

厚生労働科学研究費補助金（障害者対策総合研究事業 身体・知的等障害分野） 分担研究報告書

支援機器の実証試験における研究デザイン II 実証試験に関する論文に関する検討

研究協力者 山内 繁 NPO 支援技術開発機構 理事長

研究要旨

支援機器の実証試験における科学性は堅固な研究デザインを中心として構成されなければならない。昨年度に引き続き、生物医学領域における臨床試験の第0相、第I相、第II相、第III相、第IV相に対応した支援機器のための実証試験の各段階についての位置づけを検討した。我が国の支援機器の実証試験における問題点を明らかにするために、リハ工学カンファレンス、日本生活支援工学会誌に発表された支援機器の実証試験に関する論文について、研究デザインとエンドポイント、倫理審査の状況の調査を行った。この結果をJRRD（Journal of Rehabilitation Research and Development）所収の論文と比較した。問題点として、我が国における論文は対象のない介入の研究デザインが多く、また、主観評価に頼りがちであることが明らかになった。より具体的に適切な研究デザインの選択に資するため、支援機器に応じて適切な研究デザインの類型化を行い、仮説とエンドポイント、予測因子とアウトカム、選択/除外基準についても検討した。これらの成果に基づき支援機器開発における研究倫理マニュアルの研究デザイン部の骨格を策定した。

A. 研究目的

臨床研究における科学性を担保する基本は研究デザインにあり、適切な研究デザインを選定し、運用することが科学的に臨床評価を行うための基本である。

我が国において従来主として行われてきた研究デザインは、当事者による試用と主観評価の聞き取りであった。開発者による主観評価で済ませた場合も珍しくない。ユーザビリティやQOLへの寄与が問題になるとともに、このような傾向が増大しているように思われる。

しかし、このような主観的な評価ではバイアスが入りやすく、性能と効果・効能を客観的に主張するのは困難である。

科学性を担保した支援機器の実証試験を普及させるために、適切な研究デザインの選択と運用のための整理を行うことを目的とする。

B. 研究の方法

昨年度に引き続き、様々な研究デザインについて再検討を行うとともに必要な補足を行う。研究デザインの分類は昨年度の検討に基づき、代表的な教科書¹⁻³⁾からの研究デザインを前提とする。これに基づいて支援機器開発に関する論文を解析し、問題点を摘出する。さらに、機器に適切な研究デザインの類型化を行う。

研究デザインが適切なタイプであっても有効性を発揮するためには適切なエンドポイントと

の組み合わせが必要である。そのための指針を策定するとともに、実証試験マニュアルの研究デザイン骨格を構築する。

C. 研究成果

C-1 支援機器開発の相

昨年度の報告書においては、生物医学領域から学ぶ物として、開発段階における「相(Phase)」の概念を支援機器開発にも取り入れることを提案した。しかし、支援機器開発に即したあり方については検討が十分であるとは言えなかった。

本年度さらに検討を加えた結果、支援機器開発については、表1のようにまとめた。第II相で10名程度、第III相で20-40名程度の被験者数を想定したが、統計学的有意性の面からはサンプルサイズとして小さいとの批判を避けることはできない。しかし、支援機器の場合、期待される販売台数から考えて多数の被験者を想定することは現実的ではない。数台の試作機を使いまわして、長期貸し出しをするなら時間がかかりすぎるし、数十台の試作機を準備するのはコストがかかりすぎる可能性がある。このように考えると、上記被験者数は現実的には実行可能な最大数に近いと考える。

実証試験に関してはオランダのZuyd大学のJan Gelderblomも同様のものを発表しており⁴⁾、期せずしてヨーロッパと同様な開発の相モデルを提案したことになる。

表 1 支援機器実証試験における相の提案

相	主要点	定義
第 0 相試験	開発着手以前の観察研究	現在使用中の支援機器、支援を要する心身機能の特性、必要な支援の解明など。主として介入のない観察研究。現在使用中の機器を対象とした最小限の介入を含む。
第 I 相試験	健常成人による最初の試験	安全性の確認、安全上の問題点の把握。さらに研究を進めることが適当かどうかを決定するための探索的試行を含む。
第 II 相試験	利用者を被験者とした早期試験	想定する利用者の数名から 10 名程度の被験者によるパイロットテスト。集会などにおける不特定多数の利用者による短時間の試用による適合・選好調査を含む。
第 III 相試験	上市前の有用性・安全性の最終試験	様々な条件下で 20-40 名程度の被験者による本格試験。有用性、適応、適合の実証を目的とする。
第 IV 相試験	市販後のフォローアップ	有害事象のモニタリング、適応、応用範囲の拡大、適合のための方法の検討など

C-2 研究デザインの検討

研究デザインに関しては、本質的には昨年度の報告書におけるものと変わりはない。代表的な研究デザインを表 2 に示す。昨年来の検討の結果、変更したものは、「前後比較」の代わりに「自己対照」を用いることとした点である。これは、「前後比較」の用語に対しては何を比較し、比較しないのかが明確でなくなる場合があったためである。「自己対照」を用いることによって、同じ被験者の異なった介入に対する反応を観測することを強調することができる。表 2 に示した研究デザインの具体的内容については昨年度の報告書に記載したので再掲はしない。

C-3 わが国の実証試験の現状解析

我が国における実証試験について、ゴールが明確でない、主観評価に頼りすぎているのではないかと懸念があった。これを検証するための調査を行った。対象としたのは、リハ工学カンファレンスの 2007 年から 2011 年までの 5 年分の予稿集に収録された実証試験に関する論文 137 編、日本生活支援工学会誌の第 1 巻から第 10 巻までに掲載された支援機器の実証試験にかかわる論文 10 編である。さらに、国際的見地からの比較のために、2011 年発行の JRRD (Journal of Rehabilitation Research and Development, Department of Veterans Affairs) の中から支援機器の実証試験に関する論文 32 編を選び出した。

こうして、3 つの論文誌掲載の論文に関し、研究デザイン、エンドポイント、倫理審査に関する記載について調べた。リハ工学カンファレンス論文に関する結果を表 3 に、日本生活支援工学会誌

表 2 代表的な研究デザインのタイプ

<p>観察研究</p> <p>(1) 記述的研究</p> <p>(2) 分析的研究</p> <p>(a) コホート研究</p> <p>(b) 横断研究</p> <p>(c) ケース・コントロール研究</p> <p>(d) 複合的デザイン</p> <p>実験的 (介入) 研究</p> <p>(1) 比較対照試験</p> <p>(a) パラレル</p> <p>i. ランダム化</p> <p>ii. 非ランダム化</p> <p>(b) 逐次</p> <p>i. 自己対照 (前後比較)</p> <p>ii. クロスオーバー</p> <p>(2) 対照なしの研究</p> <p>(3) シングルケースの研究</p>

の論文に関する結果を表 4 に、JRRD の論文に関する結果を表 5 に示す。

開発段階で「開発症例」は利用者のニーズに基づいた開発で、被験者数も 1 名から数名までの小規模な開発研究である。「開発事例」はシーズから出発したもので、ハードウェアの開発に重点があり、被験者実験ではないものも含む。

表3 リハ工学カンファレンス発表支援機器開発論文(2007-2011)の解析結果

開発段階	第 I 相	第 II 相	第 III 相	第 IV 相	開発症例	開発事例	計
A. 研究デザイン							
自己対照	2	4	1	1	1	2	11
対照のない介入	16	17	5	4		20	62
記述的横断研究				2		2	4
症例報告					41		41
開発研究					3	13	16
B. 評価関数(エンドポイント)							
客観指標のあるもの	10	8	1	2	5	11	37
被験者による主観評価	5	8	3	3	7	6	32
研究者による主観評価		2	1		20	8	31
中間ユーザによる主観評価	1	1					2
エンドポイント無定義	1	1			4	3	9
評価のないもの	1	1	1	2	9	9	23
C. 倫理審査							
倫理審査の必要な研究課題							
承認済みの記載	2					1	3
インフォームドコンセント		6		1		3	10
記載なし	16	15	6	5		25	63
(ミニマムリスクと推定)	(2)	(2)					(4)
あらかじめの倫理審査が不要な研究課題							
診療					45	4	49
行政目的				1			1
機器開発						4	4
計	18	21	6	7	45	37	134

表4 日本生活支援工学会誌発表支援機器開発論文(2002-2011)の解析結果

開発段階	第 0 相	第 I 相	第 II 相	第 III 相	第 IV 相	計
A. 研究デザイン						
自己対照	2	3	2	1		8
対象のない介入		2				2
B. 評価関数(エンドポイント)						
客観指標	1	3	2	1		7
主観指標	1	2				3
C. 倫理審査						
倫理委員会承認	1	1				2
インフォームドコンセント		1				1
「実験の説明」の記載		1	1	1		3
記載なし		2	2			4
計	1	5	3	1		10

表 5 JRRD 発表支援機器開発論文 (2011) の解析結果

開発段階	第 0 相	第 I 相	第 II 相	第 III 相	第 IV 相	基礎研究	開発事例	開発症例	計
A. 研究デザイン									
RCT	1		1						2
比較対照	1	1				1			3
自己対照	1	1	2			8		1	13
クロスオーバー						1			1
シングルケーススタディ	4	1	2			1		1	9
対照のない介入			1		1		1		3
ハードウェアのみ								1	1
B. 評価関数 (エンドポイント)									
客観量	4	2	2		1	7	1	2	19
主観評価	2								2
確立したスコア	1		4			4			9
評価のないもの		1						1	
計	7	3	6	0	1	11	3	1	32

C-3.1 研究デザイン

表 3 の研究デザインにおいて、ランダム化対照試験(RCT)が 1 件もなかったのはある意味では当然であるが、自己対照の研究デザインが総数で 11 件であって、対象のない介入の 62 件と比べてあまりに少ないのは驚きであった。一方、表 4 においては、自己対照 8 件に対して対象のない介入が 2 件となっており、研究デザインへのより積極的な取り組みが見受けられる。この相違は、リハ工学カンファレンスにおいては論文審査がないのに対し、日本生活支援工学会誌においてはピアレビューのシステムによっているため、科学性の乏しい論文は掲載されないからであろう。

表 5 の JRRD の論文においては、RCT が 2 件含まれており、自己対照が最も多く 13 件となっているが対照のない介入は 3 件に過ぎず、全体として研究デザインへの積極的な配慮が伺える。当然、JRRD の論文も審査を経ているためアカデミックレベルの低い論文は掲載されないことが大きく作用していると考えられる。

C.3.2 評価関数の検討

次に評価関数について検討した。表 3 では「客観指標のあるもの」が 37 件、「何らかの主観評価」によるものが 65 件となっており、エンドポイントの定義されていないのが 9 件もある。ここで注意しておきたいのは、「客観評価のあるもの」は主要エンドポイントが客観指標となっているこ

とを意味するのではなく、何らかの客観指標を計測しているものを含ませてあり、かなりの件数は主観評価をエンドポイントとしている。一方、表 4 においては、主たるエンドポイントが客観指標であるものが 7 割に達しており、よりアカデミックな論文として仕上げる論文は科学性を重視してきたものと解釈できる。

この傾向は表 5 においてはより顕著であり、主観評価によっているものは 32 件中 2 件にすぎない。「確立したスコア」というのは、主観評価、ユーザビリティなどについて既に確立している調査票のスコアを用いるもので、主観内容を問う場合にも客観指標としてとらえていることを示している。

C.3.3 倫理審査の検討

倫理審査、倫理的配慮に関する記述を見てみると、表 3 のリハ工学カンファレンスにおいては倫理審査が必要な研究と診療などのようにあらかじめの倫理審査は不要と判断される案件とに区分して検討した。「審査不要」と判断したのは主として診療目的の行為であって症例報告と考えられたからである。なお、個々の症例の報告を超えて、一般的知見を導く内容の場合は(事後的な)倫理審査が必要であるが、この点の判断は行わなかった。

リハ工学カンファレンスにおいて、倫理審査が必要と判断された論文 76 件中審査済みとの記載

が3件であり、インフォームドコンセントに言及したのが10件に過ぎず、大多数が何も記載がなかったというのはこの分野における研究者の倫理面への関心の薄さを示している。

日本生活支援工学会誌においても、半数近くが何も記載していないのは気になることである。なお、表5は10年間近くにわたる論文を対象としているため倫理審査への関心の薄かった時代を含むことを付記しておく。

JRRDについては、投稿規定に倫理審査によって承認されていることが投稿条件になっているので、全部の論文について承認済みとの記載が含まれている。このため、倫理審査に関する集計は行わなかった。

C3.4 支援機器開発の革新のために

以上、我が国の支援機器開発における実証試験が科学性、倫理性において欠陥を有している実情を見てきた。これはそのまま放置すべきではなく、科学的実証試験の普及を急ぐ必要がある。

実際の開発現場の技術者、臨床現場の専門職と話してみると、生物医学系における研究のあり方に関する教育を受けておらず、どうして良いか判らないというのがほとんどであった。このような現状の革新のために、科学性、倫理性の両面において、問題の核心をわかりやすく解説、普及につとめることが緊急の課題である。

C.4 研究デザインの適切な選択のために

昨年度の報告において、支援機器の実証試験において一般に最も使いやすい研究デザインは自己対照のデザインであると結論した。このことはそれだけでは容易には理解していただけないことが多い。機器の類型に応じた研究デザインを考察する。

この目的のために、支援機器を以下の3つの類型に区分して考察する。

- 身体構造・機能を代替・補完する支援機器
- 特定の活動を支援する支援機器
- 心身の教育・訓練機器

C.4.1 身体構造・機能を代替・補完する支援機器

身体構造、身体機能を代替・補完する支援機器は義肢装具によって代表される支援機器であって、補聴器、人工喉頭など装着型が多いが、拡大読書機のように装着型ではない場合もある。この場合、最も使いやすい研究デザインが自己対照である点は変わらない。機器の機能について、開発した機器と既存の機器の性能を比較することに

よって、開発した機器の性能を示すことができる。すなわち、機器の機能に関する性能を代表するパラメータをエンドポイントとして既存機器と開発した機器とを比較する。

一般に機器類と身体構造とを比べたとき、機器はそれによって代替・補完した身体構造に比べてその機能が制限されることがあり、完全には代替・補完できないのが普通である。自己対照の研究デザインを策定するに当たっては、機器が人体に由来する様々なタスクを可能とするので様々なエンドポイントの選択が可能となる。主たる使用者の障害像から考えて、使用者の日常生活において重要なタスクの作業能率をエンドポイントとして選び、既存の機器による場合と開発した機器による場合の作業能率を比較する。

支援機器の機能が身体構造の機能に比べて制限される場合は、そのような制限をも視野に入れた評価が必要である。

装着型の機器の実証試験においては一般に被験者に対する適合が必要であり、測定された作業能率が機器の性能を代表しているのか適合上の問題を表しているのか区別しがたい場合もある。機能性のある義肢装具完成用部品を開発した場合、組み立てて完成したユニットを用いた試験を行うわけであるが、一般には他の部品要素の影響もあり、開発部品の性能評価を感度よく行うのは困難である。製作者の技倆によることもあるので、可能な限り条件をそろえた実験計画を立てる必要があろう。

人体機能を代替・補完する支援機器の場合、機器に求められる機能は限定されているので、代替・補完している機能を表現するパラメータをエンドポイントとして、対照となる機器によるものと比較するのが自然である。

C.4.2 特定の活動を支援する支援機器

ほとんどの支援機器はコミュニケーション、モビリティ、パーソナルケアなど特定の活動を支援するものである。

この場合も最も使いやすい研究デザインは、自己対照のデザインであって、既存の機器と開発した機器との間で、同一被験者による作業能率を比較する。この場合、比較の対象となる機器は標準的に使われているものを採用すべきであって、陳腐化してあまり使われなくなっていたりした物を採用すべきではない。

エンドポイントとしては、当然作業能率を表すパラメータが採用される。最も普通に使われるのはタスクの遂行時間である。タスクの種類によっ

てはそれ以外に巧緻性や正確さにかかわるパラメータが選ばれることもある。これらの候補の中から最も重要なものを主要エンドポイント、他のものを副次エンドポイントとして選択する。

同じ活動を支援するための支援機器が存在しない場合も考えられる。これまでにない新規の支援機器を開発した場合である。そのような場合にも似た活動を遂行するための手段が存在していることが多い。しばしば聞かれるのは、全く同じ作用をする機器は存在しないから比較の対象機器が存在しないという思い込みである。しかし、日常生活で重要な活動であれば、何とか工夫して遂行しているはずである。それが開発した機器と同じ機構で作動するとは限らない。十分な調査なしに同じ活動を支援する機器が存在しないとして対照のない介入の研究デザインを選択するのは怠惰のそしりを免れない。

C.4.3 心身の教育訓練機器

教育訓練機器は不適切な研究デザインが採用されている最も代表的な支援機器である。最も単純に「前後比較試験」として、開発された筋力訓練機器による訓練期間の前後における筋力測定によって筋力が増加したとしよう。この研究デザインによってこの訓練機の「有効性が立証」されたといつて良いであろうか？

単純に考えて、何らかの筋力訓練を行えば筋力を増強する（あるいは減退を阻害する）効果があるのは当然である。筋力増強の程度が定量的に表現されていないため、たとえば1日1万歩程度散歩していれば獲得できるほどの筋力であるのかどうか比較のしようがない。実効的には意味のない実験である。効果の程度は対象とする被験者群によっても違ってくるので、少なくとも、被験者群に対して（現在標準的と考えられている訓練法と比較することが必要である。

自己対照のデザインを採用して、標準的な訓練法の前後、開発した訓練機器による訓練の前後での成績を比較することを考えてみる。この場合、どちらを先に訓練するかで効果が変わってくる可能性がある。そうすると、2群に分けたクロスオーバーまたはパラレル対照、さらにはランダム化したRCTに依らざるを得なくなる。

この場合、よほど選択条件をそろえないとランダム要因が大きくなり、科学性を主張するためには膨大なサンプル数が必要となる可能性がある。

訓練機器の場合、さらに検討を要するのはリハビリテーション訓練を支援するための機器の場合である。急性期から亜急性期の場合は最終的に

はRCTに依らざるを得ないであろう。急速に回復している期間においてはクロスオーバーの意味はあまり大きくないであろうし、被験者要件を平均化するためにはランダム化が欠かせない。

さらに、倫理上の要請として、実証試験への参加によって現在加療中の医療の質の低下を招くことがあってはならない。たとえば、ヘルシンキ宣言第32項は次のように述べている。

新しい治療行為の利益、リスク、負担および有効性は、現在最善と証明されている治療行為と比較考慮されなければならない。

すなわち、比較すべき治療法あるいは教育訓練法は「現在最善と証明されている治療行為」なのである。これは、試験に参加することによって被験者から治療を受ける権利を剥奪することは倫理的ではないためである。通常のリハビリテーション計画の一部を新規の機器によるものに置き換えることによって被験者を不利な条件に置くことになってはならない。

上に述べた様々な要素を考慮し、科学的にも倫理的にも受け入れられ、被験者募集やコスト面でも実行可能である研究デザインを採用することになる。

【文献】

1. John I. Gallin, ed. (井村裕夫、竹内正弘、藤原康弘、渡辺亨監訳). NIH 臨床研究の基本と実際、丸善、p. 206、2004.
2. Stephen B. Hulley 他(木原雅子、木原正博訳): 医学的研究のデザイン第3版. メディカル・サイエンス・インターナショナル. 2009.
3. E. G. DeRenzo and J. M. Moss: "Writing Clinical Research Protocols: Ethical Considerations", p.57-8, Elsevier, 2006.
4. Jan Gelderblom: "Effect Evaluation of Assistive Robots", 福祉機器の実証試験を考えるシンポジウム, 国立障害者リハビリテーションセンター, 2013年2月29日.
5. 中山健夫: 健康・医療の情報を読み解く, p15, 丸善, 2008.
6. 福原俊一: リサーチ・クエスションの作り方, NPO 健康医療評価研究機構, 2008