を検出せず	技能が無関係であること	大きな標本、自宅で一人で使う場面
信頼性	不具合への不安の語り	実際の故障頻度	不具合と対処の系統的な記録
対象者	操作でき、討論に参加できた人の評価	重い認知・言語・視覚の困難がある人での評価	除かれた層での検討、SUSの適用可能性の確認
効果	早期の使いやすさの手がかり	回復、継続、費用、時間	長期の自宅利用、ランダム化比較試験

出典: 原著4.4節、4.5節と本稿の整理から作成。読み方: 右端の列のうち、大きな標本での検証とランダム化比較試験は原著が挙げたもので、それ以外はUnknown Labの整理である。

このほか原著は、臨床施設で臨床スタッフが進行したため好意的な回答に傾いた可能性、最長2.5時間の手続きによる疲労が本人の評価に影響した可能性、国ごとの人数が1区分3〜13人と少なく国別比較に耐えないこと、コード付けが各施設1人だったことを挙げている。SUS自体も、デジタルリハビリ技術での内容妥当性と、認知やコミュニケーションに困難がある人への適用可能性が確かめられていない。

透明性に関わる記載も確認した。著者らは利益相反なしと申告し、研究はHorizon Europe（助成番号101057655）の資金を受け、各施設で倫理承認を得ている。進行役は、自分たちがアルゴリズムの開発者ではないと参加者に伝えていた。データは機密保持などの理由で公開されず、合理的な請

求と施設の承認が条件となる。ソフトウェアの版、コード、保守体制は、再現できる形では本文に示されていない。参加者は研究計画の設計、分析、執筆には関与していない。

本稿の確認範囲にも限界がある。出版社の本文、全3表、宣言、参考文献は確認したが、補足資料は取得できていない。項目別の個票や追加解析を独立に再計算した解説ではなく、本文に報告された数値に基づいている。

ETIC 2.0から読む

ここからはUnknown Labの解釈である。ETIC 2.0は、AIを含むケアを、本人の意味ある作業と参加を中心に置いて設計・点検するための設計モデルで、人、環境、機器、データ、AI、ケアの調整、統治といった層の関係を見る[3]。臨床的な有効性を検証した尺度とは位置づけられておらず、本研究がETICを検証したわけでもない。

その前提で読むと、評価する人の選び方が設計の判断を変えうる点がまず目に入る。支援者の点数（75.6、等級B）と、支援者・専門職の「直感的」という語りだけを集めていれば、RGSは操作上の大きな課題のない画面として次の段階へ進んだかもしれない。本人27人が自分で操作して点数を付けたから、複雑さと扱いにくさが設計課題として残った。本人を中心に置くというETICの考え方は、この評価への参加のさせ方に具体化できる。

SUSの外から出てきた障壁の並びも示唆的である。療法士との関係、臨床家の時間、償還は、画面の出来とは別の層にある条件で、技術を支援関係や制度の中で見るという見方と整合する。ただし確かめられたのは、参加者がそう語ったことまでである。それらの条件を整えれば利用が続くかどうかは試されていない。

本研究が測っていないものは、ETICの関心からは大きい。専門職は、画面に向かう時間が日常生活の活動、人とのつながり、余暇を押しのけることを懸念し、ゲームを日常の活動へ「翻訳」することを求めた。訓練が本人にとって意味のある作業や参加につながるかという問いは、こうした発言として現れただけで、評価項目にはなっていない。使いやすい画面が生活の中の作業へ結びつくかは、この研究の範囲の外に残っている。

この試作版の改善に向けては、既存のSUSと面接に操作場面の記録を組み合わせる案が考えられる。本人の手がどこで止まり、誰が何を手伝ったかを記録し、点数と語りに照らす。本研究で不具合や手助けの実際が系統的に記録されていなかったためである。これはUnknown Labの未検証の提案である。観察記録を加えても設計上の問題の発見や改善判断が増えず、記録負担だけが増すなら、手順を縮小するか採用を見送る。追加記録の価値も、既存の評価手順との比較で確かめる必要がある。

次に読む論文

1回のセッションで分かるのは入口までである。自宅で使い続ける段階では、償還、組織の体制、担当者の時間が前面に出てくる。その条件を扱った研究として、te Boekhorstら（2025）を挙げる[4]。オランダにおける脳卒中後の運動リハビリ向け在宅技術の導入を取り上げ、22件の半構造化面接を、技術導入の複雑さを整理するNASSSという枠組みで分析した研究で、原著も考察で引用している。本稿では書誌、要旨、方法の冒頭を確認したにとどまり、結果の全文と補足は読んでいないため、数値の根拠には使っていない。本連載ではまだ取り上げていない論文である。

結論

AIリハビリは誰にとって使いやすいのか。本研究の試作RGSについて言えば、支援者には使いやすいと評価され、利用する脳卒中経験者本人にはそれより低く評価された。同じ画面を45分操作した後のSUSは本人58.9、支援者75.6で、本人27人のうち13人は最下位の等級にとどまった。

この差を設計に戻すには、周囲の人の印象で代えずに本人が実際に触れた評価を取り、あわせて人の関与、時間、費用といった利用を支える条件を直接調べる必要がある。点数と語りを並べる方法は、そのための既存の手段として使える可能性がある。差の原因、自宅での継続、機能回復への効果は未検証であり、改良後のシステムを対象とするランダム化比較試験と、本人と支援者を組にした比較が次の材料になる。

参考文献

1. Abi Saab P, Mandigout S, Danesin L, et al. Stakeholder Perspectives and Perceived Usability of an AI-Enabled Stroke Rehabilitation Platform: A Multicentre Mixed-Methods Focus Group Study With Stroke Survivors, Caregivers, and Clinicians. Health Expectations. 2026;29(6):e70947. https://doi.org/10.1111/hex.70947

2. Brooke J. SUS: A 'Quick and Dirty' Usability Scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, eds. Usability Evaluation in Industry. Taylor & Francis; 1996:189–194. https://doi.org/10.1201/9781498710411-35

3. Morimoto D. エコテクノロジカル・インテグラティブ・ケア(ETIC)2.0, Version 2.1改訂版 - AI時代の作業療法とケア生態系設計. Unknown Lab; 2026. https://doi.org/10.5281/zenodo.21848456

4. te Boekhorst KE, Kuipers SJ, Ribbers GM, Cramm JM. The complexity of home-based rehabilitation technology implementation for post-stroke motor rehabilitation in the Netherlands. BMC Health Services Research. 2025;25:21. https://doi.org/10.1186/s12913-024-12044-2