

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

支援機器の臨床評価における研究デザイン I 支援機器臨床評価に即した検討

研究協力者 山内 繁 NPO 支援技術開発機構 理事長

研究要旨

支援機器の臨床試験における科学性を担保するための基礎は堅固な研究デザインを中心として構成しなければならない。研究デザインとしては、生物医学領域で開発されてきた臨床研究における研究デザインを参考とすべきである。そのために、支援機器に関わる研究の生物医学的側面、ユーザへの適合、生命倫理、社会的側面等について検討した。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臨床試験の各段階についての位置づけを明らかにした。また、これらの相と倫理審査、迅速審査の要件についても検討した。

支援機器の臨床試験のための研究デザインの検討のために、まず、研究デザインを記述するためのパラメータとして、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、支援機器の臨床試験における意義、これらのパラメータに対する要件についての検討を行った。次いで、観察研究、介入研究において生物医学領域において用いられてきた研究デザインについて、支援機器の臨床試験に適用可能であるかどうかについて検討した。最も重要な点は、支援機器に関しては盲検化が使えないこと、群間比較を行うのも困難であることである。このため、生物医学領域で最も強力であるとされるランダム化群間比較デザインが適用できない点である。支援機器の臨床試験に使いやすい研究モデルとしては、自己対照（前後比較）試験であることを指摘するとともに、特定の個人に即した支援機器の開発においてはN of 1 デザインが有用である可能性を指摘した。

A．背景と目的

人を対象とする研究に関する倫理審査においては、研究の科学性、倫理性、安全性が評価の主要なポイントになる。このうち、工学系の研究者にとって一番困難であるのが科学性であり、そのために支援機器の、臨床評価においては「評価方法がはっきりしていない」などの問題提起がなされている。

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。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、客観性を主張するのは困難である。特に、大学や公的な研究機関において開発された機器に関して、その開発者が研究者として臨床評価を行う場合、謙遜バイアスや追従バイアスを伴いがちであり、その結果だけを信用すると、「臨床評価では高く評価されたが、結局は使われなかった」という結果を招きがちである。

支援機器の臨床評価における科学面の含有する様々な要因の理解を深めるとともに、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．支援機器開発に関わる研究の諸側面

支援機器の開発は、工学の関わる通常の機器の開発とは異なった側面を持っている。それは、生物医学的側面、ユーザへの適合の側面、生命倫理的側面、社会的側面である。

これらの側面は他の分野における機器開発においては通常は問題になることのないものであり、工学の立場からのアプローチにとっては障害となっていると考えられる。しかし、これらの側面を有することへの理解が第一に必要である。それなしには工学者にとっての障害を克服することは困難であろう。

A) 生物医学的側面

支援機器はどのようなものであれ、高齢者や障害者が使用することを想定している。それも、心身の機能障害を補い、活動の制限を取り除くとともに社会参加を推進することを目的としている。

ここで基本となるのは、第一に機能障害の補償であり、それを可能とするのが生物医学的側面で

ある。これは、ICFにおける医学的アプローチであるといつて良い。

換言すれば、支援機器の動作原理においては、生物医学的に妥当であることが第一に求められる。生物医学的側面はこの意味で研究デザインの基礎をなしている。

ここで研究と診療の関係についてのベルモントレポート¹⁾について言及しておきたい。タスキギー梅毒事件などを受けて1974年にアメリカで制定された国家研究法に基づいて「生物科学・行動科学研究における被験者保護のための国家委員会」が設置され、この委員会によって1979年に公表されたのがベルモントレポートである。現在のアメリカ政府の人を対象とする研究に関する基本原則はこのレポートに基づいている。

ベルモントレポートは、診療と研究の関係について以下のように述べている。

「診療」(practice)行為の目的は、特定の個人々人に対して診断を与え、予防的処置や治療を加えることを意味する。これに対し「研究」(research)は、仮説を検証し結論を導き出せるようにし、そこから一般化できる知見(それは例えば理論、原則、関係性についての言説などによって表現される)を見出す、もしくは見出す契機となるように考案された行為を称するものである。

支援機器の場合、「診療」には、義肢装具の処方、適合などの他、支援機器に関するあらゆるサービスが含まれる。専門職による訓練はもとより、ボランティアや専門教育の一環として提供されるサービス(パソコンボランティアや自助具の試作など)が含まれる。

「診療」に関しては、提供するサービスによっては専門的資格が求められるが、それを必要としないものもある。「診療」における倫理上の問題は生命倫理的側面で取り上げるものであって、研究倫理の問題として議論すべきではない。

これに対し、「研究」に関しては、研究倫理の問題としての倫理審査が求められる。注目すべきは、「研究」が「仮説を検証して一般化できる結論を見いだす。」ことであると定義されている点である。支援機器の開発について考えると、機器の有用性、効果に関して「一般化できる」結論を導くことが研究であるとされている点である。

有効性としての「一般化」とは、あらゆる場合に有効であるとの結論を追求することではなく、適応と適合条件、有効性の限界を含めた「一般性ある結論」を追求することを意味している。

実際には、支援機器の臨床評価においては、様々な制約のために、異論のない一般化できる結

論を導くのは困難であることが多い。

最大の要因は、一般に支援機器の市場規模が小さく、一般化が可能であると統計的に結論できる十分なサンプル数をそろえることが困難な点にある。医薬品であれば設備を整えれば数百から数千のサンプルを製造することは困難ではない。しかし、支援機器の試作品をデータロガーと併せて数十セットそろえるのはコスト面からもきわめて困難である。

開発した支援機器に適合のある多数の被験者を募集することが困難な場合も珍しくない。被験者としての可能性のある人数が多い場合にも、実際には合併症や利用環境など他の原因のために被験者としては適切ではない場合が存在するためである。

B) ユーザへの適合の側面

支援機器は一般に、あらゆる高齢者・障害者に有用ではない。利用者の心身障害の様相は様々であり、それぞれに対応した支援機器が求められる。一般に、大量生産になれてきた工学の立場からは、個別対応によるユーザへの適合には不慣れであるため、臨床評価の場面においても適合の観点を見落としがちである。

適合というとき、見落としがちなのが利用環境への適合である。実験室の制御された環境の元では問題とならなかった場合にも、実環境の元では使えないことが珍しくない。

最後に、ユーザの嗜好も無視してはならない。機能は同じでも、デザイン面などでユーザに好まれないものは、結局は使われないままで終わってしまう。ユーザビリティやQOLの向上のためには単なる機能のみにとどまらないことが求められるからである。しかし、開発した機器のデザインに対するユーザの嗜好を一般的に結論することは困難である。この点に関しては、臨床評価の中で主観評価を併用し、ユーザの好みを考慮することになる。重要なことは、そのために生物医学的側面を軽視してはならない点である。

C) 生命倫理的側面

研究倫理は臨床研究が倫理的に行われること、すなわち研究における被験者の保護を主たる目的としている。

これに対し、生命倫理においては、その機器を実用的に使用することが倫理的であるかどうかを問題とするものである。我が国においては、研究倫理と生命倫理とが明確に分離されていないことが多く、新規の診療行為の倫理性と研究計画の倫理性とを混同してしまいがちである。

宇和島徳洲会病院における病気腎移植事件の後、病気腎移植は生命倫理上許されないとしながら、臨床研究としての腎移植は許可するとしている厚労省の決定がその典型である。この場合、研究倫理のための倫理審査委員会が生体移植に関する生命倫理上の要件までも審査することになり、負担が大きい。

支援機器にかんする問題では、パロをデンマークで取り上げるに当たって、最初に生命倫理上の問題についての倫理審査が行われたとのことである。今後新規のロボットなどの導入においては、生命倫理上の検討が必要となるであろう。そのため対策も準備を始める必要があると考える。

生命倫理上のもう一つの問題として、廃用症候群の問題に注意を喚起しておきたい。障害あるいは高齢のために何らかの機能障害があり、支援機器を利用するとき、機器に頼りすぎると残存機能を喪失する廃用症候群（disuse syndrome）を引き起こす可能性がある。一方、機器に頼らず残存機能を使いすぎると残存機能の疲労のために過用症候群（overuse syndrome）を引き起こす可能性がある。

支援機器の利用に当たっては、一般には上記相反するリスクのバランスの上にあることを認識すべきであり、機器の導入によって残存機能を減殺するとすれば、それは生命倫理上も問題であり、新規機器の導入に当たってもそのような可能性についての検討が求められる。

D) 社会的側面

支援機器は高齢者・障害者の社会参加のツールとしての社会的側面を有しており、開発、評価においてもその観点を欠かすことができない。

社会参加としては、近親者、親しい友人との交わりに始まり、特に親しくはない友人から間接的にしか関わりのない人々との関係、買い物や交通機関の利用、会合への出席などがまず想定される。

その社会における文化的価値に関しても配慮が必要である。特に留意すべきはスティグマを引

き起こす可能性に対する配慮である。

また、支援機器自体社会的存在であり、給付制度や公共施設、交通機関へのアクセスなどの制度、環境との関連も無視できない社会的側面である。

C . 支援機器開発における臨床評価の諸側面

生物医学分野においては、新治療法に関する臨床試験（Clinical trial）を第 I 相から第 IV 相の段階によって区分している。第 I 相以後が治験の段階であって、表 1 に示したように一般に区別されている。

新治療法、新規医薬品の開発においては、第 I 相に先立つ予備段階があり、通常は実験室における生体外実験あるいは動物実験による生物学的知見を蓄積し、その成果にたつて安全性、有用性を確認した後、第 I 相試験を開始するのが通例である。

最近 FDA は「探索的新薬研究」の概念を提案し、ごく少数の患者に対する微量の投薬への反応を観察することによって新規医薬（特に抗がん剤）の開発が促進されることを提案した³⁾。この手法は第 0 相試験として受け容れられつつあり⁴⁾、近い将来新たな臨床試験の段階として位置づけられる可能性がある。

なお、生物医学分野においては、臨床研究（clinical research）は人または人由来試料を用いた研究を意味しており、支援機器における臨床評価についても人を被験者とした研究を意味するものとする。

生物医学分野における臨床研究の諸相にあわせて、支援機器開発における臨床評価のための臨床試験の相の考え方について整理する。以下、生物医学領域における相の定義については文献^{2,5)}によりながら検討する。

A) 第 0 相試験

生物医学研究における第 0 相試験とのアナロジーで考えると、介入のない、または最小限の介入を伴う試験を支援機器の第 0 相試験として想定す

表 1 医療における新治療法の相²⁾

相	主要点	目的
第 I 相試験	新治療法の人体・患者への最初の投与	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臨床研究
第 II 相試験	患者における新治療の早期試験	用量反応の相関、毒性の発生率、潜在的なインパクトについての洞察の提供
第 III 相試験	新治療と標準治療との大規模比較試験	現在の標準治療に置き換えるべきかの最終評価。RCT の研究モデル。
第 IV 相試験	新治療法を日常臨床で用いた場合のモニタリング	強固な副作用の発生率など、新治療の疾患に対するインパクトに関する追加情報のための市販後調査

ることになる。具体的には、利用者の障害特性と支援機器へのニーズ、機器に必要な利用者の特性（健常者を含む）に関するデータの収集、評価尺度構築などに関する研究を含む。ほとんどの場合観察研究であるが、現在使用中の機器の問題点を発見するための介入研究の場合もある。

この場合、倫理審査は不要であるとの見解も存するが、倫理審査の要否は試験の相には関係がない。機関長が倫理審査委員会への付議を要しない条件は「臨床研究に関する倫理指針、第2、3、(4)」に規定されているとおりであって、この要件は遵守する必要がある。

(4) 臨床研究機関の長は、臨床研究計画がこの指針に適合しているか否かその他臨床研究の適正な実施に関し必要な事項について、あらかじめ、倫理審査委員会に審査を行わせなければならない。

ただし、次のいずれかに該当する臨床研究計画については、この限りでない。

倫理審査委員会に属する者その他の者のうちから倫理審査委員会があらかじめ指名する者（において「あらかじめ指名する者」という。）が、当該臨床研究計画が次に掲げるすべての要件を満たしており、倫理審査委員会への付議を必要としないと判断した場合

ア他の機関において既に連結可能匿名化された情報を収集するもの、無記名調査を行うものその他の個人情報を取り扱わないものであること。

イ人体から採取された試料等を用いないものであること。

ウ観察研究であって、人体への負荷を伴わないものであること。

エ被験者の意思に回答が委ねられている調査であって、その質問内容により被験者の心理的苦痛をもたらすことが想定されないものであること。

あらかじめ指名する者が、研究者等が所属する医療機関内の患者の診療録等の診療情報を用いて、専ら集計、単純な統計処理等を行う研究であり、倫理審査委員会への付議を必要としないと判断した場合

次に掲げる事項についての規定を含む契約に基づき、データの集積又は統計処理のみを受託する場合

アデータの安全管理

イ守秘義務

第0相試験として計画した内容が人体への負荷を伴わない観察研究であって、被験者に心理的負

荷を与えない場合は上の条件を満足する可能性があるが、その場合にも「あらかじめ指名する者」による判断が必要であって、研究者による恣意的な判断によるべきではない。

B) 開発者による試験的使用の場合

開発者自身による試験的使用については異なった取り扱いを要する。研究者は被験者になることはできないことから、開発者による試験的使用の結果を機器に関するエビデンスとして外部に公表するなどは許されない。開発に当たっての参考資料として役立つにとどめるべきである。このことは、第0相試験に相当する場合に限ることはなく、あらゆる臨床試験について言えることである。

この場合、侵襲のない、体外使用の支援機器の場合には倫理審査は不要であると考える。リスクは大きくないし、探索的な試行錯誤の段階であり、結果を対外的なエビデンスとして用いるわけでもないためである。もちろん、安全性の問題は残るが、それは職場安全に関する労働安全の問題として扱うことができる。

もし、侵襲性のある機器であるならば倫理審査を必要とするのが原則である。このことは、開発者自身の保護のためでもある。しかし、試行錯誤による探索的開発の場合、設計の一部を変更するたびに倫理審査を申請するのは実質的にかなりの困難を伴う。この場合、機器の材料、動作原理と具体的な機構などリスクに関わる因子について具体的な範囲を規定し、その範囲でフレキシブルな運用が可能な倫理審査の方式を採用する必要がある。

当然であるが、開発の進展につれてリスクの再評価が必要になったときには変更申請を必要とする。この場合、リスクを再評価するので迅速審査によることはできない。

C) 第I相試験

生物医学領域における第I相試験は、初めて人体への作用を行う試験である。医薬品の第I相試験の場合は、健康なボランティアを被験者とした小規模の試験である。少量の医薬品の投与によって副作用の発生、代謝や排泄による「薬物動態」を観察し、さらに研究を進めることが適切かどうかを結論づける。この段階では、効能、効用をエンドポイントとすることはない。デザインとしてはオープンラベル（研究者も被験者も検査薬について知っている）あるいは、単盲検（被験者のみが検査薬の知識を持たない）である。

支援機器におけるアナロジーとしては、開発・

試作したプロトタイプ機について、健康なボランティアを被験者として安全性の確認を主要エンドポイントとする臨床試験である。主として少数の被験者に対して評価する。有効性の主観的評価を求めることもあるが、主要エンドポイントとするのは適当ではない。支援機器の場合、マスクングすることは不可能なので、常にオープンラベルである。

通常は健常成人を被験者とするが、場合によっては少数の障害者を含むこともある。第 I 相試験について倫理審査が不要との論もあるが、先に述べた理由によってこれは正しくない。それどころか、健康なボランティアをリスクにさらすことの倫理上の問題が重視され、研究倫理の教科書^{6,7)}においても特別の章を割いて詳細な議論がなされている。

D) 第 II 相試験

生物医学領域における第 II 相試験は、対象とする患者、障害者に初めて投与する段階である。第 II 相試験における最終目標は有効性の確認にあるが、安全性についても引き続き観察を続けてゆく段階である。医薬品の第 II 相試験においては、MTD(Maximum Tolerated Dose:最大耐量、許容できない副作用を引き起こすことなく投与できる薬物または治療の最大の用量)を求めることもある。また、2 つ以上の群に分割して比較を行う研究デザインとするのが通例である。

上記のアナロジーによって支援機器の場合を考えると、第 II 相試験において初めて対象とする高齢者・障害者による臨床試験に移行する。人数としては第 I 相よりは多数であるが、第 III 相に比べると少数の被験者を対象とする。

主要エンドポイントとしては、開発の段階に応じて、試作機の動作ならびに機能の有効性に係わる指標を設定する。第 II 相試験まではパイロットスタディと位置づけられ、数回に分けて行うこともある。それぞれの試験について、検証すべき目的を明確に定義する必要がある。検証すべき目的としては、機器としての有効性や、安全性に関わる問題を様々な側面から更に確認することにある。

E) 第 III 相試験

生物医学領域における第 III 相試験は“ 枢要な ” 相試験 (pivotal trial) とも呼ばれ、治験薬や新規治療法に関する最終的な試験である。新薬や医療機器においては治験の最終段階を構成する。第 III 相試験においては、RCT を典型とする群に分割した様々な研究デザインが採用される。

第 III 相試験によって有効性が確認されれば治療薬、治療機器として市場に出されることになる。

このアナロジーとして支援機器について考えてみる。市販前の有効性と効能にかんする最終試験の段階であることは変わらない。しかし、支援機器の場合には RCT をはじめとする生物医学領域で発展してきた強力な研究デザインの採用が不可能に近い。このための研究デザイン上の工夫が必要となる。

支援機器の場合はまた、医薬品の場合ほどの多人数の被験者による試験は不可能であって、統計的に意味のある被験者数をそろえるだけでも相当な困難があるのが通例である。また、医薬品の場合は市販時における同等の品質の製品の使用が求められる。支援機器の場合は量産体制で製造した機器を用いることは必ずしも可能ではない。このような場合、「生物学的妥当性」の判断によって検証結果を補うことになるが、そのためにはその分野を専門とする医療関連職との共同作業による判断が必要となる。

第 III 相試験は最も重要な試験であり、客観的に機器の性能を確認することが中心になる。安全性の問題については第 II 相で確認済みのはずであるが、さらに製品としての確認を行う。PIADS 等の心理スケールによる評価も併せて行われることもある。

F) 第 IV 相試験

生物医学領域における第 IV 相試験は一般的には診療環境における治療法のモニタリングと考えられており、副作用観察と効果的な治療法の開発を目指すものとされてきたが、最近注目されているのは承認された使用法以外の適応拡大のための役割である。

一方、製薬会社の利益追求を目的としていて、科学的意義を認めがたい第 IV 相試験も見受けられ、重要な結果をもたらすものとの区別が必要とされる。この区別は倫理審査によると考えられるが、診療と一体的に行われ、医師の処方権の範囲内で遂行される第 IV 相試験においては、個別的な倫理審査は必ずしも必要とは考えられていない。ただし、適用範囲の拡大について組織的な研究として第 IV 相試験を行う場合は当然倫理審査が必要である。

支援機器の第 IV 相試験においては生物科学領域における場合に比べて異なった要素があることに注意が必要である。それは、一般に第 III 相試験において十分なサンプル数をそろえることが困難であり、十分な統計的有意性を持って支援機器の有用性が科学的に確立できているとは言

い難い点にある。

このため、市販後も観察を継続し、有用性や問題点を抽出する第 IV 相試験を継続することが必要と考えられる。我が国の場合はほとんどの場合流通からのクレーム等によるフィードバックに頼っていると思われる。

これまで第 IV 相試験を積極的に推進してこなかったのは、我が国の支援機器関連企業の規模が小さく、そのための資源が乏しかった点もあろう。しかし、第 III 相試験の不十分な点を第 IV 相試験として組織化し、機器の機能、性能、適応と適合などに関してより一般化した知見として蓄積することが支援機器分野における技術力の蓄積には必要であろう。

D. 臨床研究における研究デザイン

工学系研究者が主導となって支援機器を開発したとき、困惑するのは使用者による評価法である。開発担当者への調査においては必ずと言っていいくらい「評価法が存在しないので困る」という返事が返ってくる。

評価のための方法は存在し、生物医学領域で広く使われているのであるが、それを知らないだけのことである。

一般に、工学系の研究においては生物医学系におけるようなプロトコルを作成することに慣れていないためにプロトコル作成には苦労することが多い。倫理審査のためにプロトコルを作成するに当たって、様々な苦情が寄せられるのはそのためでもあるが、以下のような要因を挙げることができる。

プロトコル（研究計画書）の作成に慣れていない。倫理審査の申請書では専門家ではない被験者のための説明文書、専門家ではない審査委員のための文書を作成するが、専門家ではない委員のための文書の作成に慣れていない。プロトコルと倫理審査に特有な用語（リスクと便益、連結不能匿名化、仮説とエンドポイント、選択/除外基準など）を理解するだけでくたびれてしまう。

倫理上の判断においては状況に依存することが大きく、場合によっては逆の判断をすることがある。

工学系の分野には被験者の尊厳、人権の尊重などの視点が工学の伝統としては存在しない。

しかし、現在のプロトコルは生物医学領域における長年の経験の中で発展してきたものであり、被験者保護に関する配慮も含め、人に関わる研究を行うに当たっては欠かすことのできない方法論である。

本章では、研究デザインに関する教科書⁸⁾に沿って生物医学領域における研究デザインの概要をとりまとめるとともに、支援機器の臨床試験に適用する場合の問題点について検討する。

A) 研究デザインを記述するパラメータ

仮説とエンドポイント

臨床試験にはそれぞれの目的に即した研究デザインが求められるが、問題点を仮説の形で表現することでポイントを明確にすることができる。すなわち、仮説は当該臨床試験によってその試験期間内に証明すべきものである。また、一般化できる知見とするためには、試験の手続きに従って抽出されたエビデンスによって証明可能な仮説であることを要する。

エンドポイントとは、仮説が成立するためのエビデンスとなる指標のことであって、「試験の目的に関連する仮説を検証するうえで臨床的に意味があり、客観的に評価できる観察・検査項目またはそれらの合成指標である。」と定義される。

エンドポイントは、対象者ごとに評価されるものであり、発生割合や有効割合など、集団について定義される指標ではない。エンドポイントのうち、試験の目的に最も合致したものを主要エンドポイントとし、それ以外を副次エンドポイントとする。

仮説とエンドポイントの設定は臨床試験における基本的要素である。臨床試験によって検証するのは、推測統計学における仮説 - 検定論理を利用するので、仮説は推測統計学で用いる仮説と同じと考えて良い。同様に、エンドポイントは仮説が成立することを検定するためのパラメータであると考えて良い。

支援機器の臨床試験の場合、第 0 相から第 II 相試験までは機器の性格によってその段階で確認することが必要な事項を仮説として選び、第 III 相試験においてはその機器が有効であることを表すために最も重要なポイントを具体的に設定し、主要な仮説として設定すればよい。

このとき、主要エンドポイントは主要な仮説が成立するためのエビデンスとして計量できるパラメータとして設定する。

しばしば見受けられることであるが、仮説の設定が適当ではないために、エンドポイントの設定が適切ではない場合がある。

車いすなどの移動支援機器の場合を例として考える。「移動の支援を可能とする。」といった抽象度の高い仮説を設定すると、せっかく開発した車いすの特長をアピールできなくなる。たとえば、片麻痺者用の足こぎ車いすの場合に、「直進走行

性が良く、小回りがきく」といった仮説を立てたとしよう。この場合エンドポイントとしては、直線走路の直進性のパラメータと回転半径の操作性にかんするパラメータをエンドポイントとして設定することになる。(どちらが重要視する方が主要エンドポイントとなる。これ以外の副次エンドポイントを設定しても良い。複数のアウトカムについて離散的スコアを定義し、それらをあわせた合成の評価スコアとすることもできる。)

仮説の成立を検証するためのエンドポイントの使い方は研究デザインによって異なってくる。たとえば上記の車いすの例で、直進性能のパラメータとして、目標とする直線と実際の走行曲線との囲む面積によってエンドポイントとしたとする。研究デザインが従来車いすとの比較デザインであれば、面積の計測値が有意に小さければ開発した車いすの直進性が比較対象のものよりも有意に良好であると結論することになる。

主として身体機能を補助する支援機器の臨床試験におけるエンドポイントは客観指標を使うべきである。これは、すでに述べたように主観評価においてはバイアスが入りやすいため、偏った評価となる可能性が高く、そのような結果を排除するためである。

予測因子

予測因子とは、時間的あるいは生物学的に先行している変数として定義される。一般には対象者の属性のほか、人為的に制御できる環境条件など、独立変数と考えられるパラメータを設定する。臨床試験では一般に何らかの介入に対する被験者の反応を計測するが、介入は特殊なタイプの予測因子と位置づけられる。

支援機器の臨床試験の場合、支援機器自体が介入であり、予測因子である。従って、そのセッティング、実験環境も予測因子として位置付ける。

実験条件を整えた実験室における実験の場合には精度と再現性の高いデータを記録することができるが、実環境でどこまでの性能が可能であるかについては結論できない。このために、実環境になるべく近い実験環境を準備することも一つの方策である。

実際には実環境における長期実験が望ましい。在宅または施設における長期実験においては、予測因子としての環境条件を統制することはできない。このような統制のない条件下での実験に際してはアウトカムの面において十分な配慮が必要である。

アウトカム

予測因子を独立変数と見るなら、アウトカムは

従属変数に対応しており、予測因子によって定められた実験条件によって与えられた帰結としてのパラメータであって、実験の結果として観測するパラメータである。

アウトカムのうち主要なアウトカム変数が主エンドポイントである。エンドポイントはアウトカム変数の組み合わせによって構成することもある。様々な課題に対する得点を組み合わせて合成指標としてアウトカム変数とすることもある。QOL 指標など心理指標や身体機能の評価指数などである。これらは既成のものをそのまま、あるいは改変して使うこともあるが、新たに作成することもある。客観的に他の研究との比較を可能とするためには何らかの方法で標準化された指標を使うことが勧められる。

支援機器の臨床試験においては主要アウトカムは客観指標を用いるべきであることはすでに述べたが、ユーザーによる受容性、ユーザビリティ等を評価するためには主観的指標を用いることになる。

新たな指標の作成、あるいは既成の指標の改変に当たって必要な要件は、

(a) 妥当性：validity

作成した指標が真の指標として妥当であることを確認する。通常はすでに確立していて、似ている指標による結果と比べて整合性によって確認する。

(b) 信頼性：reliability

異なった観察者による測定結果が整合している、ある期間を置いて計測した結果が整合している、似た項目同士で反応が一致している等の方法によって信頼性を確認する。

(c) 感受性：sensitiveness

被験者の状態変化に伴ってスコアに変化が観察されることを確認する。

主観指標を用いる際には十分な時間をとって準備する必要がある。

選択 / 除外基準

予測因子とは通常考えられてはいないが、試験条件を意図的に規制する要因であり、科学的に重要な要因であるので重要なパラメータと位置づけてとして検討しておく。

被験者の選択 / 除外基準は Inclusion/Exclusion と呼ばれ、研究計画の科学性を保証する上で重要な要素である。支援機器においては機器のターゲットとする機能障害は定義されているが、実験手順に即して具体的な被験者の選択 / 除外基準を定義する必要がある。

選択基準は研究テーマにふさわしく、研究効率の高い集団を定義することであり、「卒中による片麻痺の維持期でブロンストロ - ムステージ4以上の患者」などと定義する。一方除外基準は選択基準の中でデータの質を悪くしたり、フォローアップが難しかったり、実験が禁忌を含むなどの理由で被験者から除外する基準である。これらは研究計画書の作成時に決定していることが必要である。理工系の研究者からの申請には、「現場の理学療法士に選定してもらおう」といった記載を見受けことがある。そうすると、被験者の選択はその理学療法士の主観によって定まることになる。被験者を意図的に選択したことによる選択バイアスを疑われても反論できない。選定を依頼した理学療法士がどのような基準で判断するかを開示するのが選択/除外基準である。専門の医師に相談してあらかじめ決めておく必要がある。

選択/除外基準のうち、特に重篤な有害事象が発生する可能性のある場合、これを禁忌としての除外基準を定義する。これは科学性の担保のためであるよりは有害事象の発生予防による被験者保護を目的とするものであるが、禁忌の可能性を見落としたとすれば、実験条件の科学的検討が不十分であったとのそしりを免れないであろう。除外、禁忌などを慎重に定義する必要があるのはこのためである。支援機器の場合、これらの検討のためには医療職によるアドバイスが欠かせない。研究チームに医療職が加わっていない場合は医療職に耳を傾けることが求められる。

B) パラメータの妥当性

前節で定義したパラメータについて、それが目的とする内容を正しく表現していること、すなわち、妥当性が求められる。妥当性の評価には5つの側面がある。

内容妥当性：測定が研究目的にとって必要な全体を適切にカバーしていること。たとえば、QOLに関する質問紙調査の場合、社会的、身体的、情緒的あるいは知的機能などの質問項目がすべてカバーされていることを確認する。
構成概念妥当性：研究しようとする理論的概念を表現し得ていること。生物学的妥当性に同じ。

基準関連妥当性：新しい測定が確実性の高い既存の測定法による結果との整合性のあること。

交絡因子の排除：交絡因子が排除されていること。交絡因子とは、予測因子と関連を持つが、結果に誤りをもたらす恐れのある因子のこと。たとえば、運動の頻度の高い人が風邪

を引く度合いが小さかったとする。見かけ上は運動によって風邪の予防になるように見えるが、実際には、意識的に運動の多い人は社会経済的に生活環境、栄養状態がよいし、健康意識も高いと考えられ、その結果、運動頻度も高いし風邪の頻度も低いと考えられる。

このように、隠れた因子が真の要因となっている場合を交絡因子と呼ぶ。これを排除するためには、対象者を限定する（選択/除外要件）、対照群とのマッチングなどの手法がある。

バイアスの排除：バイアスとは、系統誤差のことで、何らかの理由で偏った結果を記録することである。バイアスを含む記録は誤った結論を導くので、科学性を担保するためにはその排除に努める必要がある。代表的なバイアスには以下のようなものがある。

(a) 測定者バイアス：測定者が故意あるいは無意識に偏ったデータを報告すること

(b) 測定手段バイアス：計測機器や質問票などによって引き起こされる系統誤差

(c) 被験者バイアス：被験者が原因となるバイアスで、被験者と測定者の関係が非対称であって、遠慮、追従、権威への追従などが想定されるとき起こりやすい。

主観評価の場合特に配慮する必要があるが、客観評価の場合にもその可能性を否定することができない。たとえば、同じ研究室の学生を被験者としたとき、研究者の期待する結果を無意識のうちに出してしまう可能性がある。このように特別な関心のある場合、期待される行動をとりやすい場合をホーンソン効果と呼ぶ。

また、主治医が勧める薬を服薬すると、偽薬であっても有効であることがある。このような現象はプラセボ効果と呼ばれる。

C) 研究デザイン

プロトコルの骨格が研究デザインである。生物医学領域の研究においては、対象が人間であるため物理化学系の研究とは異なり、確実な再現性は期しがたい。個人差に基づくバラツキを伴ったデータしか得られないのが通例である。

このため、かつての生物医学領域の臨床研究においては「比較対照のない単なる一連の観察や治療にかんする意見陳述」、「ライバルたちによる診断または治療を参考にしながらない権威者や教授の意見によって決定されていた」²⁾状態であった。

この事態を一変させたのは1940年代に確立し

た推測統計学であって、1950年代に入ってから推測統計学に基づいた「仮説 - 検定」による検証法が一般化した。

生物医学領域における臨床研究の研究デザインは、現象を記述するデザインも存在するが、それも最終的には「仮説の検証によって一般化できる知見」を得るためのデザインとなっている。

代表的な研究デザインのタイプを表2に示す。これらの概略を示すとともに、支援機器の臨床試験に対する適用に関する検討を行うものとする。

観察研究

すべての研究がそうであるように、データを収集して観察することから支援機器の開発も始まる。「介入」を伴わないことが観察研究の特長であるが、生物医学領域では「臨床現場通りに治療を行いつつ仮説を証明しようとする」のが観察研究である。

支援機器に翻訳すると、日常的に使っている機器を普段通りに使用している状況を観察、測定し、使用中の機器に関して仮説を立証しようとするのが支援機器に関する観察研究となる。

支援機器に関する観察研究はほとんどの場合、現在使用中の機器に関する問題点を把握することを目的としている場合が多いと思われる。ニーズ主導型の支援機器開発はこのような観察研究から始まる。

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>
--

一方、シーズ主導型の開発においては、開発の後でニーズを探索しがちである。シーズ主導型の

開発においても、可能な限り早い時期に実際の生活実態の観察を取り入れるべきである。

(1) 記述的研究

文字通り、ありのままに状況を記述する研究で、手動車いすユーザがどの程度片流れを経験しているか、また、その制御にどの程度困難を抱えているか等である。

(2) 分析的研究

記述的研究に引き続くのが分析的研究で、記述的研究と並行して行われることもある。分析的研究では因果関係を推論するために、変数間の関連の有無に着目する。これは時間を伴うかどうか、対照群を設定するかどうかによって以下のようなタイプに分けることができる。

(a) コホート研究

コホート研究とは被験者の群 (コホート) を一定期間にわたって追跡観察する研究である。その基本形はアウトカムの発生率の変化を記録することにある。

研究の開始に当たってはベースラインとして、アウトカムの発生に関連すると思われる予測因子を記録しておく。引き続き、定期的に出カム測定を行う。

支援機器の臨床試験においてコホート研究が有効なのは、症状が変化する場合 (認知症や ALS)、訓練効果が関わる場合、全く新規の支援機器に関して、長期使用によって引き起こされる効果を調べる場合などのために長期臨床試験を行う場合等である。

上に述べた前向きコホートの他に、後ろ向きコホートのモデルもある。これは主としてカルテの記録をさかのぼって調査する。このような記録の存在が前提となるため、支援機器の開発に関する後ろ向きコホート研究はあまり行われない。

(b) 横断研究

すべての測定をある一時点で行うのが横断研究である。対照となる因子の分布や相関を調べるには便利である。

支援機器の臨床試験としては第0相試験として現在使用中の機器の問題点についての主観評価などに用いられる。

(c) ケースコントロール研究

ケースコントロール研究のデザインは主として疾患のリスクファクターを研究しようとするときに用いられる研究デザインで、着目

する疾患の患者よりなる群をケース群と呼ぶ。

これに対し、基本属性をはじめ、その疾患に関する既知のリスクファクターの似ている条件（マッチング条件）はよく似ているが、罹患していない人よりなる群（コントロール群）を選び、可能性のある予測因子の割合を両群で比較する。有意の差があればリスクファクターと考える。

このデザインは支援機器に適用するには適していないと思われる。特別の場合、特定の機器の不適合の理由を解明するために使えるかもしれない。しかし、その場合も十分なコントロール群の選択が容易であるとは思えない。

(d) 複合的デザイン

観察研究としては、コホート内ケースコントロール研究、コホート内ケースコホート研究、ケースクロスオーバー研究など様々な複合的デザインがあるが、支援機器開発において特に有用であるとも思えないので省略する。

実験的（介入）研究

観察研究は人為的な（実験的な）操作のない状態での観察によって研究を行うが、標準的な処置を超える操作を実験的に加えることによってひきおこされるアウトカムを観測し、それによって実験の仮説を検証しようとするものである。

開発途中、あるいは開発した支援機器の試用実験は、使用者が健康なボランティアであろうと高齢者・障害者であろうと、上記意味においては「実験的」であって、「介入研究」としての取り扱いが必要である。

本報告書3.(A)に引用した「臨床研究に関する倫理指針」による「倫理審査不要の要件」からすれば、すべての介入研究は倫理審査を要する訳であり、すべての支援機器の臨床試験は倫理審査が必要である。

本筋から外れるが、迅速審査の適用条件について考えてみる。迅速審査の要件は、同指針第3(9)〈細則〉によって、次のように定められている。

研究計画の軽微な変更

共同研究であって、既に主たる研究機関において倫理審査委員会の承認を受けた臨床研究計画を他の共同臨床研究機関が実施しようとする場合の臨床研究計画の審査

被験者に対して最小限の危険（日常生活や日常的な医学検査で被る身体的、心理的、社会的危害の可能性の限度を超えない危険であって、社会的に許容される種類のも

のをいう。）を超える危険を含まない臨床研究計画の審査

この要件に照らしてみると、「最小限の危険」に相当するかどうか問題となる。試作した機器の試用は「日常的に経験するもの」ということはできず、リスクが非常に小さいと判断されるものであっても、迅速審査に相当するかどうかの判断はフル審査によらなければならない。結局日本の指針の下では支援機器に関するあらゆる臨床試験はフル審査を必要とするということになる。

以下、生物医学領域における介入研究の研究デザインについて述べ、支援機器の臨床試験への適用可能性について検討する。

(1) 比較対照試験（Controlled Trial）

介入による効果をアウトカム変数のみによって結論するのは困難であるので、異なる介入を受けた群の間の比較によって、アウトカムが有意に異なっていることによって新規の方法が優れていることを検証する研究デザインである。

比較対照試験においては、ランダム化と盲検化とが重要である。生物医学領域では盲検のランダム化試験が最も強力な研究デザインと考えられている。

盲検化というのは、処方された医薬が試験薬であるのか偽薬であるのかを知らされないことを指し、被験者、測定担当者、データ解析担当者のどこまでに知らせないかによって、単盲検、二重盲検、三重盲検と呼ばれている。

支援機器の臨床試験においては、残念ながら盲検化の手法はほとんど無力である。ほとんどの支援機器は見ただけでそれが既存のものであるか、新規に開発されたものであるかが判断できるためである。

(a) パラレル

生物医学領域において異なった患者群の間での比較を行うのがパラレル比較(parallel comparison)デザインである。介入研究のデザインとしては、最も広く用いられている。

i. ランダム化（RCT: Randomized Controlled Trial）

生物医学領域において患者を2つの群にランダムに分けて、処置（介入）に対するアウトカムを観測し、その有意の差を検定する方法であり、盲検化と組み合わせたものは最も科学的に妥当な研究デザインとして位置づけられている。

支援機器の場合への適用可能性について検討しよう。電動車いすを必要としている人

を被験者として2群に分けたとする。この2群に既存の電動車いすと新規に開発した制御装置を備えた電動車いすを割り付け、モデルコースを走行するに要する時間によって性能を評価したとしよう。

この場合の難点は、2つの群において被験者のスキルが同じであるとの保証がないことである。生物医学領域においては、生物学的な特性を同等にするためにランダムに2つの群に分割する。また、統計学的に有意であることを保証するために被験者数を十分に大きくする。

しかし、電動車いすのユーザはそう多くないし、高齢者で、介護保険で初めて簡易型電動車いすを使い始めた人と、脊損のために長年電動車いすを使ってきた人とのスキルの差は大きい。

経験年数によってスキルの程度をある程度そろえることは可能であるが、それでも、統計学的に有意性を主張できるだけの被験者数をそろえるのは容易ではない。

このようなわけで、支援機器の臨床評価においても（盲検のない）ランダム化比較デザインを採用することは可能ではあるがあまり用いられていない。

しかし、これまで全く使われたことのないタイプの支援機器の場合はアウトカム変数を適切に選択すれば有効なデザインとなる可能性もある。

ii. 非ランダム化

ランダム化のない群間比較のデザインもありうるが、一般には交絡因子を除くには困難があり、望ましくないデザインであるとされている。

(b) 逐次

同じ集団に対して、介入期とコントロール期をもうけ、アウトカムの比較を行うデザインである。自分自身をコントロールとするため、スキルに相違のある集団に対しても適用可能であり、支援機器の臨床試験においては最も適したデザインと言える。

i. 自己対照（前後比較）

生物医学領域においては、ベースライン因子とアウトカムの測定、介入とアウトカムの測定を繰り返すことによって介入による効果を比較するデザインである。

先に述べた電動車いすの例でいえば、現在使用中の車いすによる走行時間の

測定

開発した車いすによる走行時間の測定を逐次行うデザインである。仮説検証に際しては、による方が走行時間が短いことを仮説とする。検証にあたっては、両者の比、あるいは大小による2値尺度を設定し、統計的検定にかける。

支援機器の臨床試験においては、比較対象の機器が存在するときにはこの研究デザインが最も一般的に適用可能である。

注意すべきは、たとえば新規の機器に関しては一定の練習が必要であろうし、機器によっては1-2週間程度の連続使用も必要かもしれない。これらについては、必要に応じて検討する。

ii. クロスオーバー

生物医学領域においては、対照群を2群に分ける。AとBの2つの処置の比較を行うに際して、Aを先に行う群とBを先に行う群とに分けてアウトカムを測定し、一定の回復期間をおいた後で逆の介入を行う。

このデザインでは、群内比較も群間比較もできるため交絡の影響を避けることができる利点がある。

支援機器の臨床評価においては、これまで使ってきた機器との比較を行う場合にはクロスオーバーのデザインを採用する利点は小さいと思われる。これに対し、全く新規の支援機器を開発したとき、複数の設計があり得た場合の比較の場合などには有効である可能性がある。

(2) 対照なしの研究

生物医学領域では対照のない介入研究はないと考えて良い。科学的説得力があまりに弱く、医薬品の場合は偽薬を対照として選ぶことができるためである。

似た機能の存在しなかった、全く新規の支援機器の臨床評価の場合、対照とすべき機器が存在しないため、対照なしのデザインとならざるを得ない。

支援機器のない状態でアウトカムを測定し、ベースラインとする。次いで、機器を用いてアウトカムを測定し、ベースラインと比較する。この場合、プラセボ効果を伴いがちであるので、十分に注意する必要がある。

(3) シングルケースの研究（N of 1 design）

生物医学領域、心理学領域において1症例の研

究とも言われる研究デザインである。薬とプラセボの組を何組か作り、各組の中で順序をランダムに投与する。これらに対するアウトカムのスコアから統計処理によって有効性を検定する。

支援機器の臨床試験の場合、プラセボの設定ができないので、開発した機器の有効性の検証に使うのは困難がある。

特定の個人に即した支援機器の開発は伝統的に行われてきたものであり、試行錯誤で開発してきたのが通例である。

この種の開発の過程においては N of 1 の研究デザインは有効であると考えられる。しかし、そのような例は見たことがない。今後、試行錯誤による特定の個人に即した開発において、このモデルを研究デザインとして採用することも検討すべきであろう。

E . 用語について

本分担研究報告書において用いた用語のうち、留意すべきものを以下に示す。

臨床研究：主として生物医学系領域の研究であって、人を対象とするもの一般を表す。

臨床評価：支援機器について、その有効性の確認、検証を主たる目的とする臨床研究。特定の支援機器については、第 0 相試験から第 IV 相試験までを含む。

臨床試験：臨床評価において、健常ボランティアあるいは患者、高齢者、障害者等が支援機器を試験的に使用し、最終的には有効性の検証・確認を行う試験。機器の開発段階に応じて第 0 相試験から第 IV 相試験まで位置づける。第 III 相試験において市販の可否を判断するが、各相に応じて目的、ゴールは異なってくる。

【文献】

1. The Belmont Report:
<http://ohsr.od.nih.gov/guidelines/belmont.html>.
[津谷喜一郎、光石忠敬、栗原千絵子による日本語訳は
<http://homepage3.nifty.com/cont/28-3/p559-68.html> (accessed Feb 20, 2012).
2. 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 .
3. FDA:”Guidance for Industry, Investigators, and Reviewers, Exploratory IND Studies”,
<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm0>

- 78933.pdf. (accessed Feb 20, 2012)
4. “Phase 0 trials: a platform for drug development”, The Lancet, 374, p.176, 18 July 2009,<http://www.thelancet.com/journals/lancet/article/PIIS0140-6736%2809%2961309-X/fulltext>. (accessed Feb 20, 2012)
5. E. G. DeRenzo and J. M. Moss: “Writing Clinical Research Protocols: Ethical Considerations”, p.57-8, Elsevier, 2006.
6. C. S. Way: “Phase I Clinical Trials in Healthy Adults”, Institutional Review Board, Management and Function”, 2nd ed., E. A. Bankert, and R. J. Amdur Eds. Jones and Bartlett Pub., p.346-350, 2006.
7. A. R. Jonsen and F. G. Miller: “Research with Healthy Volunteer”, The Oxford Textbook of Clinical Research Ethics, E. J. Emanuel et al Eds., Oxford University Press, page 487-487, 2008.
8. 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 .
9. D. H. Barlow, M. Hersen: 一事例の実験デザイン、高木俊一郎、佐久間徹訳、二瓶社、1988
10. 岩本隆茂、川俣甲子夫: シングル・ケース研究法、勁草書房、1990