

メンタルヘルスアプリの 評価指標と留意点に関する概観



関西学院大学大学院文学研究科 博士課程前期課程
高階光梨 (たかしな ひかり)

Profile

2020年より同志社大学心理臨床センター臨床研究員, 2022年より国立精神・神経医療研究センター認知行動療法センター研究生。専門は臨床心理学。著書に『代替行動の臨床実践ガイド: 「ついやってしまう」「やめられない」の〈やり方〉を変えるカウンセリング』(分担執筆, 北大路書房), 『遠隔心理支援スキルガイド: どこへでもつながる援助』(分担執筆, 誠信書房) など。

メンタルヘルスケア×アプリの現在

情報通信技術 (Information and Communication Technology: ICT) の発展にとともに、手のひらサイズのデバイスで、いつでもどこでも、大量の情報にアクセスできるなど、私たちの生活はますます便利になっている。メンタルヘルスケアの領域でも、このような機器や技術の普及により、パソコン、スマートフォン、仮想・拡張現実等を活用した新しい支援方法が模索されている¹。

2010年以降、メンタルヘルスケアを目的としたアプリを含む、モバイルヘルスアプリ (mobile health apps) に関する研究が増加しており、関心が高まっている²。例えば、メンタルヘルスケアを目的としたアプリには、ゲームを進めることで認知行動療法プログラムを受

けられるアプリ、チャットボットとやり取りをしながら思考や感情を整理するアプリ、動画を見ながらマインドフルネス、瞑想、リラクゼーションを実施するアプリ、落ち込みや抑うつ気分などの症状や服薬を記録するアプリなどがある。国内においても、モバイルヘルスアプリをはじめとして、ICTを用いたメンタルヘルスケアのサービス (メンタルヘルステック) を扱う多様な事業者がある (図1)。

うつ病や不安症などの一般的な精神疾患に対して、アプリがどの程度効果があるのかに関する検証も行われてきているが (例えば文献⁴)、未だアプリの質や内容は玉石混濁であり、多くのアプリは有効性が未確認であることが指摘されている。そのような中には、誤った情報を提供する、効果が示されていないのに効果を主張



図1 国内メンタルヘルステックカオスマップ2022³ (emol株式会社の許可を得て転載)

する、安全性が確保されていない、個人のデータを不適切に開示する、実臨床での利用について十分な検討が進んでおらず有効性の不明な新しい技術を用いているなど、有害となり得る可能性のあるアプリもあり⁵、場合によっては利用者に健康被害をもたらす可能性も懸念される。しかしながら、自身のメンタルヘルスケアを目的とする利用者にとって、口コミやアプリサイトによる評価のみでは、それらのアプリの良し悪しを見分けることは困難であることが指摘されている⁶。そのため、有用で安全かつ効果的なアプリと、有用でなく危険かつ／または効果的ではないアプリを見分け、今の自分にとって必要なアプリを選択するための客観的な評価が可能なツールが必要となる。

アプリ選択のための評価指標

このような課題を解決するために、アプリを客観的に評価するための様々なツールが開発されている。最近のレビューによると、アプリの評価に使用するための指標として、2018年時点の系統的なレビューで45種類⁷、2020年時点では49種類の尺度が開発されていることが明らかになっている⁸。以下では、これらの評価指標について簡単に整理を行う。

これらの尺度は、大きく「アプリのユーザビリティ（使いやすさや分かりやすさ）を評価するもの」と「アプリの質を評価するもの」に分類することができる。ユーザビリティを測る尺度は23種類あり⁹、すべての尺度はアプリの開発者を使用者として想定している。これらの尺度の中で、最も使用されている尺度は「システムユーザビリティスケール（System Usability Scale :SUS）」である¹⁰。SUSはVT100というビデオ表示端末の開発にあたり、ユーザビリティを測定するために作成された10項目からなる尺度であり、アプリ評価のために作成されたものではない。評価項目には「このシステムはしょっちゅう使いたくなるだろう（I think that I would like to use this system frequently）」「このシステムは必要以上に複雑だ（I found the system unnecessarily complex）」などが含まれる。簡便

に使用できるために広く使われているが、他のユーザビリティを評価する尺度との相関が低く、むしろ利用者の嗜好と強く関連していることが指摘されていることに留意が必要である¹¹。

アプリの質を評価する尺度は、さらに「特定の対象や使用目的のアプリの質を測定する尺度」と「アプリ全般の質を測定する尺度」に分類することができる。特定の対象や使用目的のアプリを測定する尺度は7種類あり、特定の対象や使用目的として、行動変容のアプリを評価するもの、フィットネスアプリを評価するもの、出産予定日計算アプリを評価するものなどがある。これらの尺度は対象が限定的であるために、ほとんどが単一の研究のみで用いられている。アプリ全般の質を測定する尺度は18種類あり、そのなかで、「モバイルアプリ評価スケール（Mobile Application Rating Scale: MARS）」が最もよく用いられている¹²。MARSはアプリの質を評価するために開発された、23項目からなる尺度である。信頼性と妥当性についても十分に検討されており¹³、英語、イタリア語、スペイン語、ドイツ語、フランス語、アラビア語等を含む翻訳版も利用可能である。2022年には日本語版の作成と信頼性・妥当性の検証も行われている^{14,15}。MARSによる評価は、アプリの客観的品質、アプリの主観的品質、知覚的なインパクトについて行われ、客観的品質には「エンゲージメント（愛用度）」「機能性」「見た目・デザイン性」「情報」が含まれる。これらの項目に対する評価を得点化し、平均スコアを算出することでアプリを評価する。得点化可能なアプリ品質を評価する尺度の中で、唯一、日本語版が開発されており、信頼性・妥当性も確認されていることから、アプリの開発者や研究者がアプリの質の評価を行う場合に、有用な指標となるであろう。なお、MARSには利用者版（user version of MARS: uMARS）もあり、uMARS日本語版も作成されている¹⁶。

このように、多様なアプリの評価指標が開発されているが、医療技術評価に単体で使える尺度がないことや⁷、利用者が自身の状態や嗜好に合わせてアプリを選択するために使用できる

尺度が限られていることが指摘されており、利用者を対象とした、より包括性の高い尺度の開発が望まれていた。アメリカ精神医学会が提案している「アプリ評価モデル (App Evaluation Model) ¹⁷」は、これらの課題を改善した新たな評価指標である。アプリ評価モデルでは、評価を①アクセスと背景情報、②プライバシーとセキュリティ、③臨床的根拠、④ユーザビリティ、⑤治療の目標を見据えたデータの共有の5つのステップで行う (図2)。アプリ評価モデルの特徴として、アプリの開発者や研究者が指針とするだけでなく、患者や臨床家、セルフケアを行いたい者など、アプリの利用者が自身で有用で安全かつ効果的なアプリを選択するために必要な情報を収集するためのツールとして使用することが想定されている。遠隔心理支援スキルガイド¹⁹では、アプリ評価項目の評価内容および評価サンプルが日本語で参照可能である。

アプリ評価モデルの評価内容は利用者がアプリを選択する際に情報に基づいた意思決定を行うために有用なモデルを提供しているが、利用者が検討しているアプリをすべて評価しようとすると、その評価項目の数や内容に圧倒されてしまうことが懸念される。アプリ評価モデルを一般向けに、機能的かつ実用的にするために、「105の客観的な質問 (105 objective questions)」が作成されている¹⁸。本指標はアプリ評価モデルの内容を回答が2値 (例: はい・いいえ) または数値で回答出来るようにアップデートしたものであり、最終的にはデータベース化するこ

とが想定されている。実現すれば、データベース上で利用者がアプリを選択する際に必要な情報を集約し、データ駆動型で意思決定を行うことが出来るのである。それにより、利用者は情報収集をする手間を省略して情報に基づいた意思決定を行い、自分のニーズに合った、有用で安全かつ効果的なアプリを選択できるようになることが期待される。

アプリ評価指標を活かす

このように、アプリのユーザビリティや質を評価するための指標は、アプリの開発者、研究者、利用者が使いやすく、有用で安全かつ効果的なアプリを開発・選択する指針となるであろう。アプリを用いたメンタルヘルスケアはスティグマなどにより医療機関に繋がれていない者のファーストステップとして、時間的・地理的・金銭的な制約でこれまでメンタルヘルスケアサービスを受けられなかった者の新たな選択肢として、訓練を受けた実施者の不足などの問題を解消する手段の一つとして、大勢の人を対象とした予防プログラムの普及・実装戦略として、などこれまでメンタルヘルスケアが抱えていた課題を解決する有効な手段となり得る。その一方で、現在はアプリの質や内容は玉石混濁であり、健康被害が懸念されるものも存在している。日本語で利用できるアプリ評価ツールは限られているものの、上記で紹介したような指標を用いることで、客観的にアプリのユーザビリティや質を評価することが可能になり、ひい

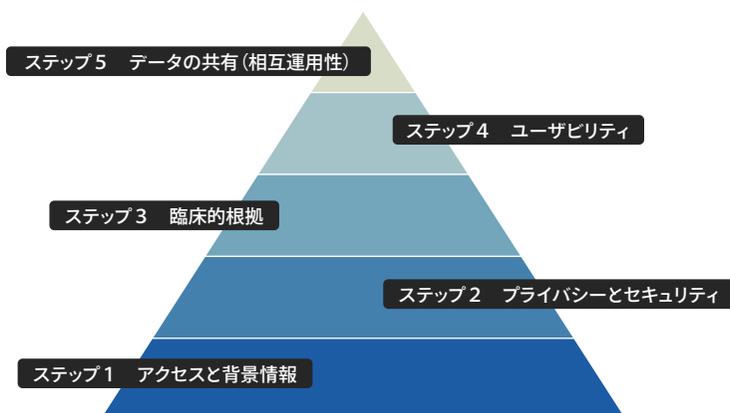


図2 アメリカ精神医学会アプリ評価モデル¹⁷の5つのステップ (文献18を参考に作成)

てはより質の高いアプリの開発や利用が促されることが期待される。

文献・注

- 1 横光健吾・高階光梨・山本哲也 (2022) 「遠隔心理支援における認知行動療法実践の基礎と課題」『認知行動療法研究』48, 135-144.
- 2 村中誠司・竹林由武 (2021) 「遠隔心理支援 (Telepsychology) におけるこれまでの検討課題: Structural Topic Model によるアブストラクト解析」『認知行動療法研究』47, 127-138.
- 3 emol 株式会社 (2022) 「『国内メンタルヘルステックカオスマップ 2022年版』を公開!」『PR Times』<https://prtimes.jp/main/html/rd/p/000000013.000043787.html> (2022年7月27日)
- 4 Miralles, L., Granell, C., Díaz-Sanahuja, L., Van Woensel, W., Bretón-López, J., Mira, A., Castilla, D., & Casteleyn, S. (2020) Smartphone apps for the treatment of mental disorders: Systematic review. *JMIR Mhealth and Uhealth*, 8(4), e14897. <https://doi.org/10.2196/14897>
- 5 高階光梨・鈴木ひかり・白塚龍太郎・大橋佳奈・宮下太陽・横光健吾 (2021) 「日本における抑うつ症状に対する心理学的支援を目的としたスマートフォン用アプリケーション・プログラムのレビュー」『認知行動療法研究』47, 1-10.
- 6 Larsen, M. E., Huckvale, K., Nicholas, J. et al. (2018) Using science to sell apps: Evaluation of mental health app store quality claims. *npj Digit. Med.*, 2, 18. <https://doi.org/10.1038/s41746-019-0093-1>
- 7 Moshi, M., Tooher, R., & Merlin, T. (2018) Suitability of current evaluation frameworks for use in the health technology assessment of mobile medical applications: a systematic review. *Int J Technol Assess Health Care*, 34, 464-475. <https://doi.org/10.1017/S026646231800051X>
- 8 Azad-Khaneghah, P., Neubauer, N., Cruz, A. M., & Liu, L. (2021) Mobile health app usability and quality rating scales: A systematic review. *Disabil Rehabil Assist Technol*, 16, 712-721. <https://doi.org/10.1080/17483107.2019.1701103>
- 9 文献8の評価時点の数であり、現時点ではさらに開発が進んでいる可能性もあることに留意が必要である。以下で出てくる開発されている尺度の数に関する記述についても同様のことに留意されたい。
- 10 Brooke, J. (1996) SUS: A 'quick and dirty' usability scale. In P. W. Jordan, B. Thomas, L. I. McClelland, & B. Weerdmeester (Eds.), *Usability evaluation in industry* (pp.189-194). London: Taylor & Francis.
- 11 Drew, M. R., Falcone, B., & Baccus, W. L. (2018) What does the System Usability Scale (SUS) measure? Validation using think aloud verbalization and behavioral metrics. *Design, User Experience, and Usability: Theory and Practice*, 10918, 356-366.
- 12 Stoyanov, S. R., Hides, L., Kavanagh, D. J., Zelenko, O., Tjondronegoro, D., & Mani, M. (2015) Mobile app rating scale: A new tool for assessing the quality of health mobile apps. *JMIR Mhealth Uhealth*, 3(1), e27. <https://doi.org/10.2196/mhealth.3422>
- 13 Terhorst, Y., Philippi, P., Sander, L. B., Schultchen, D., Paganini, S., Bardus, M., et al. (2020) Validation of the Mobile Application Rating Scale (MARS). *PLoS ONE*, 15(11), e0241480. <https://doi.org/10.1371/journal.pone.0241480>
- 14 Yamamoto, K., Ito, M., Sakata, M., Koizumi, S., Hashisako, M., Sato, M., Stoyanov, S. R., & Furukawa, T. A. (2022) Japanese Version of the Mobile App Rating Scale (MARS): Development and Validation. *JMIR Mhealth Uhealth*, 10(4), e33725. <https://doi.org/10.2196/33725>
- 15 日本語版MARSは文献14の付録よりダウンロード可能である。
- 16 Shinohara, Y., Yamamoto, K., Ito, M., Sakata, M., Koizumi, S., Hashisako, M., Sato, M., Wannous, M., Stoyanov, S. R., Nakajima, J., & Furukawa, T. A. (2022) Development and validation of the Japanese version of the uMARS (user version of the mobile app rating system). *International Journal of Medical Informatics*, 165. <https://doi.org/10.1016/j.ijmedinf.2022.104809>
- 17 American Psychiatric Association (n.d.) *The App Evaluation Model*. American Psychiatric Association. <https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/the-app-evaluation-model> (2022年7月25日)
- 18 Lagan, S., Aquino, P., Emerson, M. R., Fortuna, K., Walker, R., & Torous, J. (2020) Actionable health app evaluation: Translating expert frameworks into objective metrics. *npj Digital Medicine*, 3, 100. <https://doi.org/10.1038/s41746-020-00312-4>
- 19 前田正治・桃井真帆・竹林由武 (編著) (2020) 『遠隔心理支援スキルガイド: どこへでもつながる援助』誠信書房

謝 辞

関西学院大学 文学部 教授 佐藤寛先生, 徳島大学大学院 社会産業理工学研究部 准教授 山本哲也先生, 人間環境大学総合心理学部 講師 横光健吾先生には本記事の執筆にあたり貴重なご助言をいただきました。この場を借りて御礼申し上げます。

*COI: 本記事に関連して、著者はアプリ評価モデルおよび105の客観的な質問の日本語版の作成者である。しかし、これらの指標の販売等により金銭的な報酬は受け取っていない。また、105の客観的な質問の日本語文化適応版およびデータベース作成にむけて、クラウドファンディングを通じて寄付を受けたが、本記事の執筆にあたり寄付者は一切関与していない。