



RIETI Discussion Paper Series 16-J-059

人工知能で、人のこころは癒せるか？：

人工知能(自然言語処理)フィードバック機能搭載型のインターネット認知行動療法(iCBT-AI)の抑うつ者に対する世界初の効果検証(無作為統制試験)

宗 未来

慶應義塾大学

関沢 洋一

経済産業研究所

竹林 由武

福島県立医科大学



Research Institute of Economy, Trade & Industry, IAA

独立行政法人経済産業研究所

<http://www.rieti.go.jp/jp/>

人工知能で、人のこころは癒せるか？：人工知能(自然言語処理)フィードバック機能搭載型のインターネット認知行動療法 (iCBT-AI) の抑うつ者に対する世界初の効果検証 (無作為統制試験)¹

宗 未来 (慶應義塾大学医学部精神神経科学教室)
関沢 洋一 ((独) 経済産業研究所)
竹林 由武 (福島県立医科大学)

要 旨

背景と方法：増え続けると予測されるうつ病への対応策として、インターネットを使った自習式の認知行動療法 (iCBT) による介入が期待されている。しかし、現在利用可能な iCBT では、短期的な抑うつは改善しても、効果が長期に持続しない、脱落率が高い、社会機能の改善につながらないといった課題が残されている。近年、AI 技術の一領域である「自然言語処理技術 (NLP: natural language processing)」の進歩により、この技術を iCBT に応用して、実施者に共感を示したり適切なアドバイスを行ったりする iCBT-AI が開発されている。本研究では、通常の iCBT 群、iCBT-AI 群、待機群の 3 群の間で、どの群が最もうつ症状の軽減効果が大きいかをランダム化比較試験によって比較した。主要評価指標として代表的なうつ評価指標である PHQ-9 を用いた。

結果：iCBT-AI 群では、iCBT 群に比べてエクササイズからの脱落は有意に低かった ($p<0.005$)。全データでの結果では、通常の iCBT 群では、待機群と比べて、介入期間終了直後にうつ症状が改善する傾向 ($p=0.05$) が認められ、3 か月後のフォローアップでは有意に低くなったが ($p=0.01$)、iCBT-AI 群では有意な改善は認められなかった。一方で、PHQ-9 の得点が 10 点以上 (大うつ病性障害水準に相当) の基準を満たす者の割合は、介入期間終了直後には通常の iCBT 群、iCBT-AI 群ともに待機群に対して有意差がなかったが、3 か月後のフォローアップでは、iCBT-AI 群においてのみ、この基準を満たす者の割合が低い傾向 (改善傾向) が認められた (オッズ比 0.67; $p=0.08$)。PHQ-9 で 10 点未満の軽症うつ者に限った下位分析でみると、その減少は有意 (オッズ比 0.35; $p=0.02$) であった。以上から、iCBT-AI は非 AI 型 iCBT に比べて短期的にはマイナス効果が予想される一方で、長期的には非 AI 型には認められない将来の重症抑うつ者を減らす可能性が示唆された。我々の知る限り、諸外国でもこのような iCBT-AI の効果に関する文献報告は認めず、更なる検証が期待される。

キーワード：うつ、人工知能、インターネット認知行動療法、自然言語処理技術

JEL classification: I10, I31

RIETI ディスカッション・ペーパーは、専門論文の形式でまとめられた研究成果を公開し、活発な議論を喚起することを目的としています。論文に述べられている見解は執筆者個人の責任で発表するものであり、所属する組織及び (独) 経済産業研究所としての見解を示すものではありません。

¹ 本稿は、独立行政法人経済産業研究所におけるプロジェクト「人的資本という観点から見たメンタルヘルスについての研究 2」の成果の一部である。本稿の分析では、RIETI で実施した平成 27 年度「インターネットを使った心の健康法の効果を検証するための Web 調査」のデータを用いた。

1. イントロダクション

メンタルヘルスの改善は、医学的な視点からだけでなく、経済活性化という側面から見ても重要な課題である[1]。従業員のメンタルヘルスが企業収益に影響を及ぼすことが示唆され[2]、「健康経営」の概念が広く認知されてきたことに加えて、労働安全衛生法の改正による2015年末からの「ストレスチェック制度」の導入で、産業界におけるメンタルヘルスへの効果的かつ効率的な取り組みが急務となっている。しかし、医療費増大が問題となっている現在、医療資源は症状の重い人々に優先的に配分される必要があるため、医療の対象となりにくい軽い抑うつ状態の人々への対応が後手に回っている[3]。

メンタルヘルス上の問題の中でも中核的な問題であるうつ病を軽減するための取り組みとして、非適応的な考え方や行動習慣を修正する「認知行動療法」が注目されている。しかし、セラピスト数の少なさなどの理由により、未だ十分な広がりを見せてない。諸外国では、セラピストに頼る前にインターネットを活用した認知行動療法（internet-based Cognitive Behavior Therapy：iCBT）によって、まず自助で取り組むことが推奨されており、英国をはじめとした諸外国では、軽症うつ病の第一選択として公的医療制度にもiCBTが導入されている[4]。しかし、iCBTによる介入では抑うつ症状が短期的には改善するものの、効果が長期的に持続しない、脱落率が高い、社会機能の改善につながらないといった課題が残されている[5]。専門家のサポートを付加したiCBTは効果の増強が期待される反面、コストが高くなるという問題があり[6]、軽症のうつ症状を呈する人々の潜在的な数が極めて多いことを考慮すると、iCBTにおける専門家による関与を最小限にするための対応が必要である。

iCBTを含めた自習型の認知行動療法と比較した場合のセラピストが行う認知行動療法の特徴として、認知行動療法の実施者に共感することによって信頼関係を高めてセラピーの成功に導いたり、実施者が間違った方法で認知行動療法のエクササイズを行った場合に実施者にアドバイスして、適切な方向に実施者を誘導したりすることが考えられる。本研究の問題意識は、人工知能（AI：Artificial Intelligence）を使って、iCBTにおいて、共感やアドバイスといったセラピストの役割を代替できないかというものである。

近年、人口知能技術の多方面への応用が注目されている。人工知能とは、人工的にコンピュータ上で人間と同様の知能を実現させようという試み、或いはそのための一連の基礎技術を指し、1956年にダートマス会議でジョン・マッカーシーにより命名されたとされる[7]。しかし、その定義は広範で多様な領域から構成される。自然言語処理（natural language processing：NLP）は、人間が日常的に使っている自然言語をコンピュータに処理させる人口知能技術である[8]。日本の企業（NECソリューションイノベータ株式会社）によって、NLPを使って、iCBTのwebsiteの画面上に架空のセラピストを登場させ、このセラピストがiCBTの実施者に共感したりアドバイスをしたりするiCBT-AIが開発されて

いる²。これらの機能によって、誰にも共感してもらえない無味乾燥なエクササイズへの動機づけを高め、また間違っただけで行われていても修正されることなく終えていた従来のエクササイズと異なり適切なエクササイズの遂行を促すことが可能となった。

しかし、国際的にもこれまで人工知能機能を搭載した iCBT に対する効果検証は、2016 年 4 月に米国において iCBT-AI を用いた研究計画の論文で報告されているだけであり [9]、AI の iCBT への実際の効果に関しては十分に検証されておらず、無作為統制試験による厳密な検証が求められていると考えられた。

2. 方法

2.1 参加者とリクルートメント

本研究は、日経リサーチ株式会社（以下では「調査会社」と呼ぶ）のモニターとして登録している人々に募集をかけることによって行われた。著者（関沢）が所属する（独）経済産業研究所が調査会社と委託契約を締結することによって、本研究は遂行された。参加のための包有基準は、1) エントリー時の年齢が 20~60 歳、2) インターネット使用環境下にある、3) スクリーニング時に抑うつ尺度である PHQ-9（後述）の得点が 5 点以上、4) 日本語による説明書を理解し、同意できる判断能力を有する、とした。除外基準としては 1) 現在、メンタルヘルスの問題について専門家からなんらかの治療を受けており、主治医の承諾が得られない、2) 介入期間中に認知行動療法を専門家から受ける、3) 統合失調症に罹患している（医療機関による診断の自己申告）、4) 認知症に罹患している（医療機関による診断の自己申告）、5) 過去 12 ヶ月間に何らかの物質依存を有する（喫煙を除く）、とされた。

調査会社が自社のモニターに対して電子メールを送ることによって参加希望者を募り（募集期間は 2015 年 9 月 3 日~10 日）、参加希望者には、参加資格の有無の確認のために、オンライン上で、年齢確認を行うと共に、PHQ-9 に回答してもらった。9 月 16 日に、応募者の中で参加基準を満たした者に対して電子メールを送り、PDF 化した同意説明文書をオンライン上で示し、この文書を読んでもらった上で、自由意思による同意をオンライン上で得た。本研究への参加に同意した者には、ベースライン時点における評価指標及び属性についての質問に引き続き回答してもらった（9 月 16 日~28 日）。ベースライン時点における評価指標には PHQ-9 が含まれたため、ベースライン時点では PHQ-9 の得点が参加基準に満たない者がいることになるが、これらの者も参加者に含めることにした。

2.2 手続き

ベースラインの質問に回答してもらった者は、性別・年齢・抑うつ重症度によって本研究とは独立の研究者によって層化無作為化がなされ、iCBT-AI 群、iCBT 群、待機群に分けられた。iCBT-AI 群と iCBT 群のエクササイズ期間は 7 週間（10 月 5 日~11 月 20 日）で、

² このツールは、特許第 5831951 号「対話システム、冗長メッセージ排除方法および冗長メッセージ排除プログラム」の技術を使用したツールである。

エクササイズ期間が終了した直後に第 2 回目の評価指標のアセスメントが行われ（11 月 23 日～29 日）、更に、その 3 か月後に第 3 回目の評価指標のアセスメントが行われた（2 月 22 日～28 日）。

研究参加者への謝礼として、エクササイズ終了直後と更に 3 ヶ月後のアンケート調査に回答した者には、それぞれ全員に 500 円相当の謝礼が金券で支払われた。このことは、同意説明文書を使った説明の際に参加者に伝えられた。

本研究は特定医療法人社団 慈藻会平松記念病院倫理委員会の承認を受けて行われた。本研究は UMIN 臨床試験登録システムに登録された（ID=UMIN000019228）。

2.3 介入内容

iCBT-AI 群と iCBT 群は、エクササイズへの参加を促す電子メールが 7 週間にわたって週 2 回送付され、この間、待機群は何も行わなかった。iCBT-AI 群と iCBT 群の参加者は毎週 15 分程度の認知行動療法に関連した両群共通の心理教育的な音声動画教材を最初に視聴し、その後、無作為割付に従って、それぞれ iCBT-AI または iCBT のエクササイズサイトにアクセスするように誘導された。音声動画教材は毎週更新されたが、その後のエクササイズの内容は毎週同じであった。参加者が iCBT-AI か iCBT のどちらのエクササイズに従事しているかは、参加者には秘匿され、介入者側もブラインドであったため、この二群については二重盲検とされていた。それぞれのエクササイズの内容は以下のとおりであった。

2.3.1 音声動画教材

エクササイズ実施前に、両群共通の心理教育的な音声動画教材を最初に視聴してもらった。この音声動画教材では、エクササイズサイトにアクセスしてもらう前に、認知再構成法（後述）の方法が教示されると共に、認知行動療法に則した気分改善のためにアドバイスも提供された。

2.3.2 エクササイズサイト iCBT-AI

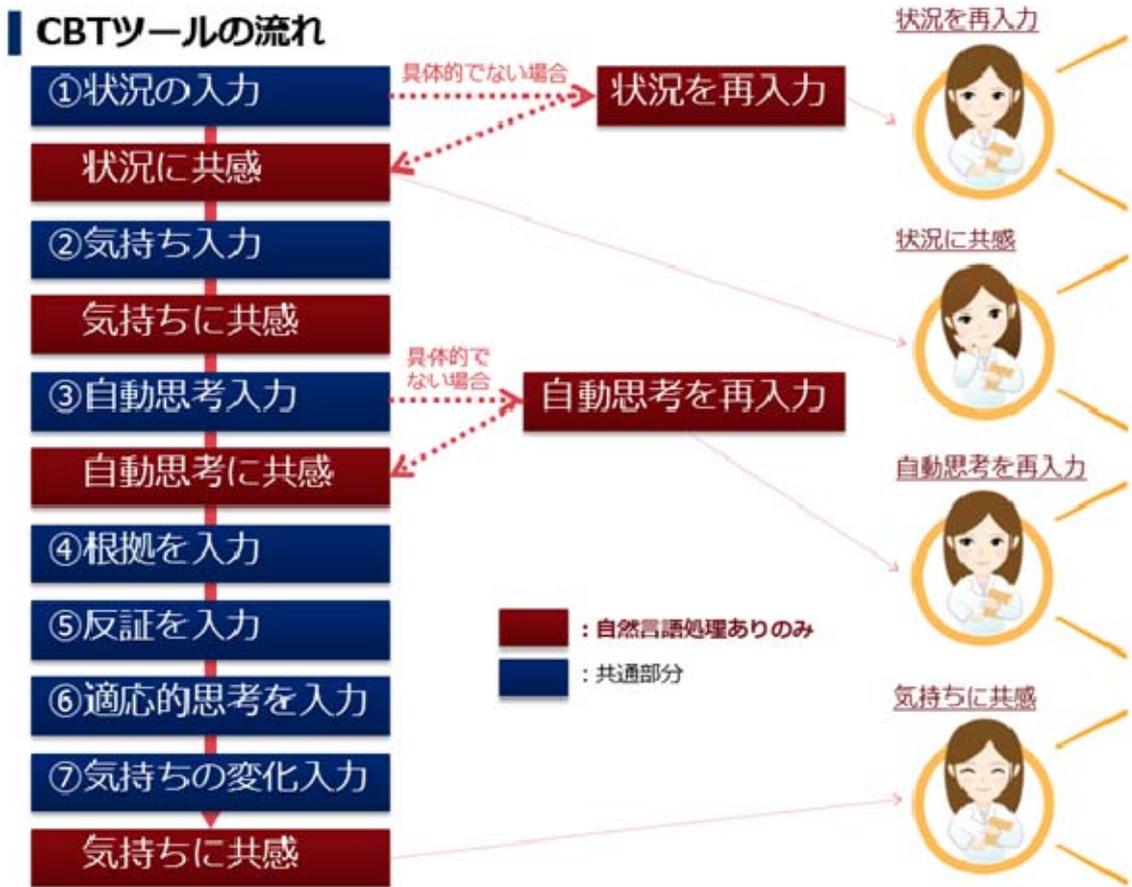
認知行動療法では、歪んだ思考（認知）がネガティブな気分の原因となっているという仮説の下で、歪んだ思考に根拠があるかどうかを検証し、反証を挙げていくことによって、代替的な思考を導くことで、気分の改善を得るというアプローチが採用される場合が多い。これは、認知再構成法と呼ばれる。本研究で採用したプログラムでは、以下の順番に従って情報を入力していき（7 ステップ法と呼ばれる）、歪んだ認知を適応的なものへと修正することを促すものとなっている。

①状況→②気持ち→③思考→④根拠→⑤反証→⑥代替的な適応的思考→⑦気持ちの変化

本研究で使用される iCBT-AI のプログラムには、従来の自助型 iCBT のエクササイズには見られない特徴が2つある。1つめは共感機能である。セラピストが行う認知行動療法のセッションでは、クライアントが自分の直面した状況や感情をセラピストに説明すると、セラピストがクライアントに共感することにより、クライアントとセラピストの間の信頼関係が高まって、セッションの効果が高まることが期待される[10]。ところが、従来の iCBT による認知再構成エクササイズは、ヒトによるフィードバック機能がないために、実施者が書き込んだ状況や感情に対して、プログラム側からの反応がなかった。このため、iCBT が無味乾燥な一人での自習作業となり、セラピストによる iCBT の持つ交流的側面が欠けていた。これに対して、iCBT-AI では、実施者が書き込んだ状況や感情に関する記述が NLP で解釈され、適宜、その内容を反映した共感の言葉が画面上に出現してフィードバックが得られ、また、それに連動してナビゲーター役の女性キャラクターの表情も変化するように設計されていた。

iCBT-AI の2つめの特徴は実施者にアドバイスを行うことである。認知再構成エクササイズでは、ストレスとなっている状況とその状況に対応して生じる思考（自動思考）を具体的に書き込むのが望ましいが、従来の iCBT では、実施者の書き込んだ内容が具体性を欠いていても書き直しするようにアドバイスすることはできなかった。これに対して iCBT-AI では、状況や思考の記述が具体性に欠いていると NLP によって判断された際に、書き直しをするようにアドバイスする注意喚起メッセージが画面上に表示されることで、より適切なエクササイズの実施へと導く機能を有していた。その結果、本来のプロトコルに従ったエクササイズの実施が可能となって効果の増強が期待されると予想された（図1参照）。例えば、「困った」といったように抽象的で漠然とした現状認識（前操作的思考）の習慣がある者は、具体的で実効性を有する問題解決を得られにくく、その結果、それが抑うつ慢性化の要因となると考えられている[11]。特に、つらい気持ちが生じるきっかけとなる出来事の状況を 5W1H 近い形で具体的に記述することは抽象的思考を防ぎ、結果的に効果的な認知再構成を行うための有用なプロセスともされる。そのため、状況と思考の記述を具体化することが認知行動療法の成功の鍵だと認識されており、iCBT-AI のアドバイス機能はこれに資するものと予想された。

図1 AIを活用したインターネット認知行動療法（iCBT-AI）の流れ



この前提となる NLP による対話エンジン技術は、過去の認知再構成法 28,718 事例の悩み文を元に、悩みや気持ち、主観的な表現の抽出を通じてコーパスが作成され（図 2）、この対話エンジンを用いた対話支援の違和感の有無に関する定性評価試験の結果では、違和感を抱いた人々の割合は、気持ちを推定した共感機能に関しては 9%、気持ちと考えのズレを検知後の確認では 4%と少ないものであった（図 3）。

図2 NLP の悩み文用コーパス作成における表現抽出例

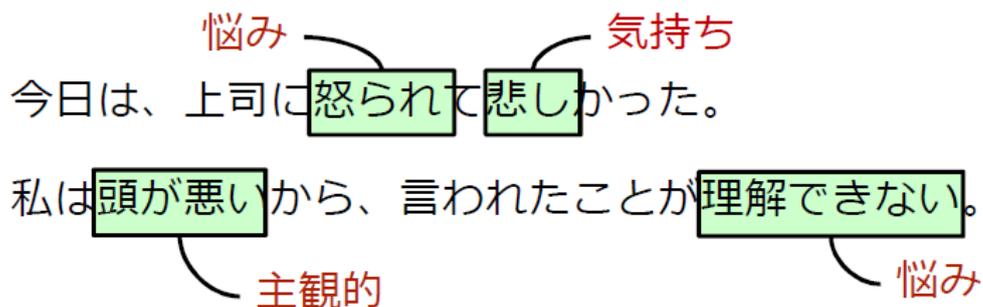
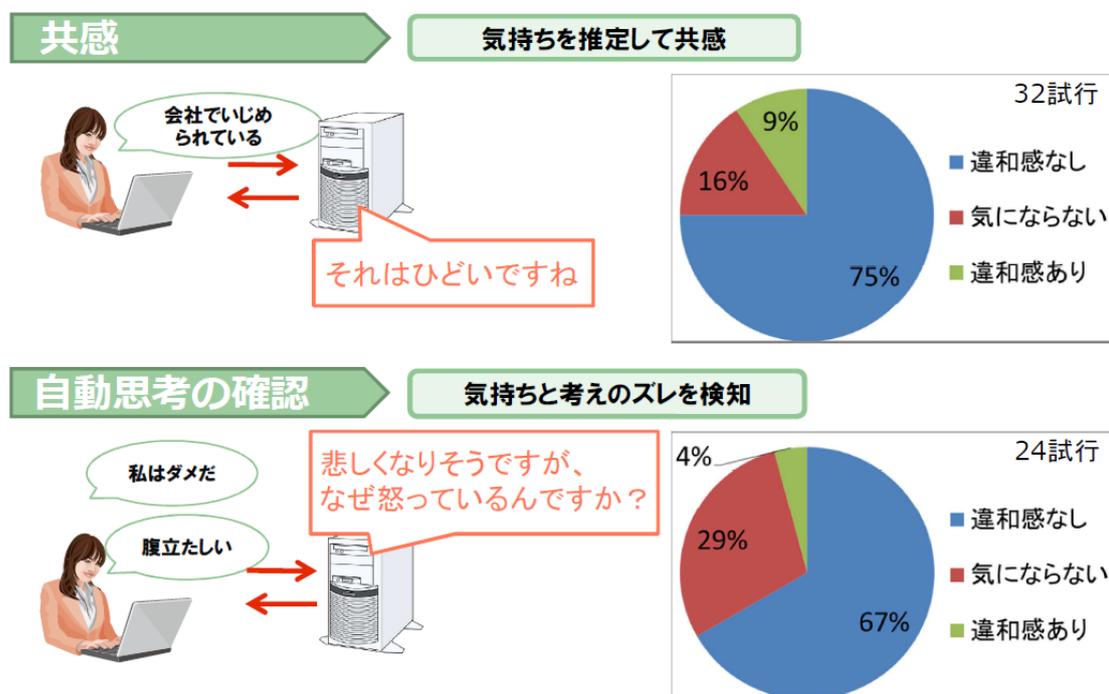


図3 NLPによる対話エンジンを用いた対話支援の定性評価結果



2.3.3 エクササイズサイト iCBT

iCBTのエクササイズサイトは、共感機能とアドバイス機能が存在しない他は、iCBT-AIと全く同じものとなっている。

2.4 評価指標

(1) PHQ-9

主要評価指標として、本研究では Patient Health Questionnaire (PHQ-9)を採用した。PHQ-9は、大うつ病性障害等の診断のために開発された質問票で[12]、日本語版は村松らが作成している[13]。多忙なプライマリケア医が短時間で精神疾患を診断・評価するためのシステムである PRIME-MD (Primary Care Evaluation of Mental Disorders) を Spitzer R. L らが開発し、さらに実施時間の短縮化のために PRIME-MD の自己記入式質問票版として Patient Health Questionnaire (PHQ) を開発した。PHQ はプライマリケア医が日常診療において遭遇する 8 種類の疾患の診断・評価ができるようになっている。PHQ の中から、大うつ病性障害モジュールの 9 個の質問項目を抽出したものが PHQ-9 である。日本語版での信頼性や妥当性は、再翻訳法によって検証されている。英国の国立医療技術評価機構 (NICE; National Institute of Health and Clinical Excellence) のうつ病治療のガイドラインや米国精神医学会 (American Psychiatric Association APA) は、うつ病の評価尺度とし

て PHQ-9 を推奨している。DSM-5 (The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition) に対応する 9 個の質問から PHQ-9 は構成されており、過去 2 週間について、「全くない=0 点」「数日=1 点」「半分以上=2 点」「ほとんど毎日=3 点」となっている。合計点は 0~27 点で、0~4 点はうつ状態でない、5~9 点は軽度のうつ、10~14 点は中等度のうつ、15~19 点は中等度~重度のうつ、20~27 点は重度のうつとなる。

(2) Quick Inventory of Depressive Symptomatology (QIDS-J)

オリジナルの QIDS は、Rush らによって抑うつ症状評価尺度である 30 項目の Inventory of Depressive Symptomatology (IDS) として開発され、その後に QIDS-SR16 として 16 項目の簡易抑うつ症状自記式評価尺度としてまとめられた[14]。抑うつの重症度を評価できるほか、アメリカ精神医学会の診断基準である DSM-IV の大うつ病性障害（中核的なうつ病）の診断基準に対応しているという特長を持っている。QIDS-SR16 日本版は藤澤らによって QIDS-J として妥当性信頼性が検証されている[15]。本研究では、抑うつの主要評価項目である PHQ-9 が日数を点数化した頻度基準であることを考慮して、重症度を点数化した程度基準である本尺度を抑うつの副次的評価として採用した。

(3) Generalized Anxiety Disorder-7 (GAD-7) 日本語版

GAD-7 は、全般性不安障害を簡易に評価するための質問票として開発されたもので[16]、日本語版は村松らが作成している[13]。GAD-7 は 7 つの質問から構成されており、過去 2 週間について、「全くない=0 点」「数日=1 点」「半分以上=2 点」「ほとんど毎日=3 点」となっている。合計点は 0~21 点で、0~4 点は全般性不安障害がなく、5~9 点は軽度、10~14 点は中等度、15~21 点は重度と評価される。英国における国立医療技術評価機構 (NICE; National Institute of Health and Clinical Excellence) ガイドラインでは全般性不安障害の評価尺度として、GAD-7 を推奨している。

(4) Sheehan Disability Scale (SDISS) 日本語版

Sheehan Disability Scale は、機能障害評価尺度の一つである。この指標は疼痛評価における視覚的評価スケール (Visual Analog Scale : VAS) と同様に社会機能障害を数字で評価することができる。その簡便さと治療効果に対する正確性ゆえに、米国においてこの指標が広まりつつある[17]。日本語版においては、吉田らにより信頼性および妥当性検証が行われ、高い信頼性が示されている[18]。また、SDISS 日本語版と Global assessment of functioning (GAF) との相関でも高い併存妥当性を示し、SDISS 日本語版は高い信頼性・妥当性を有し、機能障害を測定する簡易尺度として有用であることが示唆されている。

(5) CSQ-8-J

CSQ-8-J (Customer Satisfaction Questionnaire-8-Japanese version) は、プログラムを実施してもらった人々にその評価を書き込んでもらう指標であり、プログラムに対する満足度を調査するためのものである[19]。本研究では、介入2群の参加者に、エクササイズ期間終了直後の時点1において、エクササイズについての満足度をCSQ-8-Jにて評価された。

(6)他の評価指標

毎週エクササイズサイトにアクセスする度に、抑うつと不安の簡易尺度である ODSIS および OASIS の施行後でなければ、エクササイズに進めないように設定され、毎週の抑うつや不安の変化も追跡できるようにも設定された。さらに、全員に期間内のコストの群間比較を行うため、介入期間中における諸コストもアンケート形式で聴取された。しかし、本研究は原則として解析者がブラインドで行った解析のみを対象とすることとし、これらの結果は、今回の報告対象外とした。

2.5 分析手法

本研究においては、評価指標の計測は開始時、介入終了時（開始7週間後）、3か月後フォローアップ時（介入終了後12週間後）の3時点で行われるため、その3時点の評価対象とした。データ解析は、既存の研究[20]を踏まえて、以下の2つの方法によって行った。第一に、PHQ-9の得点が10点以上が大うつ病水準の病態とされていることから、PHQ-9の得点が10点以上と10点未満の二値アウトカムを作り、これを従属変数として、ロジスティック回帰分析により、介入終了時と3か月後フォローアップ時において、各群間で大うつ病の基準を満たす者の割合に違いがあるかを検証した。解析では、各時点毎に、独立変数を、群（iCBT-AI群、iCBT群、待機群）、開始時の評価指標（PHQ-9、GAD-7）、性別、年齢として、オッズ比を求めた。

次に、PHQ-9得点を加工せずに連続値として扱い、混合効果モデルによる反復測定データ解析法（Mixed Model Repeated Measures, MMRM）を行った。他の評価指標についても同様の分析を行った。従属変数は評価指標（PHQ-9、GAD-7、QIDS-J、SDISS、CSQ8-J）であり、独立変数は、時点（開始時、介入終了時、3か月後フォローアップ時）、群（iCBT-AI群、iCBT群、待機群）、時点と群の交差項、及び、ベースラインの評価指標、性別、年齢とした。各時点における効果を見るために、時点はカテゴリカル変数とした。また、エクササイズの脱落を評価するために、第2週目以降のエクササイズへの参加有無を第6週目まで毎週サイト上で記録し、それらの平均をt検定によって比較検討した。さらに下位集団分析として、うつ症状の重さに応じた効果の違いを見るために、ベースラインのPHQ-9の得点が5点未満の非うつ群のみ、5点以上で10点未満の軽症うつ群、10点以上の重症うつ群グループに分けての分析も行われた。解析は、全参加者のデータを用いたITT（Intention-to-Treat）解析で行われた。

各群の参加者の基本統計量の違いは、分散分析及びカイ二乗検定によって検証した。分析はプロトコルの作成に関わらない第 3 者が割り付けをマスクキングされた状態で実施された。分析の有意水準は 5%とし、ブラインド解除前の一次的解析では「R」、ブラインド解除後の二次的解析では「STATA 13 (Lightstone Corp)」が解析ソフトとして用いられた。

3 結果

3-1. 参加者の属性

参加者の属性を表 1 に示した。基準を満たし本研究の参加に同意した対象者 1187 人が、性別、年齢、抑うつ度によって三群に層別無作為割付がなされた。開始時における iCBT-AI 群は 43.6 歳で男性は 58.6%、iCBT 群は平均年齢 43.2 歳で男性 58.4%、待機群は 43.6 歳で男性は 59.4%であった。四大卒以上が約 6 割と多数を占めており、勤労者が約 7 割を占めている。約半数は結婚しており、4 割弱が未婚者となっている。他の抑うつ、不安、婚姻状況、学歴、就業状況も含めた全属性において有意な群間差は認められなかった。

表 1 基本統計量

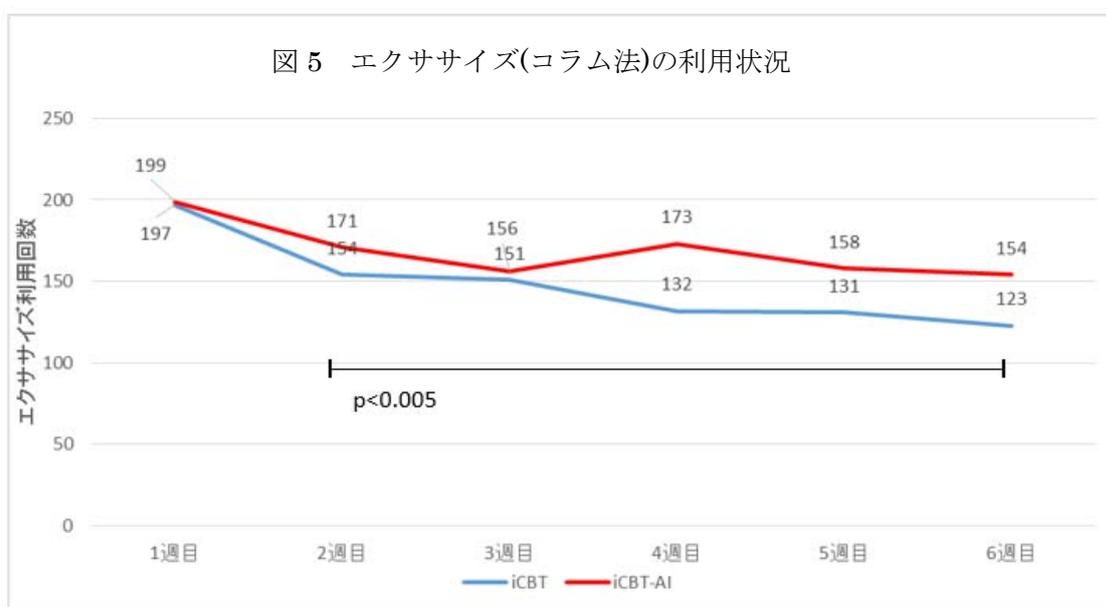
		iCBT群	iCBT-A群	待機群	全体	
性別	男性	232 (58.4%)	232 (58.6%)	234 (59.4%)	698 (58.8%)	
	女性	165 (41.6%)	164 (41.4%)	160 (40.6%)	489 (41.2%)	
年齢	平均 (標準偏差値)	43.2 (9.9)	43.6 (9.5)	43.6 (10.1)	43.5 (9.8)	
婚姻状況	既婚	219 (55.2%)	226 (57.1%)	220 (55.8%)	665 (56.0%)	
	離婚	17 (4.3%)	29 (7.3%)	30 (7.6%)	76 (6.4%)	
	死別	4 (1.0%)	2 (0.5%)	1 (0.3%)	7 (0.6%)	
	未婚	157 (39.5%)	139 (35.1%)	143 (36.3%)	439 (37.0%)	
最終学歴	中学校	1 (0.3%)	3 (0.8%)	3 (0.8%)	7 (0.6%)	
	高校	84 (21.2%)	87 (22.0%)	68 (17.3%)	239 (20.1%)	
	短大・高専・専門学校	73 (18.4%)	71 (17.9%)	72 (18.3%)	216 (18.2%)	
	大学・大学院	239 (60.2%)	235 (59.3%)	251 (63.7%)	725 (61.1%)	
勤労状況	働いている	324 (81.6%)	320 (80.8%)	328 (83.2%)	972 (81.9%)	
	無職 (休職中)	30 (7.6%)	27 (6.8%)	22 (5.6%)	79 (6.7%)	
	無職 (休職中でない)	43 (10.8%)	49 (12.4%)	44 (11.2%)	136 (11.5%)	
評価指標のベースライン得点	PHQ-9	平均 (標準偏差値)	8.6 (4.9)	8.8 (4.9)	8.8 (4.7)	8.7 (4.8)
	QDS-J	平均 (標準偏差値)	8.5 (5.2)	8.8 (5.2)	8.9 (5.1)	8.7 (5.2)
	GAD-7	平均 (標準偏差値)	5.9 (4.7)	6.0 (4.4)	6.2 (4.7)	6.0 (4.6)

3-2. 参加者のエクササイズ遂行状況

参加者のフローチャートを図 4 (最終頁) に示した。介入終了時のアセスメントに回答したのは、iCBT 群が 56.2% (=223 人/397 人)、iCBT-AI 群が 61.4% (=243 人/396 人)、待機群が 86.0% (=339 人/394 人)、さらに 3 か月後フォローアップ時では、アセスメントに回答したのは、iCBT 群が 60.7% (=241 人/397 人)、iCBT-AI 群が 65.4% (=259 人/396 人)、待機群が 81.0% (=319 人/394 人)であった。介入終了時、フォローアップ時のいずれにおいても、

待機群に比べて介入2群で得られた回答数は有意に低かった ($p < 0.01$)。iCBT と iCBT-AI の二群間においては、介入終了時および3か月後フォローアップ時における回答数に有意差は認められなかった。

認知再構成法のエクササイズに実際に従事した人数を毎週毎に集計すると、週平均で iCBT 群では 148 ± 26.9 人/週に対して、iCBT-AI 群では 168.5 ± 16.9 人/週であり、2週目参加以降の平均人数は、図5のようになり、iCBT-AI 群は iCBT 群より第2週目以降のエクササイズ実施数が有意に多かった ($p < 0.05$)。



3-3. 分析結果

3-3-1. 主要評価指標

大うつ病性障害水準とされる PHQ-9 の得点が 10 点以上の場合、開始時ではいずれの群も 36%が大うつ病性障害水準であり、介入終了時では iCBT 群が 29%、iCBT-AI 群が 34%、待機群が 32%となり、3か月後フォローアップ時では iCBT 群が 28%、iCBT-AI 群が 25%、待機群が 31%となった(表2)。

表2 PHQ-9の得点が10点以上の参加者数(割合)

	iCBT群	iCBT-AI群	待機群
開始時	143/397 (0.36)	142/396 (0.36)	143/394 (0.36)
介入終了時	65/223 (0.29)	82/243 (0.34)	109/339 (0.32)
3ヶ月フォローアップ時	67/241 (0.28)	66/259 (0.25)	100/319 (0.31)

(注) 分母が回答者数で、分子がPHQ-9の得点が10点以上の人数。

主要評価指標である PHQ-9 での 2 値アウトカムでは、介入終了直後には iCBT-AI 群および iCBT の両群は待機群に対してオッズ比で有意差は認められなかった(表3)。しかし、3か月後フォローアップ時点では、PHQ-9 で 10 点以上の大うつ病性障害状態を呈している

者のオッズ比は、iCBT 群では待機群と比べて有意差は認められなかったが、iCBT-AI 群では待機群に比べてオッズ比が 0.67 (95%CI:0.43 to 1.04, $p=0.08$) と有意傾向にあることが認められた。iCBT-AI は短期では効果は認められないが、時間経過と共に長期的な重症うつ者を減少させる可能性が示唆された。

表3 ロジスティック回帰分析の結果 (オッズ比は重症うつ状態についてのもの)

	オッズ比 (95%信頼区間) :P値		
	iCBT群 vs iCBT-AI群	CBT群 vs 待機群	iCBT-AI群 vs 待機群
介入終了時	0.82(0.51 to 1.33):0.43	0.82(0.51 to 1.32):0.41	1.00(0.63 to 1.59):1.00
3ヶ月後フォローアップ	1.36(0.84 to 2.20):0.21	0.89(0.57 to 1.38):0.61	0.67(0.43 to 1.04):0.08

(注) 重症うつ状態は、PHQ-9の得点が10点以上。

表4に示した通り、PHQ-9の連続アウトカムでは、AI機能のないiCBT群では介入終了時では待機群と比べて得点が減少する傾向が見られ(-0.54, 95%CI:-1.08 to -0.01, $p=0.05$)、3か月後フォローアップでは有意な改善が認められたが(-0.69, 95%CI:-1.22 to -0.15, $p=0.01$)、iCBT-AI群では、介入終了時、3か月後フォローアップのいずれにおいても有意な改善はなかった。しかし、iCBT-AI群とiCBT群の間の直接比較では、短期および長期においても、両群間に有意差は認められなかった。

表4 混合効果モデルによる分析結果 (PHQ-9の連続値)

	平均値 (95% CI)		群間差 (95%CI) :p値	効果量 (95%CI)
	iCBT群 (or iCBT-AI群)	iCBT-AI群 (or 待機群)		
開始時				
iCBT群 vs iCBT-AI群	8.71(8.39 to 9.03)	8.80(8.48 to 9.12)	-0.09(-0.54 to 0.36):0.71	-0.66(-0.2 to 0.08)
iCBT群 vs 待機群	8.69(8.37 to 9.01)	8.76(8.44 to 9.08)	-0.07(-0.52 to 0.38):0.77	-0.07(-0.21 to 0.07)
iCBT-AI vs 待機群	8.91(8.59 to 9.24)	8.90(8.58 to 9.22)	0.02(-0.44 to 0.47):0.95	-0.01(-0.15 to 0.13)
介入終了時				
iCBT群 vs iCBT-AI群	7.72(7.29 to 8.14)	7.89(7.49 to 8.30)	-0.18(-0.76 to 0.41):0.55	-0.09(-0.27 to 0.09)
iCBT群 vs 待機群	7.70(7.28 to 8.12)	8.23(7.89 to 8.58)	-0.54(-1.08 to 0.01):0.05	-0.11(-0.28 to 0.06)
iCBT-AI群 vs 待機群	8.01(7.60 to 8.41)	8.37(8.02 to 8.72)	-0.37(-0.9 to 0.17):0.18	-0.03(-0.19 to 0.14)
3ヶ月後フォローアップ時				
iCBT群 vs iCBT-AI群	7.13(6.72 to 7.54)	7.36(6.97 to 7.76)	-0.23(-0.80 to 0.33):0.42	-0.12(-0.29 to 0.06)
iCBT群 vs 待機群	7.11(6.71 to 7.52)	7.80(7.44 to 8.15)	-0.69(-1.22 to -0.15):0.01	-0.21(-0.38 to -0.04)
iCBT-AI群 vs 待機群	7.48(7.08 to 7.78)	7.94(7.58 to 8.30)	-0.46(-0.99 to 0.07):0.09	-0.09(0.25 to 0.08)

3-3-2. 副次的評価項目

不安の程度を示すGAD-7では、介入終了時にはiCBT群およびiCBT-AI群の両者とも有意な効果は認められなかったが、3か月後フォローアップではiCBT群では有意な改善が認められ(-0.53, 95%CI:-0.99 to -0.06, $p=0.03$)、iCBT-AI群でも改善傾向が認められ(-0.42, 95%CI:-0.86 to -0.02, $p=0.06$)、即効性が低い一方で両群共に遅発性の効果出現の傾向が示唆された(表5)。iCBTとiCBT-AIとの間にはどの時点においても有意な差は認められなかった。

表5 混合効果モデルによる分析結果 (GAD-7の連続値)

	平均値 (95% CI)		群間差 (95%CI) ; p値	効果量 (95%CI)
	iCBT群 (or iCBT-AI群)	iCBT-AI群 (or 待機群)		
開始時				
iCBT群 vs iCBT-AI群	6.04 (5.77 to 6.30)	6.05 (5.78 to 6.31)	-0.01 (-0.38 to 0.36); 0.96	-0.01 (-0.15 to 0.13)
iCBT群 vs 待機群	6.11 (5.83 to 6.39)	6.17 (5.89 to 6.45)	-0.06 (-0.45 to 0.33); 0.76	-0.06 (-0.20 to 0.08)
iCBT-AI vs 待機群	6.15 (5.88 to 6.41)	6.21 (5.95 to 6.48)	-0.07 (-0.44 to 0.31); 0.73	-0.06 (-0.20 to 0.08)
介入終了時				
iCBT群 vs iCBT-AI群	5.27 (4.92 to 5.62)	5.33 (5.00 to 5.67)	-0.07 (-0.55 to 0.42); 0.78	-0.03 (-0.22 to 0.15)
iCBT群 vs 待機群	5.34 (4.98 to 5.71)	5.52 (5.22 to 5.82)	-0.18 (-0.65 to 0.29); 0.46	-0.05 (-0.22 to 0.12)
iCBT-AI群 vs 待機群	5.43 (5.10 to 5.77)	5.56 (5.28 to 5.85)	-0.13 (-0.57 to 0.31); 0.57	-0.02 (-0.19 to 0.14)
3ヶ月後フォローアップ時				
iCBT群 vs iCBT-AI群	4.80 (4.47 to 5.14)	4.93 (4.61 to 5.26)	-0.31 (-0.60 to 0.34); 0.59	-0.06 (-0.24 to 0.11)
iCBT群 vs 待機群	4.88 (4.53 to 5.23)	5.41 (5.10 to 5.71)	-0.53 (-0.99 to -0.06); 0.03	-0.19 (-0.36 to -0.02)
iCBT-AI群 vs 待機群	5.03 (4.70 to 5.36)	5.45 (5.15 to 5.75)	-0.42 (-0.86 to 0.02); 0.06	-0.14 (-0.30 to 0.03)

QIDS-Jでの連続アウトカムでは、iCBT-AI群は、介入終了時および3か月後フォローアップの両方で有意な効果は認められなかった(表6)。iCBT群では介入終了時で有意な改善効果を示したが(-0.81, 95%CI: -1.40 to -0.23, $p=0.01$)、3か月後フォローアップでは有意差は認めなかった(-0.50, 95%CI: -1.08 to 0.08, $p=0.09$)。また、介入終了時におけるiCBT群とiCBT-AI群の間の比較では、通常iCBT群がiCBT-AIより優れている可能性が示唆された(-0.58, 95%CI: -1.21 to 0.05, $p=0.07$)。

表6 混合効果モデルによる分析結果 (QIDS-Jの連続値)

	平均値 (95% CI)		群間差 (95%CI) ; p値	効果量 (95%CI)
	iCBT群 (or iCBT-AI群)	iCBT-AI群 (or 待機群)		
開始時				
iCBT群 vs iCBT-AI群	8.72 (8.36 to 9.08)	8.83 (8.47 to 9.19)	-0.11 (-0.62 to 0.39); 0.66	-0.05 (-0.19 to 0.09)
iCBT群 vs 待機群	8.73 (8.38 to 9.08)	8.72 (8.36 to 9.07)	0.01 (-0.48 to 0.51); 0.96	-0.05 (-0.19 to 0.09)
iCBT-AI vs 待機群	8.96 (8.61 to 9.31)	8.84 (8.49 to 9.19)	0.12 (-0.37 to 0.61); 0.62	0.00 (-0.14 to 0.14)
介入終了時				
iCBT群 vs iCBT-AI群	7.33 (6.88 to 7.78)	7.91 (7.47 to 8.34)	-0.58 (-1.21 to 0.05); 0.07	-0.17 (-0.35 to 0.02)
iCBT群 vs 待機群	7.35 (6.90 to 7.80)	8.16 (7.79 to 8.54)	-0.81 (-1.40 to -0.23); 0.01	-0.17 (-0.34 to 0.00)
iCBT-AI群 vs 待機群	8.04 (7.61 to 8.46)	8.29 (7.29 to 8.66)	-0.25 (-0.81 to 0.31); 0.39	0.00 (-0.17 to 0.16)
3ヶ月後フォローアップ時				
iCBT群 vs iCBT-AI群	7.36 (6.92 to 7.80)	7.50 (7.07 to 7.92)	-0.14 (-0.75 to 0.47); 0.66	-0.07 (-0.24 to 0.11)
iCBT群 vs 待機群	7.38 (6.94 to 7.81)	7.78 (7.49 to 8.26)	-0.50 (-1.08 to 0.08); 0.09	-0.16 (-0.33 to 0.00)
iCBT-AI群 vs 待機群	7.62 (7.21 to 8.04)	8.00 (7.62 to 8.38)	-0.37 (-0.94 to 0.19); 0.19	-0.10 (-0.26 to 0.07)

CSQ-8-JおよびSDSにおいては群間の有意な差は認められなかった。

3-3-3. 下位集団分析

対象が臨床患者群ではない本研究では、特に軽症うつ者にどのような効果が得られるのかという点が大きな焦点であった。また参加募集時に行ったPHQ-9の得点が5点以上だと参加資格があったが、開始時のPHQ-9の得点で選別を行っていないため、開始時に含まれてしまった募集時に5点未満の得点を有する参加者(うつでない人々)も効果検証に含ま

れ、バイアスとなっていた。そこで、開始時における主要アウトカムのPHQ-9が、非うつ者（5点未満）、軽症うつ者（5点以上で10点未満）、重症うつ者（PHQ-9が10点以上）を層別化して下位集団分析を行った。

うつの重症度別に分けられた下位集団のデータでは、PHQ-9の二値アウトカムでは、非うつ者、および重症のうつ状態を呈する者においては、iCBT群およびiCBT-AI群の両群とも、介入終了時および3か月後フォローアップ時において、待機群に対して有意差は認められなかった（表7）。軽症うつ者では、介入終了時には各群間で有意差は認められなかった。3か月後フォローアップ時においては、iCBT群と待機群の間では有意差が認められなかったが、iCBT-AI群においては、待機群に比べてオッズ比が0.35(95%CI:0.14 to 0.86, $p=0.02$)と有意に低かった。このことから、軽症うつ者において長期的にはうつ重症者を減らす効果がiCBT-AIだけにおいて認められ、iCBT群や待機群では認められなかったことが示唆された。

表7 ロジスティック回帰分析の結果（うつのPHQ-9得点による重症度別）オッズ比

開始時の得点		オッズ比(95%CI):P値		
		iCBT群 vs iCBT-AI群	iCBT群 vs 待機群	iCBT-AI群 vs 待機群
5点未満	介入終了時	0.77(0.11 to 5.45):0.79	3.27(0.27 to 39.37):0.35	6.09(0.53 to 70.59):0.15
	3ヶ月フォローアップ時	0.97(0.13 to 7.45):0.98	1.29(0.17 to 9.89):0.81	1.15(0.13 to 10.26):0.90
5点以上 10点未満	介入終了時	0.67(0.29 to 1.54):0.35	1.06(0.47 to 2.39):0.89	1.37(0.66 to 2.84):0.40
	3ヶ月フォローアップ時	1.89(0.73 to 4.93):0.19	0.79(0.36 to 1.75):0.57	0.35(0.14 to 0.86):0.02
10点以上	介入終了時	0.81(0.42 to 1.57):0.54	0.59(0.31 to 1.10):0.09	0.71(0.37 to 1.35):0.29
	3ヶ月フォローアップ時	1.10(0.60 to 2.01):0.77	0.86(0.49 to 1.53):0.61	0.81(0.46 to 1.43):0.46

(注) 性別、年齢、ベースラインのPHQ-9、GDA-7の数値を調整している。各時点毎に算出。

PHQ-9での連続アウトカムでは、非うつ群ではiCBT群もiCBT-AI群もどちらも有意な効果は認められなかった（表8）。一方で、開始時に重症うつ状態を呈していた者は、介入終了時にはiCBT群、iCBT-AI群それぞれにおいて、-1.65(95%CI:-2.72 to -0.58, $p<0.01$), -1.15(95%CI:-2.22 to -0.09, $p=0.03$)で有意な抗うつ効果が認められた。しかし、3か月後フォローアップ時ではどちらの効果も消滅していた。それに対して、軽症うつ群では、iCBT群およびiCBT-AI群の両者において、介入終了時には有意な効果を認めていなかったが、3か月後フォローアップ時には両者とも有意な効果が、それぞれ-1.21(95%CI:-1.89 to -0.52, $p<0.01$)および-1.03(95%CI:-1.69 to -0.37, $p<0.01$)と出現していた。

表 8 混合効果モデルによるPHQ-9の推移（うつのPHQ-9得点による重症度別）

		平均値 (95%CI)		
		iCBT群 (or iCBT-AI群)	iCBT-AI群 (or 待機群)	群間差 (95%CI) : P値
ベースライン 5点未満	開始時			
	iCBT群 vs iCBT-AI群	2.59 (2.03 to 3.15)	2.68 (2.11 to 3.24)	-0.09 (-0.89 to 0.71) : 0.83
	iCBT群 vs 待機群	2.56 (2.01 to 3.10)	2.73 (2.16 to 3.30)	-0.17 (-0.96 to 0.62) : 0.67
	iCBT-AI群 vs 待機群	2.83 (2.27 to 3.38)	2.86 (2.28 to 3.45)	-0.04 (-0.84 to 0.77) : 0.93
	介入終了時			
	iCBT群 vs iCBT-AI群	4.22 (3.48 to 4.96)	4.36 (3.63 to 5.09)	-0.14 (-1.18 to 0.90) : 0.80
	iCBT群 vs 待機群	4.19 (3.48 to 4.90)	3.56 (2.96 to 4.16)	0.63 (-0.30 to 1.56) : 0.19
	iCBT-AI群 vs 待機群	4.51 (3.79 to 5.23)	3.71 (3.10 to 4.32)	0.80 (-0.15 to 1.74) : 0.10
	3ヶ月後フォローアップ時			
	iCBT群 vs iCBT-AI群	3.12 (2.42 to 3.82)	3.25 (2.53 to 3.98)	-0.13 (-1.14 to 0.88) : 0.80
	iCBT群 vs 待機群	3.11 (2.43 to 3.78)	3.21 (2.56 to 3.86)	-0.10 (-1.04 to 0.84) : 0.83
	iCBT-AI群 vs 待機群	3.41 (2.69 to 4.12)	3.39 (2.73 to 4.06)	0.02 (-0.96 to 1.00) : 0.97
ベースライン 5点以上10点未満	開始時			
	iCBT群 vs iCBT-AI群	6.91 (6.52 to 7.30)	7.00 (6.61 to 7.39)	-0.09 (-0.64 to 0.46) : 0.74
	iCBT群 vs 待機群	6.92 (6.52 to 7.31)	6.95 (6.56 to 7.34)	-0.03 (-0.59 to 0.52) : 0.91
	iCBT-AI群 vs 待機群	7.03 (6.63 to 7.43)	7.02 (6.63 to 7.42)	0.00 (-0.56 to 0.57) : 0.99
	介入終了時			
	iCBT群 vs iCBT-AI群	6.27 (5.74 to 6.81)	6.34 (5.84 to 6.84)	-0.07 (-0.80 to 0.67) : 0.86
	iCBT群 vs 待機群	6.32 (5.78 to 6.86)	6.53 (6.11 to 6.95)	-0.21 (-0.90 to 0.48) : 0.55
	iCBT-AI群 vs 待機群	6.38 (5.86 to 6.89)	6.60 (6.17 to 7.03)	-0.22 (-0.89 to 0.45) : 0.51
	3ヶ月後フォローアップ時			
	iCBT群 vs iCBT-AI群	5.54 (5.02 to 6.06)	5.80 (5.32 to 6.27)	-0.25 (-0.96 to 0.45) : 0.48
	iCBT群 vs 待機群	5.58 (5.05 to 6.10)	6.78 (6.35 to 7.22)	-1.21 (-1.89 to -0.52) : <0.01
	iCBT-AI群 vs 待機群	5.82 (5.33 to 6.31)	6.85 (6.40 to 7.29)	-1.03 (-1.69 to -0.37) : <0.01
ベースライン 10点以上	開始時			
	iCBT群 vs iCBT-AI群	14.28 (13.65 to 14.92)	14.39 (13.75 to 15.03)	-0.11 (-1 to 0.79) : 0.82
	iCBT群 vs 待機群	14.17 (13.53 to 14.81)	14.23 (13.59 to 14.87)	-0.06 (-0.97 to 0.84) : 0.89
	iCBT-AI群 vs 待機群	14.51 (13.86 to 15.15)	14.43 (13.78 to 15.07)	0.08 (-0.83 to 0.99) : 0.86
	介入終了時			
	iCBT群 vs iCBT-AI群	11.28 (10.47 to 12.09)	11.75 (10.96 to 12.53)	-0.47 (-1.59 to 0.66) : 0.42
	iCBT群 vs 待機群	11.17 (10.36 to 11.99)	12.82 (12.13 to 13.52)	-1.65 (-2.72 to -0.58) : <0.01
	iCBT-AI群 vs 待機群	11.86 (11.06 to 12.66)	13.01 (12.31 to 13.71)	-1.15 (-2.22 to -0.09) : 0.03
	3ヶ月後フォローアップ時			
	iCBT群 vs iCBT-AI群	11.12 (10.33 to 11.9)	11.61 (10.83 to 12.38)	-0.49 (-1.59 to 0.61) : 0.38
	iCBT群 vs 待機群	11.01 (10.21 to 11.8)	11.59 (10.9 to 12.28)	-0.58 (-1.63 to 0.47) : 0.28
	iCBT-AI群 vs 待機群	11.69 (10.91 to 12.48)	11.78 (11.09 to 12.48)	-0.09 (-1.14 to 0.96) : 0.87

(注) 性別、年齢、ベースラインのPHQ-9とGAD-7の得点を調整している。それぞれ、2群×3時点のモデルとしている。

4. 考察

本研究では、AI機能の一領域であるNLPを用いた自習型インターネット認知行動療法(iCBT-AI)について、通常のインターネット認知行動療法(iCBT)、及び、待機群と比較して、抑うつ症状の改善の程度について差が見られるかを無作為比較統制試験によって検証した。その結果、iCBT-AIはiCBTに比べて、エクササイズの実施回数が有意に多く、iCBTで懸念とされている高い脱落率を減らす可能性が示された。その一方で、全体での抑うつ症状の改善効果は、PHQ-9およびQIDS-Jの両者において、通常のiCBT群の方が、iCBT-AIよりも待機群に対して有利な結果が示された。特に、QIDS-Jでは、介入直後においてiCBT群とiCBT-AI群の差に有意傾向まで認められた。iCBT-AIとiCBTの違いは、AI機能の作動以外にはないデザインのため、これらの抗うつ効果の差は、iCBTが本来有するはずの介入効果がAI機能によって減弱された可能性が高いと考えられた。一方で、PHQ-9にお

る二値の結果では、長期的に iCBT-AI 群のみにおいて、待機群と比べて、PHQ-9 で 10 点以上の重症のうつ者の割合を約半減させる可能性（有意傾向）が示唆され、これは PHQ-9 で 5 点以上 10 点未満の対象者に下位集団分析で絞ってみた場合、3 か月後フォローアップ時における、うつ重症者の割合が約 1/3（実際にはオッズ比 0.35）に有意に抑える効果が認められた。不安については両者とも介入直後には効果を認めないが、3 か月後フォローアップでは改善傾向が見られた。iCBT と iCBT-AI では、連続アウトカムでは大きな効果の差が認められなかったものの、iCBT-AI では有意に低い脱落率が得られたことを考慮すると、より多くの抑うつ者に効果を提供できる可能性は示唆された。

本考察では、1) なぜ、iCBT-AI では脱落が低かったのか、2) なぜ、iCBT-AI が iCBT の短期的効果を毀損したのか、3) なぜ、iCBT-AI では iCBT には認められなかった 3 か月後フォローアップ時の重病者割合を防ぐことに寄与したのか、の 3 点について言及したい。

1) なぜ、iCBT-AI の脱落が低かったのか？

脱落をいかに抑えるかは、iCBT の普及という観点からも極めて重要な課題と考えられている。脱落の理由として、先行研究は、使い方が難しい、コンピュータがいやだ、課題が多すぎる、症状が改善しない、他の治療を受けることになった、気分がよくなってしまった、時間がない、プログラムがよく理解できないといった理由を報告している[21-23]。筆者(宗)による先行研究における iCBT 群における終了後アンケートで最も多かった自由回答は、自分でやっていることが本当に正しいのか不安があり、できればもっとフィードバックが欲しいというものであった。また先行研究では、脱落が高ければ効果量が低下するとされており、本研究においても、効果が感じられれば利用者の動機づけが高まり脱落率は下がることが想定されていた。しかし、実際には、iCBT-AI では iCBT 群と異なって全体での抗うつ効果が得られなかったのにも関わらず、iCBT 群よりも脱落率が低かったという結果が得られた。これは非常に興味深い点である。

本研究における通常の iCBT と iCBT-AI の差は、NLP による共感機能の有無か事実認識の明確化促進機能の 2 点しか存在しない。そのため脱落に差が出る原因としても、それらのいずれか、もしくは両者が脱落低下への寄与をもたらしたと考えられた。

たとえば、AI の有無により満足度に差が出た可能性も考えられたが、満足度の指標である CSQ-8-J の得点においては iCBT-AI 群と iCBT 群の間には有意差が認められなかったため、AI 機能が満足度に寄与した可能性は低く、満足度の差が脱落の差に反映されたとは考えにくかった。そのため、AI 機能の共感、事実の明確化、もしくはその両者がどのように通常の iCBT の脱落を低下させることに寄与したのかが関心事となるが、現状ではどちらの効果もより主として脱落低下に寄与したかははっきりせず、得られたデータから今後に予定している二次解析の結果が待たれる。

2) なぜ、iCBT-AI が iCBT の短期的効果を毀損したのか？

今回の結果より、iCBT の持つ短期的な抗うつ効果が AI 機能の存在によって毀損されていた可能性が示された。繰り返しになるが、本研究における通常の iCBT と iCBT-AI の差は、NLP による共感機能の有無か、事実認識の明確化促進機能の 2 点しか存在しない。この点を踏まえて、iCBT-AI が通常の iCBT と比べて効果を上げられなかった原因についていくつか推測すると、以下のものが考えられる。第一に、AI による共感の言葉が検討違いだったために利用者が失望感を抱き、モチベーションを低下させた可能性が考えられる。セラピーに限らず、現実の我々の人間関係においても共感は会話の潤滑油のようなもので、会話に共感が不十分、もしくはその共感が検討違いであるならば、相手に対してはそれ以上コミュニケーションを継続しようというモチベーションが著しく阻害されることは心理学的にもよく知られ、それは心理療法でも鍵となる部分である[24, 25]。

第二に、NLP による事実の明確化作業が利用者に対する何かしらのジャッジメント（口語的に言えば「ダメ出し」）として利用者に受け止められ、不快感を生じさせた可能性が考えられる。たとえば人と人による対面での認知行動療法ではクライアントがたとえ間違えた理解をしていたとしても、セラピストは安易にその間違いを指摘するのではなく、適切な質問を繰り返す中でクライアント自らがその間違いに気づけるように導くアプローチが重視され、「誘導された発見」や「ソクラテス的問答」と呼ばれる[26]。これは、個人の有する考えや価値観を他人が一方向的に議論や説得で変更を促すやり方ではうまくいかないことが知られているからである。本 AI におけるフィードバックでもいきなり指摘をするようなことはせず、「そのような記載よろしいのでしょうか？」と問いかけるシステムにはなっていたが、それでも「誘導された発見」や「ソクラテス的問答」といったアプローチを中核技術として重視する認知行動療法の観点からは、配慮が不十分な問いかけに対して利用者がきつい「ダメ出し」と受けとった可能性は考えられる。

第三に、事実の明確化作業自体が認知行動療法の効果を毀損した可能性が考えられる。必ずしも状況を明確化すること自体が認知行動療法における認知再構成において正しい作業とは限らず、利用者に自由に記載してもらった方が少なくとも短期的な効果を得るのには有効なのかもしれない。逆に、それが長期的な効果の取得には不可欠とも考えられる。これは、認知行動療法そのものにとっても興味深い点であり、どちらにしろ、このあたりも予定されている二次解析により詳細な解明が期待される。

3) なぜ、iCBT-AI では iCBT には認められなかった 3 か月後フォローアップ時の重病者割合を防ぐことに寄与したのか？

本研究で非常に興味深いもうひとつの点は、短期的には効果の得られていなかった

iCBT-AI が介入終了後 3 か月というタイムラグを経て遅発性のうつ病の重症者の減少が認められたという点である。元々、iCBT は先行研究におけるメタ解析でも長期的な効果では有意差が示されておらず、iCBT の大きな課題として考えられていた[5]。どうしてこのようなタイムラグを有する遅効性効果が iCBT-AI 群のみに生じたのであろうか？

ひとつには、エクササイズ実施者間に脱落の差があったことから、介入終了後の 3 か月フォローアップ期間中における自主的なエクササイズ、いわば自主練の実施数に差を生じていた可能性が考えられた。本研究では、介入終了後にも 3 か月間のフォローアップ期間中、iCBT-AI 群および iCBT 群は、それぞれ自由にエクササイズが可能であった。しかし、調べてみるとフォローアップ期間中の両者の使用頻度には有意な差は認めず、単純に AI 機能がモチベーションを高めることでフォローアップ期間中のエクササイズ回数を増やした結果、iCBT-AI 群の“自主練”が iCBT 群に比べて多かったから、とは単純に原因づけにくいと考えられた。

他には、AI 付きのエクササイズが CBT 学習において質的な差を生じせしめ、それが効果発現の差異に至ったという考え方もできる。しかも、それは短期的には即効的な効果が出ないものであり、かつ、長期的に遅効性的効果として表れるものということになる。加えて、それが PHQ-9 における 10 点以上のような抑うつ重症化予防効果にもつながっているというトリッキーな性質のものである可能性が示唆された。

ひとつの仮説として、対面におけるうつ病への心理療法として認知行動療法と並んで高いエビデンスを有するものにクライアントの対人関係的な要因に焦点化する対人関係療法の効能特性が参考になるかもしれない。両治療は、同等の効果を得られるにも関わらずその回復プロセスには違いがあることが知られている。一般に認知行動療法が、比較的即効性はあるがその後の効果の伸びは平坦化しプラトーに達しがちなのに対して、対人関係療法には当初の回復スピードは遅いが徐々に効果を認め出し終了後にもさらに時間差をおいて効果が認めるといふ、いわゆる“*slower action*”という現象が知られている[27-29]。いわゆる対人関係的な要因が何かしらの遅効的な効果発現に寄与した可能性はあるかもしれない。

また、近年、うつ病を発症させないための重要な要因として、逆境で折れない力とも言われる“レジリエンス”という概念が注目されている。米国心理学会 (APA) は、“レジリエンス”を構築するための 10 の方法の中で[30]、考え方や問題解決といった行動面に先んじて人間関係を一番目に挙げている。本研究では、iCBT 群と iCBT-AI 群の差においては、明確化機能によって考え方のスキルに学習強化がなされたか、共感機能のいずれか、もしくは両者によって AI 機能がレジリエンス力構築への働きかけが起きた可能性も示唆される。このあたりも、何がどのようなプロセスでこの遅効性に寄与したかも今後の二次解析では検討したい点である。

一方で、逆に将来的に遅効的な効果を呈しうる素因のある参加者だけがこの AI 機能によって残ったことによる選択バイアスを生じた、いわゆる *reverse causality* (逆の因果関係)

の可能性もある。しかし、本研究は ITT 解析で脱落の影響は最小化されている研究デザインの上、実際にエクササイズにおける iCBT-AI 群の脱落は iCBT に比べて有意に低く、また、これまでそのような先行研究での報告も認めず、積極的には考えにくいかもしれない。これも、今後の二次解析において、エクササイズに残った参加者の iCBT-AI 群と iCBT 群の属性比較も含めて二次解析での検討が待たれるところである。

本研究の限界は、主に3つ挙げられる。ひとつめには、臨床対象群ではなく調査会社を用いた研究デザインということもあり、参加者全体の呈する抑うつが全体的に軽症であったという点である。うつ症状に対する介入においては重症者が多いほど介入への反応性も大きく出現する傾向があるため、このような研究では重症者を多く含めた方が介入に有利な結果が得られると一般的に考えられている。しかし、そのために本研究では層化した下位集団分析も行って見たが、iCBT-AI の iCBT に対する利点は、重症のうつ者よりも、開始時における軽症うつ者の重症者を減らすという点にあることから、軽症うつ者において最もその利点が発揮されることが示されていた。

2つ目の限界としては、高い脱落率が挙げられる。本研究では、研究データ回収率という広義の脱落と、実際の iCBT-AI または iCBT エクササイズ参加からの脱落という狭義の脱落の両方が検討される。脱落率が高いというのは iCBT においては重大な課題と考えられている。本研究では、ITT 解析の趣旨に従った広義の脱落率であるデータ回収率では、iCBT 群での未回収率は 43.8%、iCBT-AI 群での未回収率は 38.6%であり、どちらも待機群に比べて有意に高い未回収率であった。しかし、Richardson らの報告によれば、本研究のような人手によるサポートが皆無である完全自助型の iCBT における介入直後での未回収率は、平均で 74%という高い数となっている[31]。治療者によるサポートが入ると未回収率は 28%、非専門家による支援では 38%まで下がることも示されている。それらに比較すると、本研究における未回収率は、どちらも 5 割を下回っており決して悪いとは言えない。Cochran handbook[32]では、20%を超える未回収率は ITT 解析であっても結果に影響を与える可能性が高く望ましくないとしているが、本研究では解析においても MMRM を採用しているため欠損処理は施行しない解析が可能となったため、従来のように欠損データの影響を最小限に抑えた解析がなされてはいる。特に、初回にエクササイズに参加した人数を母数とした時の平均参加率は、iCBT では 75.1%、iCBT-AI では 84.7%であり、一度でもエクササイズを経験したものが非常に低い脱落に抑えられていたとも解釈できた。これらの原因としては、プログラム自体の問題もあるが、そもそもの母集団が調査会社の登録者である上、開始時の抑うつの程度が軽い参加者が多かったために介入の必要性が低く、他の臨床的集団に比べると参加へのモチベーションが低かった可能性が考えられた。また、3つめの限界としては、参加募集対象者がリサーチ会社登録者であることから、今回得られた結果の一般化するに際しては十分な注意が必要と考えられた。

5. 今後に向けて

本研究から、iCBT の効果を iCBT-AI が毀損していた可能性が示唆されたことから、この点についての何らかの工夫（一律に AI を作動させるのではなく、最初の数セッションは、全員または逆効果が予想される一部のユーザーで AI 機能をオフにする）が今後必要になるかもしれない。本研究では、母集団が多かったために大量のデータの利用が可能なることから、下位集団分析において iCBT-AI における AI フィードバック機能のどの要因が効果を高め、どの要因が効果を下げたのかについて、今後の解析が望まれる。

将来的には、今回のような NLP に限らず、さまざまな AI 技術の広い応用が iCBT においても予想され、米国でも 2016 年 4 月に iCBT-AI を用いた研究の研究計画の論文が報告されている[9]。本研究は、今後多くの研究が登場すると思われる AI 技術を使った iCBT の効果を世界で初めて検証したものであり、その第一歩としての意義があると考えられる。一方で限界も少なくない。今後、NLP 以外の AI アプローチも含めて、さらに AI を用いた心理的介入に関する多様な試みとその検証が期待される。

6. 利害相反：本演題に関して、筆頭研究者に開示すべき利益相反はない。

7. 謝辞

本研究の遂行にあたり、多岐にわたるご助言を頂いた一般社団法人認知行動療法研修開発センター理事長の大野裕氏に深く感謝申し上げます。

引用文献

1. 厚生労働省. 自殺・うつ対策の経済的便益 (自殺やうつによる社会的損失). 2010 [cited 2016 6/19]; Available from: <http://www.mhlw.go.jp/stf/houdou/2r9852000000qvsy.html>.
2. 黒田祥子 and 山本勲. 企業における従業員のメンタルヘルスの状況と企業業績－企業パネルデータを用いた検証－. 2014 [cited 2016 6/19]; Available from: <http://www.rieti.go.jp/jp/publications/dp/14j021.pdf>.
3. 宗未来 and 渡部卓. 未病うつ (*Non-clinical depression*) に対する低強度メンタルヘルス・サービスにおける積極的な民間活力導入の提案：趣味を実益に変えて、医療負担から戦略的事業へ *RIETI* ポリシーディスカッションペーパー. 2014; Available from: <http://www.rieti.go.jp/jp/publications/pdp/14p001.pdf>.
4. NICE, *Depression in adults: recognition and management*. National Institute for Health and Clinical Excellence. 2009.
5. So, M., et al., *Is computerised CBT really helpful for adult depression? A meta-analytic re-evaluation of CCBT for adult depression in terms of clinical implementation and methodological validity*. BMC Psychiatry, 2013. **13**(1): p. 1-14.
6. 宗未来. 成人うつに対するコンピュータ認知行動療法 (*CCBT*) の臨床効果、及び費用対効果についての系統的レビュー *RIETI* ディスカッション・ペーパー. 2013 [cited 2016 6/19]; Available from: <http://www.rieti.go.jp/jp/publications/dp/14j003.pdf>.
7. Lifschitz, V., *John McCarthy (1927-2011)*. Nature, 2011. **480**(7375): p. 40-40.
8. Mitkov, R., *The Oxford Handbook of Computational Linguistics*. 2005: OUP Oxford.
9. Piette, J.D., et al., *Patient-Centered Pain Care Using Artificial Intelligence and Mobile Health Tools: Protocol for a Randomized Study Funded by the US Department of Veterans Affairs Health Services Research and Development Program*. JMIR Res Protoc, 2016. **5**(2).
10. Burns, D.D. and S. Nolen-Hoeksema, *Therapeutic empathy and recovery from depression in cognitive-behavioral therapy: A structural equation model*. Journal of Consulting and Clinical Psychology, 1992. **60**(3): p. 441-449.
11. McCullough, J.P., Jr., *Treatment for chronic depression using Cognitive Behavioral Analysis System of Psychotherapy (CBASP)*. J Clin Psychol, 2003. **59**(8): p. 833-46.

12. Kroenke, K., R.L. Spitzer, and J.B. Williams, *The PHQ-9: validity of a brief depression severity measure*. J Gen Intern Med, 2001. **16**(9): p. 606-13.
13. 村松公美子, *Patient Health Questionnaire (PHQ-9, PHQ-15) 日本語版および Generalized Anxiety Disorder -7 日本語版—up to date—*. 臨床心理学研究, 2014. **7**: p. 35-39.
14. Rush, A.J., et al., *The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression*. Biol Psychiatry, 2003. **54**(5): p. 573-83.
15. 藤澤大介, *日本語版自己記入式簡易抑うつ尺度 (日本語版 QIDS-SR) の開発*. ストレス科学, 2010. **25**: p. 43-52.
16. Spitzer, R.L., et al., *A brief measure for assessing generalized anxiety disorder: the GAD-7*. Arch Intern Med, 2006. **166**(10): p. 1092-7.
17. Sheehan, D.V., *The Anxiety Disease*. 1983: Scribner.
18. 吉田卓史, et al., *Sheehan Disability Scale(SDISS)日本語版の作成と信頼性および妥当性の検討*. 臨床精神薬理, 2004. **7**(10): p. 1645-1653.
19. 立森久照, *日本語版 Client Satisfaction Questionnaire 8 項目版の信頼性および妥当性の検討*. 精神医学, 1999. **41**(7): p. 711-717.
20. Gilbody, S., et al., *Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial*. BMJ, 2015. **351**.
21. Andersson, G., et al., *Internet-based self-help for depression: randomised controlled trial*. Br J Psychiatry, 2005. **187**.
22. Proudfoot, J., et al., *Clinical efficacy of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial*. The British Journal of Psychiatry, 2004. **185**(1): p. 46-54.
23. Warmerdam, L., et al., *Internet-based treatment for adults with depressive symptoms: randomized controlled trial*. J Med Internet Res, 2008. **10**.
24. Moyers, T.B. and W.R. Miller, *Is Low Therapist Empathy Toxic? Psychology of addictive behaviors : journal of the Society of Psychologists in Addictive Behaviors*, 2013. **27**(3): p. 878-884.
25. 堀越勝, *ケアする人の対話スキル ABCD*. 2015: 日本看護協会出版会.
26. Braun, J.D., et al., *Therapist use of Socratic questioning predicts session-to-session symptom change in cognitive therapy for depression*. Behav Res Ther, 2015. **70**: p. 32-7.

27. Stangier, U., et al., *Cognitive therapy vs interpersonal psychotherapy in social anxiety disorder: A randomized controlled trial*. Archives of General Psychiatry, 2011. **68**(7): p. 692-700.
28. Elkin, I., et al., *National Institute of Mental Health Treatment of Depression Collaborative Research Program. General effectiveness of treatments*. Arch Gen Psychiatry, 1989. **46**(11): p. 971-82; discussion 983.
29. Agras, W.S., et al., *A multicenter comparison of cognitive-behavioral therapy and interpersonal psychotherapy for bulimia nervosa*. Arch Gen Psychiatry, 2000. **57**(5): p. 459-66.
30. American Psychological Association (APA). *10 ways to build resilience*. 2014 [cited 2016 6/19]; Available from: <http://www.apa.org/helpcenter/road-resilience.aspx>.
31. Richards, D. and T. Richardson, *Computer-based psychological treatments for depression: a systematic review and meta-analysis*. Clin Psychol Rev, 2012. **32**(4): p. 329-42.
32. Higgins, J.P.T., J.J. Deeks, and D. Altman. *Intention to treat issues: Special topic in statistics.* " In *Cochrane handbook for systematic review of interventions: version 5.0.2*. 2009 [cited 2016 6/19]; Higgins, J. P. T & Green, S.: [Available from: <http://www.mrcbsu.cam.ac.uk/cochrane/handbook/>].

図 4 試験のフローチャート

