

## 様式1 倫理審査申請書

申請書のカバーシートで、申請内容の主要な内容を記載する様式です。審査に当たっては、その申請の主要なポイントを審査委員が読み取ることができるように作られています。

## 様式2 支援機器の実証試験計画書

審査の対象となる実証試験の研究計画の詳細を記載する様式です。科学面と倫理面の両方について審査のポイントにあわせて記入するようになっています。審査では科学面が特に問題になります。「被験者実験の科学的な研究計画」というのはエンジニアにとっては理解しやすいわけではありませんが、被験者実験の基礎ですから、「研究倫理入門」および参考資料を熟読の上で計画書を作成する必要があります。

## 様式3 支援機器の実証試験にご参加いただくための説明文書

インフォームド・コンセントに際して、被験者候補が実験参加の意義を理解し、自律的な同意のもとに参加するための説明文書です。技術内容の予備知識のない一般の方に読んでいただくのですから、そのような読者を想定して記載する必要があります。一度書き上げたら、予備知識のない家族の方や事務職の方に読んでもらって理解できることを確認されることをお勧めします。

## 様式4 同意書

書面による同意を必要とする手続のための同意書の様式です。説明文書と同じ項目番号を参照しています。説明文書の項目番号を変更したときは項目番号が整合するように調整が必要です。

## 様式5 倫理審査申請書（変更申請）

承認済みの研究計画を変更するときの申請書です。

## 2. 実証試験のための研究倫理入門

### 2.1 はじめに

工学系のカリキュラムにおいては、研究倫理はほとんど扱われていない。そのため、開発を担当するエンジニアにとって研究倫理には戸惑うことが多い。ここでは、エンジニアの読者を想定して研究倫理の基礎について解説する。なお、最近では研究倫理や倫理審査関連の参考書 [1; 2; 3; 4]、臨床研究の教科書 [5; 6; 7; 8; 9] など入手が楽になったので、参考にさせていただきたい。

#### 2.1.1 支援機器と実証試験

支援機器の実証試験に関わる研究倫理を検討するに先立って、支援機器および実証試験の定義について考えてみたい。

支援機器の定義であるが、支援機器を表すためには様々な用語、定義がある。ここでは最新の定義として ISO 9999:2011 [10] (支援機器分類) によるものを引用しておく。

支援機器：障害者によって使用される機器（用具、器具、機具、機器、ソフトウェア）であって、特別に製造されたものであると、汎用製品であると問わず、以下の目的で使用されるもの。

- － 参加の支援
- － 心身機能／構造及び活動に関する保護、支持、計測、代替
- － 機能障害、活動制限、参加制約の予防

この定義は ICF (International Classification of Functioning, Disability and Health; 国際生活機能分類) [11] との整合性を考慮して定義されたもので、障害者の心身機能、活動と参加すなわち生活機能を支援する役割を有しているものと理解できる。

支援機器の実証試験として広く受け入れられている定義はないが、ISO 14155:2011 [12] (人を被験者とした医療機器の臨床研究) を参考にすると、以下のように定義することができる。

実証試験：支援機器の性能の評価を目的とし、一人以上の被験者を用いた体系的に計画された研究

支援機器の実証試験においては、性能の評価と安全性の確認が目的であり、体系的な研究計画によって遂行するとともに、結果を性能と安全性に関するエビデンスとすることがポイントである。

もう一つ重要なポイントは、支援機器の実証試験が支援機器開発における不可欠の構成要素であるが、それ自身独立した一つの研究であるという点である。つまり、開発における他の活動とは別に、実証試験としての独立した研究計画を持ち、性能評価に関する独立した結論を導くことができるように組織されていることが必要であることである。

ここで、実証試験が「研究」であることに付言しておきたい。研究倫理においては研究の範囲が重要になる。特に、医学研究においては研究と診療の区別が重要である。診療名目で安全性や治療効果の確

立していない療法の研究を行うことは倫理的に許されない。この問題の解決のためにベルモントレポート [13]が明確にしたのが研究と診療の区別である。研究は「仮説の検証により一般化できる知見を導くこと」と定式化されている。ここで「一般化できる知見」がポイントである。実証試験においても機器の性能並びに支援機能について一般化できる結論（特定の個人への効能でなく、特定の障害者群一般を対象とした効能として）を導くことを目的としているからである。実証試験を「研究」として取り扱うのは「一般化できる知見」を追求する活動であることによる。

機器開発としては工学原理を駆使した機器開発が中心であるが、実証試験においては被験者の反応に関する実験から、障害者の支援に関する機器の性能を一般化できる知見として導くことになる。エンジニアにとっては日頃扱っている機器とは違って、人の反応の解析が中心となる。この点、慣れるまでは発想の違いに戸惑うかもしれない。

### 2.1.2 研究と倫理

研究の関わる倫理問題としては、研究倫理以外に生命倫理、研究不正と呼ばれる倫理問題がある。これらを混同しないことが肝要である。これら 3 つの倫理問題の相違点を表 1 に示す。

生命倫理はバイオテクノロジーの発展により提起されたもの [14]であるが、支援機器開発に即して考えると、生命倫理は開発中の支援機器を実際に使用した場合に発生する倫理問題である。典型的な事例に、機器の導入によって拘束が問題となるような場合がある。

支援機器と生命倫理に関してさらに付言すると、支援機器そのものが生命倫理上のジレンマを抱えている。車いすを例にとると、車いすを使い続けると廃用症候群によって歩行機能が減退する可能性がある。車いすを使わないと十分なモビリティを確保できない。このようなジレンマが移動支援機器には本質的に存在していることは意識しておく必要がある。

研究不正は有効性を示す実験データを捏造したり、改ざんしたり、盗用したりする場合 [15]で FFP とまとめて表現されることもある。主として学術論文として発表する場合に問題となる。機器の効用を謳うデータについても同じ問題がある。

研究倫理を一口に言うと、被験者実験としての妥当性と被験者保護の問題である [3; 16]。そのために必要な配慮が研究倫理であり、実証試験も直接の対象となっている。この手引きは実証試験のためにあらかじめ心得ておくべき研究倫理についてまとめたものである。

表 1 研究に関わる倫理問題

<p><b>生命倫理</b>：バイオテクノロジーの発展によってもたらされた先端医療技術の応用の倫理性に関する問題。</p>
<p><b>研究不正</b>：FFP すなわち捏造(Fabrication), 改ざん(Falsification), 盗用(Plagiarism)などの研究上の不正に関する問題。出版倫理を含むのが通例。</p>
<p><b>研究倫理</b>：「ヒトを対象とした研究」すなわち、被験者実験としての妥当性および被験者の人権や尊厳に関わる問題。</p>

### 2.1.3 実証試験と倫理指針

研究倫理とはいうものの、そもそも倫理とは何だろうか。岩波書店の哲学小辞典 [17]によれば、「倫理学」とは道徳の起源・発達・本質・規範についての学、となっており、「道徳」とは「その社会で一般に承認されている規範の総体」であるとされている。これらから、研究に関する倫理とは、「研究に関する社会規範」であると解釈することができる。すなわち、研究倫理は人を対象とする実験における被験者保護のための社会規範である。

ところで、社会規範には道徳と法律とがある。法律は明確な規定がある上に、遵守することを義務づけている。一方、道徳は必ずしも明文化されてはおらず、明確な罰則もない。倫理についても一般には規則として明文化されているわけではない。むしろ、倫理あるいは道徳律は状況依存性が大きく、確定した規則として記述することは必ずしも適切ではない。このために、具体的問題を道徳的に議論する際に道徳的ジレンマとして示す場合や決疑論と呼ばれる特別なアプローチによって示す場合がある。このため、研究に関わる倫理を体系的に述べるに当たっては、基本原則、規則、規範などとして述べられる。そして、研究者が順守すべき研究倫理は「倫理指針」として示されているのが通例である。倫理原則を「法律」とすることが必ずしも適切ではないためである。

我が国における研究倫理としての「倫理指針」はヒトゲノム・遺伝子解析研究に関する倫理指針(2001年) [18]、疫学研究に関する倫理指針(2002年) [19]、臨床研究に関する倫理指針(2003年) [20]の3つが示されている。これらは今世紀に入って相次いで制定されたものであり、直近の改訂はそれぞれ2013、2008、2008年である。

これらの指針のうちどの指針が支援機器の実証試験に適用すべきであるかを検討する。このうち、支援機器は通常ヒトゲノムや遺伝子解析には関係ないし、疫学研究は「明確に特定された人間集団の中で出現する様々な事象の頻度及び分布ならびにそれらに影響を与える要因を明らかにする研究」(臨・第1・3・(1)・③)であって、開発した機器の実証試験とは直接の関係はない。適用する可能性のあるのは「臨床研究に関する倫理指針」である。

支援機器の実証試験が「臨床研究に関する倫理指針」の対象であるかどうかは同指針の解釈にかかってくる。すなわち、実証試験がこの指針の「臨床研究」の定義である「疾病の予防方法、診断方法及び治療方法の改善、疾病原因及び病態の理解並びに患者の生活の質の向上を目的として実施される次に掲げる医学系研究」に含まれるかどうかの問題である。事情を複雑にしているのは、この項の細則に含まれる『「医学系研究」には、医学に関する研究とともに、歯学、薬学、看護学、リハビリテーション学、予防医学、健康科学に関する研究が含まれる。』との注記である。

この細則により、支援機器の実証試験が「リハビリテーション学」に含まれるなら実証試験はこの指針の適用範囲に含まれる。問題は指針に述べられた「リハビリテーション」の範囲にかかってくる。伝統的な「リハビリテーション」の概念は医学的、教育的、社会的、職業的の4つの主要分野よりなるとされてきた。これにリハビリテーション工学を含めるのが通例である。この観点からは支援機器の実証試験はリハビリテーション学の研究に含まれる。一方、一般には「リハビリテーション」を機能回復訓練のための医学的リハビリテーションに限定して解釈することも広く行われており、この立場からは支援機器はこの指針の対象範囲外であると解釈される。

指針の主要な対象は侵襲性のある介入研究であり、支援機器の実証試験が主要な対象として視野に入っていないことは明白である。しかし、直接に侵襲性のある臨床研究にかかわる項目を除外すれば、支

援機器の実証試験にも適用可能な内容となっている。

結局、現在の我が国において支援機器の実証試験に対してこの指針を適用すべきかどうかについては指針上だけでは確定することができない。当面の結論としては、臨床研究に関する倫理指針を「準用」することが最も穏当であろう。準用というのは単に文字上の規定を当てはめようとするのではなく、それらの規定の背景、根底にある原理に敬意を払い、現実の問題の解決に生かそうとすることである。

## 2.2 研究倫理の意義

### 2.2.1 研究倫理の沿革

被験者実験に関して倫理規範が必要であるとされるようになったのはさして昔のことではない。被験者実験の倫理問題が初めて問題となったのは第二次大戦後のニュルンベルク裁判におけるナチスの人体実験であった。このために制定されたのが研究倫理の最初の重要文献であるニュルンベルク綱領 [21] (1947 年) である。これに引き続いて世界医師会によるヘルシンキ宣言 [22] が 1964 年に制定され、1978 年にベルモントレポート [13] が公表された。この 3 つの文書は研究倫理が形成される里程碑となった。

#### (1) ニュルンベルク綱領

第二次大戦の終了後ナチスの戦争犯罪を裁くためにニュルンベルク国際軍事法廷が開設された。ナチスの犯罪を裁くためのゲーリングやヘスの裁判の終了後、12 のニュルンベルク継続裁判が開かれたが、その一つが「医師裁判」である。ナチスの医師によって医学実験の名のもとに行われた人体実験が裁かれた。この裁判の過程でまとめられたのが 10 項目にわたるニュルンベルク綱領 [21] である。

ここに盛り込まれた考え方は必ずしもニュルンベルク綱領で初めて現れたのではなく、類似の要件は既に当時散見されていたようであるが [23]、被験者実験の要件としてまとめられたのはこれが初めてである。研究倫理がニュルンベルク綱領に始まるといわれるのはこのためである。

ニュルンベルク綱領の意義は、医学研究に関する倫理原則として世界的に影響を持った点にある。医療における倫理原則としてはヒポクラテスにまでさかのぼることができるが、医学研究を対象とした倫理原則としての意義は大きい。

この 10 項目のうち一番重要な第 1 項は、被験者の自発的同意が「本質的に絶対に必要」とであると規定している。これを可能とするためには、被験者が「理解したうえで間違いのない決断を下せる」ことが必要である。換言すれば、同意能力が不足した被験者や昏睡状態にある患者の場合は自発的同意を得ることが不可能であり、代諾の手続きも準備されていない。

ニュルンベルク綱領における同意要件の重要性は、インフォームド・コンセントという言葉が出現したのが、裁判後 10 年経過してからであった [24] ことから見て取れる。

#### (2) ヘルシンキ宣言

世界医師会は 1964 年のヘルシンキにおける総会で「一 ヒトを対象とするバイオメディカルの研究に携わる医師のための勧告 一」を採択した。この決議がヘルシンキ宣言 [22] と呼ばれるようになった。ヘルシンキ宣言はその後改訂を重ねてきた。単にヘルシンキ宣言というときは最新のバージョンによるべきである。現在のバージョンは 2008 年ソウル大会で決議されたものである。

1964 年版は序文に続いて第 1 章基本原則 (5 項目)、第 2 章専門的医療行為を伴う臨床的研究 (2 項

目)、第三章治療目的でない臨床的研究(4項目)からなっている。序文において強調されているのは臨床研究の意義が「学術知識を深めることによって人を助けるためには、研究室での実験結果をヒトに応用することが必須である」ことを明確にした点にある。この点はその後の改訂においても継承され続けている。

この宣言ではさらに「被験者に対する十分な説明と書面による自主的な同意」という表現でニュルンベルク綱領の自発的同意を継承したほか、法的代理人による代諾を可能とすることによってニュルンベルク綱領における過剰な制限の緩和も図っている。

その後の改訂のうち重要なのは1975年の東京大会における改訂で、「インフォームド・コンセント」が正式に用語として取り入れられ、「特別に任命された独立した委員会」という表現によって倫理審査委員会を要件として明確にした。

ヘルシンキ宣言はニュルンベルク綱領や次項ベルモントレポートのような歴史的文書ではなく、常に改訂が重ねられている文書であり、引用する場合は最新版によることとされている。

### (3) ベルモントレポート

研究倫理における次のステップは米国の「国家研究法(1974年)」およびそれに基づいて策定された「ベルモントレポート(1978年公表)」[13]であった。国家研究法に先立って、1960年代のアメリカでは生物医学領域のみならず行動科学においても倫理的に問題とされた研究が相次いだ。その結果、医学研究のみならず人文科学、社会科学、行動科学を含むあらゆる科学分野における被験者実験を伴う研究の倫理性が問題とされた。このため、国家研究法によって「生物医学・行動研究における被験者保護のための国家委員会」が設立された。これによって倫理審査を義務づける体制が確立した。これに先立ち、NIHでは1966年からNIHの資金による被験者実験に対して倫理審査が制度化されていた[25]。

アメリカで倫理審査が義務づけられ、倫理審査委員会の活動が監査の対象となっているのは連邦

政府の資金供与になる研究活動に限られる。これとは別に医薬品および医療機器の治験に際しても倫理審査は義務づけられている。この点は日本、ヨーロッパとも同様である。

ベルモントレポートで確立したのが先に述べた「研究と診療」の区別である。診療は「特定の個人に対して診断を与え、予防的処置や治療を加えること」であり、研究は「仮説を検証し結論を導き出すようにし、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」である。この区別はアメリカ以外の諸国の指針でも採用されており、研究倫理を適用すべき研究の定義となっている。

#### 参考 (1970年代に問題となった主な研究スキャンダル) ウイローブルック肝炎研究

ウイローブルック州立学校はニューヨーク市のスタンテン島に作られた州立の知的障害児施設。定員は4000名であったが最盛時は6000名を収容。1950年代に肝炎ウイルスの伝搬の研究のため、健康な児童に肝炎の児童の糞便から作った溶液を飲ませた。親にはこの実験に同意しない限り入所できないと説明していた。エドワード・ケネディはこの施設をsnake pit(蛇穴)と呼んだ。70年代になってその他の虐待とともに報道され、全国的なスキャンダルとなった。

#### タスキギー梅毒研究事件

タスキギーはアラバマ州メーコン郡中心の町。人口1万人程度。アフリカ系アメリカ人の小作人300名の梅毒患者の自然史の研究が1930年から72年まで公衆衛生局の資金で継続された。脆弱な層を対象として、可能な治療をせずに経過観察のみを続けた。全国的なスキャンダルとなり、1972年に中止された。

#### T-room 交渉研究

社会学者ハンフリーズは公衆トイレでの男性同性愛をのぞきで観察、自動車ナンバー、住所、氏名を調査した。対象者は観察されていることを知らされず、論文には対象者を同定できる詳細データが含まれていた。

## 2.2.2 研究倫理の原則

前節で述べたアメリカの国家委員会の目的は研究倫理のための広範で一般性ある倫理原則を示すことを含んでいた。このためにまとめられたのがベルモントレポートの倫理原則である。この原則は現在でも研究倫理の基本原則として位置づけられている。

表 2 ベルモントレポートの倫理原則

原則	適用
人格の尊重	インフォームド・コンセント
善行	リスクとベネフィットの衡量
正義	被験者の選択

ベルモントレポートの倫理原則と呼ばれるのは人格の尊重、善行、正義の 3 つの原則である。(委員会のスタッフとしてベルモントレポートを執筆したトム・ビーチャム [26]はヒポクラテス以来の無危害の原則を加えて生命倫理の 4 原則としている。 [27])

この 3 つの原則とそれより導かれる適用を表 2 に示す。「人格の尊重」の原則は人間としての自律性の尊重を求めるものであり、主たる適用はインフォームド・コンセントのプロセスにおける自律性と人間の尊厳である。インフォームド・コンセントは必要な情報が被験者にとって理解しやすい形で呈示され、それに基づいた自発的な同意を意味する。これには、同意の撤回や同意能力が不足している場合の尊厳が含まれる。

「善行」の原則は責務としての善行として取り上げられ、被験者実験に伴うリスクと負担の最小化、無危害の原則を含むと理解される。さらに、リスクとベネフィットを系統的に評価するとともに、リスクに見合うベネフィットを要求する。

ベルモントレポートはさらに研究を正当化できる 5 つの要件を示している。

- ① 被験者に対する過酷な扱いや非人間的扱いは道徳的に正当化できない。
- ② リスクは目的の達成に必要な限度まで減少させるべきである。
- ③ 深刻な障害をもたらす可能性があるリスクについては、委員会は格別に慎重に検討すべきである。
- ④ 弱者を被験者にするとき、そのこと自体の適切性が検証されていなくてはならない。
- ⑤ 研究に伴うすべてのリスクとベネフィットはインフォームド・コンセントに際してすべて開示しなくてはならない。

「正義」の原則は被験者の選択が公正に行われることを要求している。これは、特定のグループ（人種、経済的要件、施設入居者等）のみが被験者としての負担を課されるならばそれは不正義として排除されることを意味する。研究の成果もすべての人々がひとしく享受できることを要求する。特に、弱者の優先的選択は避けるべきである。社会レベルでは、移民や少数民族などのマイノリティ、施設入所者などを挙げることができるし、個人レベルでは研究者が特定の人たちを恣意的に選択してリスクにさらすことになるなどである。

## 2.3 実証試験の計画

### 2.3.1 実証試験の目的

支援機器の開発は基礎研究に始まり、プロトタイプの開発から数回の試作を経るのが普通である。図 1 に示したように、開発一試作を繰り返して、試作した機器の有効性を確認する。この確認のためには、対象とする障害者による試用と評価が含まれる。評価の結果は支援機器の有効性のエビデンスとする。

実用化のため開発の継続、商用化の判断に際しても実証試験の結果を参考にする。

実証試験の目的は第一に目的とした機能が実現していることの確認である。支援機器は障害者の心身機能、活動と参加に関する支援を目的としている。従って、支援機器開発においては第一にそれらの支援機能が求められる。実証試験ではその機能を検証する。

機能の確認以外に実証試験の目的として重要な意義を持つのは、改良・改善を要する問題点の発見および解決策の把握である。適応・適合のために必要な問題点の把握や適応・適合要件の確認など実用の場で問題となる事項を把握することも重要である。

実証試験において、機器の機能のみならずユーザビリティや QOL 評価なども最近では行われるようになってきた。単にニーズに見合った機能だけではユーザーの満足が得られにくく、ユーザビリティや QOL への視点も求められるようになってきたからである。しかし、基本はあくまで所期の機能であり、ユーザーを支援する機能のない機器のユーザビリティ評価は意味がない。

実証試験のもう一つの目的は安全性の確認である。しかし、被験者を実際に危害にさらし、危害の発生を判断基準とする検証は倫理的であるとは考えがたく、安全性の主要な作業はリスクマネジメントによるべきで、実証試験における安全性の問題はその結果を確認することにあると位置づける。

### 2.3.2 被験者実験の特質

物理学や化学をバックグラウンドに持つエンジニアが初めて被験者実験を行うとき、別の世界に迷い込んだ感覚を持つかもしれない。エンジニアにとって通常の決定論的アプローチが支配する無機物の世界から、確実性の乏しい人間の挙動の観察と確率論の世界に入ることによる困惑である。

被験者実験においては、一人一人の反応は一般には同じではない。異なっているのが当たり前の世界である。物理的世界の言葉を使えば、系を支配するパラメータの数が多すぎるため、すべてを制御・観測することができず、決定論の論理がそのままでは使えない。

具体例として電動車いすの新規の制御装置を開発した場合を想定する。この装置を用いて被験者による走行テストを行うとする。簡単な場合として、直進走行とスラローム走行の走行時間を計測したとする。被験者が脊髄損傷であったとしよう。操縦する手の障害の程度、巧緻性にはバラツキが大きい。走行に要する時間は手の巧緻性のみならず姿勢の安定性、座の適合の度合いの等の影響を受ける。電動車いすの使用歴にも依存する。その結果、走行時間は機器の性能よりも被験者の身体条件に強く依存する。このようにばらついたデータから機器の性能を表現するデータをどうすれば抽出できるであろうか？

個人差に基づくバラツキを伴ったデータしか得られなかった伝統的な生物医学領域の臨床研究にお

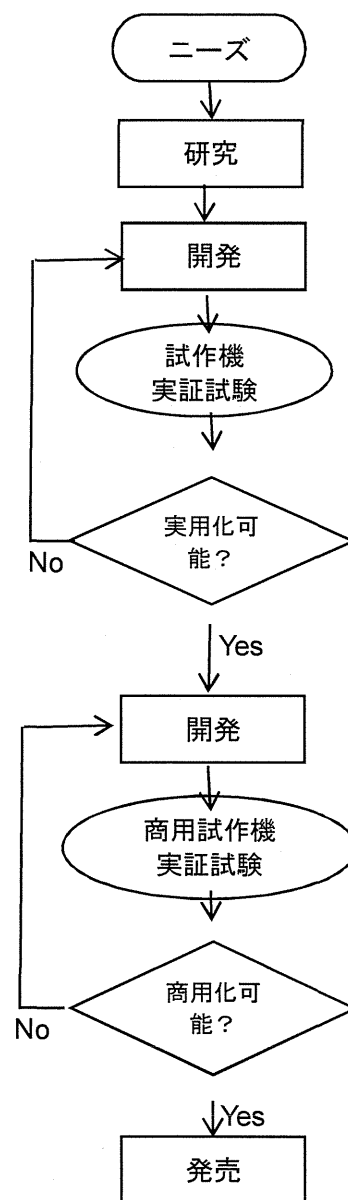


図1 ニーズの把握から研究・開発、商用化まで



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」 [28]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る。1950 年代に入ってから推測統計学に基づいた「仮説-検定」による検証法が普及した。

先の電動車いすの走行時間のデータに関しては、それによって検証すべき仮説を設定する。たとえば、普段使っている電動車いすによる走行時間に比べて、試験用の電動車いすによる走行時間が短い等である。測定データについて、これを仮説とした検定を行うことによって検証する。

このように、支援機器の実証試験においても生物医学系の方法論である推測統計学、特に仮説・検証論理による統計的検定がツールとして用いられる。ところが、支援機器の実証試験の場合は統計学が要求するだけのサンプル数をそろえることは必ずしも可能ではない。これには、特定の被験者を集めるのが容易ではないことの他に、コスト面からも多数の試作機による実証試験は困難であるし、少数の機器を使い回すにしても、一定期間で試験を終えるためには試験項目を減らすしかない。

いずれにせよ、質のそろった被験者を数十名集めたデータを得ることは容易ではない。通常は 20-40 名程度のサンプル数を得るのがやっとである。これではサンプル数が少なく統計処理のみでは説得力ある結論を得ることが困難となることもある。

生物医学的研究においては、統計学的に得られた結論は心身機能に関する生物学的・医学的知見と併せて考察する。一般に生物学的妥当性とも呼ばれるこの方法をとるためには、その分野の医学的知識を駆使する必要がある。これは、エンジニアだけでは一般には困難である。このような目的のためには専門医の協力とアドバイスが欠かせない。

### 2.3.3 優れた実証試験の条件

実証試験の実験を構想するにあたっては前節で述べた特質に留意しつつ、優れた研究計画を策定することが必要である。生物医学領域における優れた研究計画策定のためには、適切な問題提起が必要になる。この点で参考になるのが、EBM (Evidence Based Medicine) でよく使われる PICO による問題の整理である [29; 30]。PICO は P(patient or problem)、I(intervention)、C(control)、O(outcome)の頭文字を連ねたものである。これらの要因を明確に定義しておくことが、実証試験によって検証すべき問題点を明らかにするために有効である。特に、目的とゴールを明確にするための手がかりとなる。

#### (1) P: patient、対象とすべき障害の特性

Patient は機器の対象とする障害を明確にすることである。たとえば「下肢障害」等の抽象的な設定では明確にしたことにはならない。第一には、その機器の開発ターゲットとしての機能に関連した障害像が問題となる。同じ下肢障害でも様々な機能障害が含まれるし、同じ機能障害でも原因疾患には様々なものがあり、一般には原因疾患によって支援すべき機能障害のポイントが異なってくる。これらの要因を詳細に検討し、今回の実証試験においてフォーカスすべきポイントを絞っておく。原因疾患ごとに適応・適合の観点から検討すべき問題についても明らかにしておく。

これらは開発の企画から研究段階で十分な検討を済ませておくべき問題である。開発初期の検討が不足している場合には、実証試験の策定に当たって必ず検討を済ませておく。

これらの検討を経て、計画中的実証試験において検証すべき知見を具体的に絞り込んでおくことが優れた実証試験を策定する第一歩である。付録の iBOT の場合、選択基準、除外基準を詳細に記載してあ

るが、これは iBOT の多様な機能に対して被験者候補をバイアスなしに絞り込むために必要な条件として設定したものである。また、個別機能に関する除外基準も定めてある。対象とすべき障害像、想定すべき合併症などを検討し、部分的に適合しない要件を明確にすることによって実証試験において検証すべきポイントを絞り込んだものである。

#### (2) I: intervention、介入

支援機器の実証試験における介入とは、支援機器を実際に使用する実験そのものである。研究計画は当然支援の対象となる障害とそれに対する支援効果にフォーカスした介入として設計すべきである。開発した機器を漫然と「使ってみる」だけでは実証試験から得られる知見は貧しいものに留まる。対象とする障害特性に基づいて、検証すべき知見を絞り込んだ実験手続を立案すべきである。

具体的には、**patient** としての検討から絞り込んだ検証すべきポイントにそった実験項目の検討を行う。iBOT の場合、パイロット試験 1 においては在宅での 1 週間の試用、パイロット試験 2 においては、就労の場における 4 時間以上の試用、パイロット試験 3 においては 6 時間の試用、本格試験においては 2 週間の試用としている。それぞれ必要な試験・測定を行っている。

#### (3) C: control、比較対照

生物医学領域においては適切な比較対照とすべき介入を選定することが必須であると考えられている。場合によっては、介入のないことを比較対照とすることもあるが、最も多いのは現在の標準治療を比較対照とすることである。支援機器に置き換えると、現在の標準的な支援機器あるいは(人力による)支援手法を比較対照とすることになる。

比較対照が優れた実証試験の要件であるというのは決定論的世界からは奇異に見えるかもしれない。しかし、決定論的世界においても、温度計などの計測器の較正は、たとえば水の三重点やアルミニウムの凝固点などの物理量の定義に立ち戻って行うことは滅多に無く、より信頼性の高い測定器と比較するのが普通であり、絶対的な物差しの代わりに対照 (**control**) を用いると解釈することもできる。

iBOT の臨床試験においても、被験者が日常的に使用している電動車いすによる実験を比較対照としている。支援機器の実証試験における典型例である。

#### (4) O: outcome、アウトカム

アウトカムの一般的な定義は、介入の帰結を表現するパラメータのことである。アウトカム変数のうち、検証すべき知見を直接表現するものがあればそれをアウトカムとして選択することになる。直接表現するパラメータがないときは、最も適切なパラメータを「主要アウトカム」と定義する。たとえば、iBOT の本格試験においては、主要アウトカムとして **Community Driving Test** スコア、副次アウトカムとして **SSFS** を記録し、その他の測定項目として時間と距離のデータログ、モビリティ記録、主観評価などを記録している。

### 2.3.4 開発の段階と実証試験

機器の開発は様々な段階を経過する。実証試験もそれぞれの段階に応じたものとなる。実証試験の内容は機器と開発計画によって異なる。生物医学分野においては、新規治療法、医薬品、医療機器の開発にさいしては臨床試験の相を表 3 のように定義し、各段階における臨床試験を位置づけている [31; 32]。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、

表 3 臨床試験の相

相	治療法 [31; 32]	
	主要点	目的
第 0 相試験	2006 年に FDA が提案。末期がん患者に少量の処方をし、薬物動態の観察を行うことによって新薬の開発を促進。	
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。 医薬品の開発の場合、健康成人に少量の投与により安全性や薬物動態を調べる。
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。医薬品の場合、比較的少数の患者を対象に有効性、副作用を検討。用法、用量を決定する。
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。 <u>RCT</u> の研究モデル。多数の被験者を対象とする。
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

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）の開発を促進することを提案した [33]。この手法は第 0 相試験として受け入れられつつあり [34]、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。

支援機器においても実証試験は開発段階に応じて位置づけることができる。我々は開発段階に応じた PICO による整理のためにも医薬品の場合と同様な「実証試験の相」の概念が有用であると提案してきた。研究計画の立案においてフォーカスすべきポイントを絞るための立ち位置が明確になるためである。表 4 に支援機器の実証試験に適用するときの相の定義を示す。ヨーロッパにおいても同様の提案がなされており [35]、今後普及するものと思われる。

第 I 相試験が基本機能と安全性に関する確認、第 III 相試験が市販のための最終試験である点は医薬品の場合と同様である。別の定義による混乱を避けるためである。第 0 相試験は FDA の提案からの連

表 4 支援機器実証試験の相

相	支援機器
第 0 相試験	現在使用中の機器の使用状況の観察、支援すべき身体機能の特性測定などを目的とし、介入のない、または最小限の介入（現在使用中の機器の問題点を確認・探索する場合など）を伴う試験。
第 I 相試験	試作した支援機器の基本機能と安全性に関する健常者による確認。
第 II 相試験	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。有用性・ユーザビリティ・適合・環境との整合性などの確認を目的とする。集会などで不特定多数の利用者による短時間の試用を含む。
第 III 相試験	市場に出す前の最終確認。様々な条件下で 20-40 名程度の利用者による本格試験。有用性・ユーザビリティ・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。有害事象の解明、適応の拡大などを目的とする。

想によるものであるが、機器開発の基礎段階における被験者実験を漫然としたものにしなないための提案である。開発、改良に着手するための検討段階を独立した研究として位置づけ、効率的に推進することを狙いとしたものでもある。第0相試験を定義することの是非については、さらに今後の検討が必要である。

第IV相試験はこれまでは組織的にはあまり行われてこなかった。2.3.2で述べたように支援機器の実証試験においては統計学的に有意性を主張できるだけのサンプル数をそろえることが必ずしも可能ではない。新規機種の発売時に1年間程度は改修が必要となるのは珍しくない。「モニター販売」としてディスカウント販売することもある。これらのフォローアップを組織的な研究として取り上げることを意図している。

## 2.4 研究計画の科学面

### 2.4.1 科学面とリスク・ベネフィット

#### (1) 実証試験における科学的取り組み

支援機器開発の伝統的な手法は、対象となる障害者のニーズを直接に支援する機器を試作し、対象者の試用状況の観察、使用感の聞き取りに基づいて改良を重ねてゆく手法であった。この手法では、特定の個人に対しては最適化が可能であっても、適応、適合条件を明らかにした一般化したエビデンスとするのは困難である。

我が国では最近まではユーザーの使用感やセラピストの直感による有効性の判定が普通であった。変化を促したのは医学研究の変革である。1990年代から医療分野においてEBM（Evidence-Based Medicine: 根拠に基づいた医療）が提案され、医学研究におけるエビデンスが重要視されるようになった。今世紀に入ってから科学的な研究法に関する日本語の参考書も入手できるようになり、今では、「効果・効能および適応・適合のエビデンス」を明らかにすることが支援機器の実証試験の主要な目的となっている。実証試験を科学的研究として計画することはこのようなエビデンスとして「一般化できる知見」を確立することである。

#### (2) ベネフィットの根拠としての科学性

研究計画が科学的に基礎づけられたものであること、科学的エビデンスの追求が実証試験のみならずあらゆる被験者実験あるいは人を対象とする研究における研究倫理の基本的要件であることは、工学系の研究者には必ずしも広く認識されてはいない。実証試験の科学性あるいは実証試験が科学的に妥当であることが必要であることはニュルンベルク綱領でも萌芽的には表現されていた。ヘルシンキ宣言1975年修正において明確な位置づけが与えられ、ベルモントレポートにおいて確立した。

ベルモントレポートの善行の原則は将来にわたって期待されるベネフィットがリスクを上回ることが研究を正当化するための要件であることを示したが、そのためにはベネフィットとしての期待に十分な説得力のあることが必要である。たとえば、「水がしゃべる」類の研究、「足裏の親指を刺激した。手のしびれがなおった。足裏が利いた。」という類の研究は科学的根拠がなく、主張されるベネフィットがいかに大きかろうと研究を正当化するためにリスクとの比較を行うことは意味がない。

なお、後者の論法は「雨乞いをした。雨が降った。利いた。」と同じであることから「雨乞い3た論法」[36]とも呼ばれる。雨乞いや足裏の例はわかりやすいが、「開発された機械で訓練した。足が強く

なった。「機械が利いた」ではどうであろうか。このような論法は訓練機器の実証試験でしばしば見受けられる。第一に「足が強くなった」といっても、どのような計測法によってどの程度強くなったのか明確ではない。第二に、比較対照が定義されていないために PICO の要件を満たしていない。場合によってはベースラインとして訓練前の筋力を訓練後の筋力と比較することもある。しかしこれでも十分な比較対照とはみなせない。対照となるのは異なった介入であり、筋力の増強の程度の介入による相違を比較すべきであるからである。これは、訓練効果が孤立した生物学的効果とみなすべきではないからでもある。少なくとも、社会的に意味のある訓練法は、現在普通に入手できる訓練手段と同程度かそれより大きいこと、あるいは効果は同程度でもコストの小さいことが求められる。このような効果の比較によらなければ客観的な意義は減殺される。標準的に普及している訓練法よりも筋力の増強度が小さかったとすればそのような訓練法に意義があるであろうか？それでも意義があると主張するためには、計測した訓練効果以外の点でのベネフィット（たとえばコストが低い）が求められる。このような理由で、「訓練した。強くなった。利いた。」とするのは社会的意義としては「雨乞い3た」論法に同じである。

以上述べたように、実証試験の研究計画が科学的に根拠づけられることは研究倫理の基本的要請であり、科学的に基礎づけられない研究計画は倫理的でない。この点特に銘記すべきである。

### (3) ベネフィットの評価

一方、ベネフィットについては、支援機器の開発に即してベネフィットを構成する価値観を整理すると以下ようになる。

- ① 工学的価値（機器の機構、工学上のブレークスルーなど工学の専門家集団の評価する価値）
- ② 医学的価値（その機器が人体機能に及ぼす効果によってユーザーが得られる効果に関して医療関連専門職の評価する価値）
- ③ 利用者にとってのベネフィット（その機器を利用することによって期待される心身機能の支援および ADL 改善効果）
- ④ 社会にとってのベネフィット（障害者がその機器を活用することによって社会に及ぼす効果）

これらの中で研究倫理の立場から優先されるべきは社会にとってのベネフィットであり、利用者にとってのベネフィットがこれに続く。工学的、医学的価値観に基づいたベネフィットを主張するならば、それが社会にどのように貢献するかを示さなければならない。支援機器開発においては、アカデミックな工学的価値のみでは実証試験におけるベネフィットと位置づけることはできない。研究倫理においてリスクとの衡量が求められるのは社会に対するベネフィットであるから。

工学上のベネフィットに関しては、工学上のブレークスルーによって利用者にもたらされる新たなベネフィットを第 II 相試験で検証することもあり得る。この場合、両者の関連を述べただけでは実証試験に即したベネフィットを表現したことにはならない。開発プランとしては開発着手時に工学上のブレークスルーが達成すべき工学上の目標を設定し、それによって利用者にもたらされるベネフィット、期待される社会的効果、付随する問題とその解決策、その見通しなどを明確にしておくべきである。実証試験としては工学上の目標が達成された段階で機器が意図通り動作することを検証する。これとは別に、工学上の目標は未達成でもその意図がある程度達成された段階でそれが利用者に意図した効果をもたらす可能性を検証する場合もある。いずれにせよ機器が意図通りに動作したときに期待される効果をベネフィットとして評価する。

支援機器が医学上のベネフィットをもたらすのは訓練機器によるリハビリテーション効果であろう。

これは当事者にとっても利益になることであるから利用者にとってのベネフィットとして位置づけることができる。社会にとってのベネフィットは直接的には機器の支援による参加のもたらす効果と考えて良いが、障害者の支援によって間接的にもたらされる社会への効果も含まれる。

以上、実証試験によるベネフィットについて見てきた。くどのようなであるが、その実現性が科学的に裏付けられたものでなければ評価に値しない。

#### (4) 安全性とリスクマネジメント

被験者実験においては被験者に対するリスクが最小化されており、安全であることが求められる。研究倫理はこれまで主として医学領域を対象としてきたので最近の安全工学の成果を取り入れることは希であった。しかし、医療機器のリスクマネジメントに関しては ISO [37]、JIS [38]、CEN [39]等 で標準化され、最近の支援機器の JIS にもリスクマネジメントが規定されるようになった。

このような状況を考えると、将来的には実証試験に際してもリスクアナリシスが必要になると思われる。リスクアナリシスの結果に基づいた主要な残留リスクが中心となろう。

研究倫理においては被験者の被るリスクとしてかなり定性的な表現で済まされることが多いが、リスクマネジメントにおいては、危険源 (hazard) によって危害 (harm) がもたらされる。危害のひどさとその発生頻度の組み合わせがリスクである。リスクはこの両者によって表現する。

リスクマネジメントの立場からは、安全は受け入れ不可能なリスクがないこと、あるいは残留リスクが許容可能なリスク以下であることと考えられている。「許容可能」というのは安全対策の手段がない場合やコストがかかり過ぎる場合などを含むので、残留リスクはゼロではない。ここに安全上の配慮が必要となる。

なお、危害のひどさについては、たとえば、死亡あるいは入院・通院を要する日数、現場での応急処置などで表現できるし、頻度については、たとえば、頻繁、時々、たまに、殆どなしなどで表現できる。臨床研究ごとに包括的なリスクマネジメントを行うのは困難であるとしても、せめてリスクは危害のひどさと頻度によって表現するようにしたい。

なお、リスクマネジメントについて、2種類の場合の区別に注意喚起しておきたい。通常の商品開発の場合は開発した商品の使用場面におけるリスクを対象とする。しかし、実証試験におけるリスクマネジメントは、設定した実験条件のもとにおけるリスクマネジメントである。機器の設計時に行った使用場面を想定したリスクマネジメントだけでは必ずしも足りるわけではない。たとえば、試作機では機器としての完成度が低いために完成機では発生しない事故の可能性もある。このため、実証試験ではより慎重な安全対策を取るのが通例である。

逆に、実証試験においてはスタッフを配置して事故に対処することもできる。機器を製品として利用する場合に比べると、機器そのものの残留リスクは大きくてもリスクを明確にし、スタッフの配置による残留リスクの低減も可能である。

支援機器の実証試験において特記したいのは、リスクが被験者の障害特性にも依存する点である。支援機器は何らかの機能障害を支援するものであるから、被験者は当然何らかの機能障害を有している。被験者の機能障害は機器が目的とする機能障害に限定しているわけではない。車いすの場合、片麻痺患者に半側空間無視があるかもしれない。その場合、環境条件によっては衝突などのリスクが無視できなくなる。この場合、実証試験の目的によっては半側空間無視のある被験者は除外するか、衝突などの事故が発生しないような対策を講じてリスクを低減させる。

容易に理解できるように、これらのベネフィットと被験者の受けるリスクとは方向の違ったベクトルに例えることができるし、同一尺度で定量化できるわけではない。そうすると、リスクとベネフィットを比較衡量するとはどのようなことになるか？定量的な比較は不可能でも定性的にはその大小を考察することはできる。

実際には、おおまかにはベネフィットが科学的に根拠づけられており十分に大きいと判断されたところで、リスクが最小化されているかどうかに着目する。リスクマネジメントによってリスクを許容リスク以下に低減させることが求められる。

## 2.4.2 科学性を裏付けるツール

### (1) 予測因子とアウトカム

実証試験においては様々な変数を制御、観測する。このとき、生物医学領域では人為的に制御する変数、時間的、生物学的に先行する変数を予測因子、介入の結果を表現する変数をアウトカムと呼んで区別する。介入としての実験の設定はそれ自身を予測因子と位置づける。予測因子は独立変数、アウトカムは従属変数と考えて良い。予測因子とアウトカムの適切な選定は実証試験を科学的に裏付けるための必須要件であり、選定が不適切であると科学的に妥当であるとは評価されない。独立変数と従属変数と呼んでもよさそうであるが、ここでは生物医学における用語をそのまま使うことにする。

この区別は、前提条件・原因と結果の取り違いを避けるために重要である。「予測因子: **predicative variables**」という用語は、結果の予測が可能であるという意味を持っており、性別、年齢、障害特性、等の個人の属性から介入条件の設定などを含む。

「アウトカム: **outcome**」という英語は結果事象を意味するが、介入の結果を観測するための変数と理解して良い。アウトカムは **PICO** の **O** を表しており、観測可能な変数のうち主要なアウトカムとしての選定が優れた実証試験であることの要件となる。

### (2) 仮説検定とエンドポイント

決定論的アプローチがそのままでは使えない被験者実験においては、生物医学領域で発展してきた推測統計学によることについては 2.3.2 で述べた。具体的には、推測統計学の仮説検定アルゴリズムを適用する。たとえば、支援機器がある機能障害の影響を受ける特定のタスクを支援することを検証するものとしよう。支援機器の使用によってスキルが向上するものと仮定する。スキルを表現するパラメータをアウトカムとして選定する。実験条件として、機器を使った場合と使わなかった場合のアウトカムを測定する。実証試験で主張したいのは機器を使用した場合のほうがしなかった場合よりはアウトカムが大きいという仮説である。この仮説を検証するために、この逆の仮説を帰無仮説として立てる。すなわち、アウトカムは使用しなかった場合の方が大きいという仮説である。この仮説に統計的検定を行い、帰無仮説を棄却して仮説の証明を行うものである。

このような背理法のアルゴリズムによって検証するものであるが詳細は推測統計学の教科書を参考にしていきたい。

上で、仮説を検証するためのアウトカム変数を測定すると述べたが、仮説の検証のために測定するアウトカム変数は特別にエンドポイントと呼ばれる。エンドポイントの本来の意味は、文字通り観測を終了するパラメータを表しているが、広義には目的を表現するパラメータとしてアウトカムと同義に使われることもある。仮説検証型の研究計画においては、仮説の検証のための鍵となるアウトカム変数とし

て扱われることが多い。

この場合、観測可能なアウトカムの中で実際に測定するものの選定、さらにその中からエンドポイントの選定は実証試験の質の決め手となる。当然仮説の検証に最も適したアウトカムによってエンドポイントを定義する。エンドポイントはアウトカム変数をそのまま使う場合もあるが、いくつかのアウトカム変数を組み合わせた合成指標とすることもある。

最も多いのはいくつかのエンドポイントを定義し、その中から主要エンドポイント、副次エンドポイントを定義することである。iBOT の本格試験においては、主要エンドポイントとして **Community Driving Test**、副次エンドポイントとして **SSFS** を定義している。付録の iBOT 紹介の解説ではエンドポイントの代わりに「アウトカム」と記載してあるが、これはオリジナル表記に従ったためである。

重要な点は、エンドポイントは可能な限り客観指標として定義すべきであることである。支援機器の評価には被験者や開発者の主観評価に頼ってきたために、つい主観評価に頼りがちである。「可能な限り」客観的に観測できる量をエンドポイントとして観測すべきである。iBOT の場合では、**Community Driving Test** は 15 のタスク達成のための自立度（介助の必要度合）によって 7 段階のスコアを割り付けている。副次アウトカムの **SSFS: Subject Specific Functional Scale** は被験者が自分の車いすで困難に感じている事項をタスクの自立度によって記録し、2 週間の使用後同じタスクについて iBOT での自立度を記録した。両者のスコアの比較が **SSFS** である。**SSFS** では困難なタスクについては主観的に決定するが、スコア自体はタスクの自立度として客観化された指標になっている。

ユーザビリティについても、設問項目を選びそれについて数段階の **VAS (Visual Analog Scale)**、その項目に関する主観的判定を数段階の数直線上の値で判定する) によって表現するのが通例である。このように試験と対照に関して主観の関与するスコアを記録する場合に注意すべきことは、独立事象としてスコアを記録すべき点である。時として比較した結果を **VAS (A と B の比較においてどの程度 A が優れているかなど)** で記録するプロトコルを見かけるが、記録対象である主観内容を比較する操作が含まれている。このとき混入するバイアスを否定できない。従って、質問項目を独立に計測することによって客観性を担保し、独立事象として記録、検定する。独立に測った結果を比較することに意義がある。

## 2.4.3 研究デザイン

### (1) 研究デザインの種類

生物医学領域で決定論的アプローチが使えないために、推測統計学による仮説検定のアルゴリズムによることは前項で述べた。このためのもう一つのツールが研究デザインである。一言で言えば、研究デザインは仮説を構成するための論理構成であり、比較対照の選定や被験者の選定／除外基準などにかかわってくる。

表 5 に生物医学領域で用いられる代表的な研究デザインを列挙した。

観察研究はプラクティスに変更を加えることなく観察する研究であり、症例報告など事実の記述を主と

表 5 代表的な研究デザインのタイプ

- |                  |
|------------------|
| ① 観察的研究          |
| (1) 記述的研究        |
| (2) 分析的研究        |
| (a) コホート研究       |
| (b) 横断研究         |
| (c) ケースコントロール研究  |
| ② 実験的 (介入) 研究    |
| A) 比較対照試験        |
| (a) 群間比較 (同時対照)  |
| i. ランダム化         |
| ii. 非ランダム化       |
| (b) 群内比較 (逐次対照)  |
| i. 自己対照 (前後比較)   |
| ii. クロスオーバー      |
| iii. シングルケースデザイン |
| B) 対照なしの研究       |



する記述的研究と観察した事実を分析して何らかの結論を導く分析的研究とに分けられる。分析的研究としては、時系列に沿っての変化に注目するコホート研究（前向きと後ろ向きとがある）、横断研究（ある時点における集団中の存在率のスナップショット）、ケースコントロール研究（有病者群と無病者群の2群の過去の暴露歴を比較する）等がある。

これら観察研究の研究デザインは、支援機器に関する被験者実験では第0相試験として観察研究を計画するとき必要となるが、第I相以後の介入試験に当たってはほとんど使わない。

## (2) 支援機器実証試験に用いられる研究デザイン

「介入」とは「通常の診療を超えた研究目的の診療行為」と普通には理解されており、支援機器に当てはめると、「通常の支援を超えた研究目的の実験」と位置づけることができる。新規開発した支援機器を試用することは通常の支援としてのプラクティスを超えているので、支援機器の試用による被験者実験は介入に相当する。臨床研究に関する倫理指針には「2つ以上のグループに分けてグループ間で効果を比較する」ものも介入研究として定義してあるが、2群に分割しての比較が通常の診療を超えているためと思われる。支援機器実証試験においては新規の機器を試用することだけで通常の支援を遙かに超えるので、2群への分割の有無にはかかわらず介入研究と位置づける。

介入研究は比較対照試験と対照なしの研究に分類され、比較対照試験には群間比較と群内比較がある。群間比較（同時対照）には、被験者をランダムに分割する場合とランダムではない場合とがある。群内比較（逐次対照）には自己対照とクロスオーバーがある。

### A) 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を比較し、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的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、データ解析担当者にどこまでに知らせないかによって、単盲検、二重盲検、三重盲検がある。

支援機器の実証試験においては、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) 群間比較試験 (Between-group Comparison)

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もので、同時対照(Concurrent Control)デザインとも呼ばれる。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手法であり、盲検化と組み合わせたものは最も科学的に信頼できる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ランダムに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のに要する時間によって性能を評価したとする。

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群に分割する。つまり、ランダムに分割することによって平均的に2つの群のスキルが同じになると仮定する。このことは、系統誤差の原因となる未知の因子（交絡因子）をならしてキャンセルすることをも意味する。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ーはそう多くはいないし、高齢で、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すぎる。また、既存の車いすとして普段使っているものと比較するとすれば、それらの性能の相違も無視できない。

選択基準において経験年数がある範囲に限定するなど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。このように限定すると、統計学的に有意性を主張できるだけ多数の（2群の平均的スキルが同じと見なせるだけ多数の）被験者をそろえるのは容易ではない。

支援機器の実証試験においてランダム化を採用せざるを得ないのは訓練機器の場合である。同一人物に対する異なった機器での訓練効果を比較することのできない場合、たとえば卒中の急性期には次に述べる群内比較が意味をなさない。それまでの訓練効果を消去（wash out）することができないためである。このような場合を除くと、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ない。

#### ii. 非ランダム化 (Non-Randomization)

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) 群内比較試験 (Within-group Comparison)

逐次対照試験 (Sequential Control Trial) とも呼ばれ、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、異なる介入に対するアウトカムの比較によって意味のある仮説検定が可能であり、支援機器の実証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 (Self-control Design)

前後比較試験 (Time Series Design) とも呼ばれるが、生物医学領域においては、同じ群に対して順次異なった介入を行い、それぞれに対する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、特定のテストコースについて、

① 現在使用中の車いすによる走行時間の測定

② 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、たとえば、②による走行時間が短いことを仮説とする。検証にあたっては、両者の値の差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。

支援機器の実証試験においては、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。比較対照の機器がない場合には、機器のない場合のスキル、あるいは現在一般的に行われている手段によるスキルとの比較も可能である。

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。また、訓練機器の場合は特に注意を要する。卒中の回復期の被験者の訓練の場合は、複数の訓練機器のうち、どちらの訓練を先に行うかによって効果が違って来るかもしれない。そのような場合には次項のクロスオーバーのデザインが有効かもしれない。しかし、急性期の場合には、被験者の障害特性の影響が大きすぎる可能性もある。対照とする訓練の特性、それに対する訓練機器の特性、障害と病歴、それまでの訓練歴などの関連を考察して適切な研究デザインを選択することになる。

#### ii. クロスオーバー (Cross-over Design)

対照群を2群に分けて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研究デザインである。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。

支援機器の実証試験においては、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、前項で述べた新規の訓練機器のように、複数の介入の比較の場合などには有効である可能性がある。

#### iii. シングルケースデザイン (Single Case Design)

N-of-1 デザインとも呼ばれ、一人の患者に試験薬とプラセボとを一定期間ずつランダムに投与し、個人内での薬効を比較し、統計検定にかけられるデザインである。行動科学の分野でも推測統計学には基づかない研究デザインとして注目されている。

広義では特定の個人の特定のニーズに即した試行錯誤による支援機器の開発もシングルケースデザインに含まれると考えられる。しかし、このような開発はあくまで特定の目的のための「支援」であり、プラクティスであるので研究とはいえない。個人に即しての単なる試作一試用の繰り返しだけでは研究デザインとは言えない。

いくつかのタスクに着目し、ベースラインと試作・改良した機器による場合のスコアを比較するなど、組織的な研究計画が必要である。支援機器の場合、観測したスコアのみを頼りに一般化するのは困難であり、生物学的妥当性による考察が欠かせない。

#### B) 対照なしの研究 (Uncontrolled Trial)

生物医学領域では対照のない介入研究はあまり使われない。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実証試験の場合、対照とすべき機器や比較すべき手段が存在しないため、対照なしのデザインとならざるを得ない場合もある。この場合、データの解釈には十分な注意が必要である。

以上見たように、支援機器の実証試験として通常最も使いやすい研究デザインは自己対照試験である。計画の策定において注意すべきは比較すべき機器の選定と比較すべきアウトカム変数の選択であり、結果に及ぼすバイアスの影響を小さくできる選択／除外基準および予測因子である。

なお、一般的ではないが、同時対照試験が可能であれば、RCTによる方が手間とコストはかかるがより説得力の強い研究デザインとすることができる。

### 2.4.4 科学性に影響する因子

## (1) 選択／除外基準

被験者の選択／除外基準は **Inclusion/Exclusion Criteria** 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がターゲットとする機能障害は一般に機器ごとに定義されているが、実験手順に即してより具体的に被験者の選択／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、研究効率の高い被験者集団を定義することであり、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、実験に際してリスクが大きく有害事象の可能性があるなどの理由で被験者から除外する基準である。理工系の研究者の作成した研究計画の中には、現場の理学療法士に選定してもらうといった選択／除外基準をしばしば見受ける。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り、意図的な選択バイアスを疑われても反論できない。専門の医師に相談してあらかじめ明確に決めておいた選択基準、除外基準に即した被験者の選択は理学療法士に依頼して良い。

選択基準は理工学における材料の組成、処理、加工に相当する。例えば半導体デバイスの研究の場合、組成、組成分布、幾何学的形状、熱履歴などを厳密に規定する必要がある。被験者の選択基準も同様で、必要な程度に厳密に定義されているかどうかでその研究の質が定まる。支援機器の場合は対象とする障害像を想定して開発を進めるわけであるが、想定した障害像を純粹に有する障害者は希である。障害の程度、態様は様々であり、様々な合併症、二次障害を伴っているのが通例である。

WHO の生活機能分類 (ICF) に即して考えたとき、心身機能・身体構造・活動制限のどのレベルで選択／除外基準を定義するかは機器の特性による。活動制限のレベルで定義したときは、それに関わる心身機能・身体構造について明確にしておくべきである。

選択／除外基準の具体例としては、iBOT の本格試験の選択基準、除外基準が参考になる。選択基準は対象とする障害、実験を可能とする要件から比較的容易に決まる。除外基準では、被験者候補に発生する可能性のある二次障害、合併症、既往症等に関わるリスクについて広い配慮をしている。

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を防ぐこと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とすれば、実験条件の科学的検討が不十分であるとのそしりを免れない。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協力を求め、耳を傾けることが求められる。

## (2) バイアス

研究デザインの科学性の担保のためには、バイアス(系統誤差)への配慮が不可欠であり、実証試験の科学性の検討においては欠かせない視点である。また、バイアスを排除するために様々な工夫がなされている。

バイアスには大別して選択バイアス(被験者の選択の偏りによってもたらされる系統誤差)、測定バイアス(実験、調査で集めるデータに偏りがある場合の系統誤差)、交絡バイアス(真の原因が観察因子の後ろに隠れていて因果関係が適切に把握されていない場合)等がある。

選択バイアスを排除するとともに研究デザインの生物学的妥当性を保証するのが選択／除外基準である。測定バイアスは測定操作に関わるバイアスで、面接者と被験者のやりとりの中に系統誤差が混入することによって引き起こされる。記憶バイアス (面接者が被験者の記憶を聞き出すのに、対照群に比