

大学院医学研究科シリーズ

人を対象にした研究のデザイン

伊木 雅之

近畿大学大学院医学研究科疫学・健康科学

1. 緒言

知識には使える知識と使えるまでには至らない知識とがある。教科書や論文から学んだ知識はすぐには目の前の患者には使えない。上級医の指導や同僚の協力があってはじめて使うことができる。使った知識は経験という糧を得て次第に使える知識に育っていく。このような豊富な経験がなければ、臨床医学は成立しない。臨床医が経験に支えられているのは当然のことだ。しかし、人間は成功体験を採用し、失敗体験を忌避する。「使った、治った、効いた」の成功体験が必ずしも正しくないことは自然治癒があることを指摘するだけで明白だ。臨床研究に求められていることは経験以外に頼る術のない局面をできるだけ減らすこと、そのためのエビデンスの積み重ねである。それには人を対象にしながらも、できるだけ科学的な判断をすることのできる研究をデザインできなければならない。無作為割付比較試験(randomized controlled trial, RCT) がその代表だが、いつもそのようなデザインが採用できるわけではないし、最適なわけでもない。目的と状況に応じて柔軟な対応を取る必要がある。本稿では、臨床研究でよく見られる目的に応じた代表的な研究デザインを取り上げ、解説する。

2. 人を対象にした研究の目的

人を対象にした研究、中でも臨床研究の目的の多くは以下の3点に集約される。

・薬剤など治療の有効性の確認

臨床研究の最も多くを占めるのが治療の有効性評価である。治験で行われる第I相試験からIV相試験がその代表的なものだ。第I相試験は薬剤の体内動態の把握や重大な有害作用を起こさない忍容量の決定を目的に行われる少人数の試験、第IV相試験は薬剤の認可後、期待される有効性が発揮できているかどうかや希な有害作用

を発見するためのサーベイランスとして行われるもの。本稿では第II相、第III相で行われる薬剤の有効性を評価する研究デザインを取り上げる。

・検査の有効性の確認

検査の目的は疾患の診断と疾患の治療経過のモニターに大別される。本稿では前者の診断的検査の有効性評価のデザインを扱う。

・予後予測因子の検索

診断が確定すれば、次なる関心事は治療の選択である。EGFR 遺伝子変異陽性非小細胞肺癌にはゲフィニチブ、HER2 陽性の転移を有する乳癌にはトラスツズマブ、といった個別化医療を推進するには、その治療の有効性に大きく影響する要因を把握することが必要である。そのような予後予測因子を発見するデザインを扱う。

3. 治療の有効性を評価するためのデザイン

(1) Before-After design

問題とする疾患に対して従来から行われている医療行為について、有効性データ (before) がすでに得られている場合、新しい医療行為の有効性データを得る臨床研究 (after) のみを実施し、両者を比較して、後者の有効性を評価するもの。治療の前後での患者の状態の比較を意味しているのではない点に留意されたい。

本デザインの概要は図1に示した通りで、実際に行うのは図の中央から右の部分で、ここから得られるアウトカムをすでに得られている標準治療のアウトカム (図の中央から左の部分) と比較する。

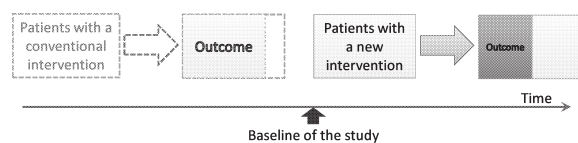


図1 Before-After design

具体的な手順は表1に示す。①は本デザインを実施する前提となる事項だが、実はこれが無い場合、すなわち有効な治療法がなく、標準治療と言えるものが定まっていない場合にもこのデザインは採用される。標準治療は定まらないが、例えば、1年生存率は20%程度というような信頼できるデータがあればよい。⑤のアウトカムは①で得られるデータと比較できる指標を主要アウトカムとし、追跡期間も①に応じて設定する。それ以外に意味のある指標は副次的アウトカムとしてモニターする。⑥の予定症例数は指標によって計算方法が異なるが、ロジックは以下の通りである。①で得られた基準値に対して、新薬に対しては、他の研究データから実際に期待できる値を設定するか、臨床上有意義な有効性を上積みした値、すなわち、これだけ従来薬を上回らないと臨床意味はないと考えられる値を設定する。この新薬での期待値と基準値との差をある検出力（たとえば80%）で有意差（5%）として検出できる症例数を計算する。

このデザインの利点は、1群の患者だけで実施できるので、費用や労力が少なく済むことだ。しかし、重大な欠点は、過去の有効性データと新薬の有効性データが比較性可能であるという保証がないことである。過去の研究で用いられた患者群と実際に調査された患者群が同等であるという保証がない。診断基準や組み入れ基準、除外基準を合わせていても、得られた患者群の年齢、性別、重症度などが一致しているとは限らない。また、補助療法や一般的なケアの方法が同じかどうかはわからない。一般的にケアは時代と共に向上し、それが予後にも影響しているはずである。となると、当該治療が標準治療と同等であっても、他のケアの時代的な向上の結果、後で行われた研究における予後は過去のデータより

改善するだろう。Before-After design では以上のような評価できない諸々の差が、結果の信頼性を低下させている。これらの弱点を克服するには、比較性の保証された標準治療群を自ら設定する他なく、それは次項の Concurrent design を用いることである。

では、Before-After design を用いる意味はないのかということそうではない。費用と労力がかかり、多くの患者を巻き込む Concurrent design を実施するには、ある程度の有効性が期待できねばならない。また、有効性があり、かつ有害作用が少ない投与量や投与方法が決定されていなければならない。そのような準備研究として行うには適したデザインである。

(2)並行群間比較デザイン Concurrent design

新しい治療だけでなく、比較対照となる標準治療についても平行して有効性を評価し、両者を比較するもの。まったく新しい治療で、比較できる標準治療が無い場合には偽薬を比較対照とする。本デザインの概要は図2に示した通りで、患者を新治療群と対照治療群のいずれかに割り付け、それぞれ割り付

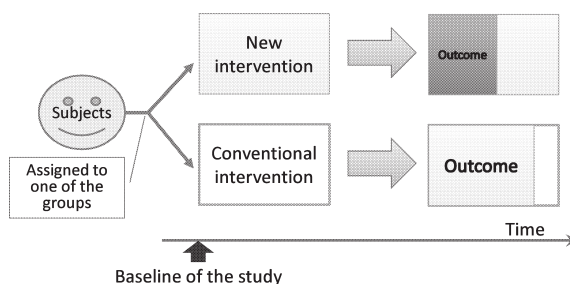


図2 平行比較介入デザイン

表2 並行群間比較デザイン Concurrent design の実施手順

表1 Before-After design の実施手順

- ①当該医療行為と比較する医療行為について比較に耐えうるデータがあることを確認する。
- ②当該疾患の診断基準を確定する。
- ③研究参加を求める患者の組み入れ基準と除外基準を作成する。
- ④当該医療行為を標準化する。
- ⑤アウトカムを決定し、追跡期間とモニターの仕方を決める。
- ⑥予定症例数を決定する。
- ⑦該当患者を登録し、必要な介入前検査を行う。
- ⑧介入を開始し、予定された追跡期間中のアウトカム発生をモニターする。
- ⑨予定した症例数が登録されたら、登録を終了する。
- ⑩予定された追跡期間が終了したら、集計し、アウトカムを①のデータと比較する。

- ①当該治療と比較する対照治療を特定し、標準化する。
- ②当該疾患の診断基準を確定する
- ③研究参加を求める患者の組み入れ基準と除外基準を作成する
- ④アウトカムを決定し、そのモニターの仕方を決める。
- ⑤予定症例数を決定する。
- ⑥該当する患者があったら登録し、いずれかの治療群に割り付ける。
- ⑦必要な介入前検査を行う。
- ⑧介入を開始し、予定された追跡期間中のアウトカムをモニターする。
- ⑨予定した症例数が登録されたら、登録を終了する。
- ⑩予定された追跡期間が終了したら、集計し、アウトカムを2群間で比較する。

けられた治療を実施し、アウトカムを追跡する。対照治療についても当該臨床研究の一部として、新治療と同時並行して実施される点が Before-After design と異なる。患者を新治療と対照治療とのいずれかに割り付ける方法によって2つに分かれる。

共通する手順を表2に示す。介入内容の異なる2群を並行して追跡し、アウトカムを比較すること以外は Before-After design と大きくは変わらない。

①の対照治療についても標準化しておき、対照群では同じ治療が行われるように管理することが必要だ。標準治療と呼べるものがない場合には、実薬と区別がつかない偽薬を用いることもある。一方、④のアウトカムの設定は過去のデータに縛られることがないので、最も必要なものを自由に設定することができる。⑤の予定症例数の決定のロジックは同じだが、すでに定まった標準治療の値と比較する Before-After design と異なり、比較対照群も設定して調査をするので、必要な例数は Before-After design の2倍よりも多くなる。⑥の割り付け法が非無作為割り付けか、無作為割り付けかで2つのデザインに分かれるが、それ以外の手順は基本的には同じである。

① 非無作為割り付けデザイン Non-randomized design

無作為ではない何らかの方法で患者を新治療と比較対照に割り付けるもの。日常診療では医師が患者に最適と考え、患者がそれを受け入れた治療が行われるが、そこでの治療方法の選択を研究上の割り付けとして採用することが多い。

[利点と欠点]

本デザインの利点としては、当該治療以外はまったく同じように実施できれば、新治療と対照治療がある程度比較でき、割り付けが日常診療内で行われるので、診断から治療に入るまでの研究もたらす時間のロスがないこと、割り付けにかかる手間が少ないことが挙げられる。

欠点としては、二群を並行して調査するので、Before-after design の2倍以上の患者が必要で、費用も2倍以上かかる。さらに重要なことは、比較する2群の患者特性が同じである保証がないことだ。医師が医療行為を選択するには選択するだけの理由があり、患者の選択にもそれなりの理由がある。したがって、選択された治療内容によって患者の特性が異なる可能性は十分にある。その特性が予後に影響するものであった場合、新法と従来法の有効性に差があっても、それが医療行為の違いに起因するのか、その理由に起因するのかが判断できなくなる。また、治療法が異なれば、治療に付随する検査やケ

アの内容が異なることは十分に考えられ、それがアウトカムに影響する可能性もある。

非無作為割り付けデザインでは多くの場合、医療内容が患者、主治医のいずれにも知られているので、患者には偽薬効果 (placebo bias)、すなわちまったく効果のない偽薬であっても治療を受けているという安心感で症状が多少とも改善する効果が現れるが、その大きさは医療行為の内容によって異なる可能性がある。また、主治医が割り付けを知っていると割り付けによって効果判定の仕方に影響がでるいわゆる観察者バイアス (observer bias) が生じてしまう。

[対策]

これら欠点に対して、まったく対策がないわけではない。患者の特性の違いについては、測定できる差、たとえば、性別、年齢、進行度、組織型などであれば、統計解析段階でマッチングや多変量解析を用いてその差を調整できる可能性はある。しかし、測定できないものやそもそも未知の特性に差があるかもしれない、これらは調整のしようがない。プロトコルで検査やケアの群間差についてはそれらの実施法を定め、できるだけ同じように扱うことが必要である。また、2群間の偽薬効果の差はいかんともしがたいが、患者の評価における観察者バイアスは、割り付けを知らない別の医師が行うか、主観の入らないアウトカムを用いることによって除くことができる。

② Randomized design 無作為割り付けデザイン

患者の試験治療群か対照治療群への割り付けを、くじ引きで決めるように、被験者、実施者のいずれの意図も交えずに、無作為に行うもの。

[利点と欠点]

利点は、非無作為割り付けデザインの最大の欠点であった両群間の患者特性の差がなくなると期待されることだ。無作為割り付けデザインでは患者の割り付けに何の意図も入らない。ただ偶然によっていずれかの群に割り付けられるのだから、割り付けに起因する患者特性の系統的な差が生じるはずはない。これは測定可能な要因の群間差に止まらない。測定できない要因や未知の要因も含めて両群間に差はないと考えることができる。これによって、当該治療以外に2群間の違いはなくなるので、アウトカムの差は介入の違い、すなわち試験治療と対照治療の違いによるものと結論される。無作為割り付けデザインは、治療の有効性を証明する単独の研究デザインとしてはもっとも強力なものである。

欠点は非無作為割り付けデザインと同様に、Before-after design の2倍以上の患者が必要で、費用も2倍以上かかることと、通常、無作為割り付けは患者

と主治医の間では決められず、患者を登録した後で、研究の管理者が患者の的確性を確認した上で、無作為割付けし、それを主治医に伝えるという手順をとる。そのため、その場で治療を開始できないし、いずれの治療が受けられるかが分からないという点に対する患者の理解も得にくく、説明に通常より時間がかかってしまう。特に対照治療に後述する偽薬が使われる場合はなおさらである。

[無作為割付の種類]

最も簡単な無作為割付の方法は封筒法と言われるものである。割付を記した紙の入った封筒を多数つくっておき、該当患者が出現する度にその封筒の束から1枚引き抜き、開封して、記された群に割り付ける方法だ。これがきちんと行われれば、無作為割付となり、その場で割付が分かって治療も開始できる利点はあるが、開封結果を見てから患者が研究参加を取りやめたり、希望する群が出るまで封筒を引き直すような違反行為をなくすることが難しい。単純な無作為割付はできても、次に述べる層化無作為割付法は実施困難というような問題もある。

封筒法や患者登録センターで乱数を発生させて、たとえば奇数ならば試験治療群、偶数ならば対照治療群と割り付けるのが単純無作為割付法である。しかし、この方法では偶然の所産として2群間の年齢に差が出たり、介入前検査に差が出ることもある。そこで、重要な予後決定要因については予め層別し、各層の中で無作為割付を行う層化無作為割付法がある。これによれば、重要な要因についての群間差は出なくなる。この方法は標本数が少ない場合には実施しにくく、小型の無作為割付デザイン用には最小化法が考案されている。

[割付重視の解析 Intention-to-treat analysis]

無作為割付デザインが最強の研究デザインであるためには、研究の最後の最後まで無作為割付された状態を維持しなければならない。これはそれほど簡単なことではない。患者の脱落や群間での乗り換えが生じてしまうからだ。研究に参加した患者が何らかの理由で参加そのものを取りやめることはよくあるし、副作用や効果が不十分なためにプロトコル治療を中止したり、新たな治療を併用せざるを得ないこともままある。脱落には脱落する理由があるわけで、試験治療に副作用が高頻度に出て、それによって患者が脱落すると、試験治療群では副作用に強い患者が多くなって、対照治療群と患者特性が異なってしまう。しかもそれが治療効果と関連していれば、アウトカムに本来の治療効果以外の影響が混入してしまう。では、脱落した患者をどのように扱うか。脱落しなかったものとして、最初に割り付けられた

群に残して最後まで解析する割付重視の解析 (Intention-to-treat (ITT) analysis) が行われる。

図2のように各群に割り付けられた患者が最後まで追跡できて、全員のアウトカムが判明すればよいが、現実には図3のように脱落 (dropout) が出てしまう。これを除いてアウトカムが確定できた患者のみの集計をするのがプロトコル重視の解析 (Per protocol set (PPS) analysis) で、脱落した症例も元々の群に残して解析するのがITT解析である。後者では無作為割付によって得られた2群間の同等性が維持され、解析結果は介入の違いを反映するものとなる。しかし、脱落の中には、通常、試験治療を止めて標準治療に移行した者が多く含まれる。となると、ITT解析では試験治療の効果に標準治療の効果が含まれ、純粋に試験治療の効果を見ているのではなくてしまう。それに対してPPS解析では試験治療を完遂できた場合の効果を示している。これは理想的な条件下で試験薬が発揮できる最大の効果、すなわち効能 (efficacy) を表し、ITT解析では、副作用による脱落は試験治療の失敗なので、その失敗を含めて、治療全体の現実的な効果 (effective-

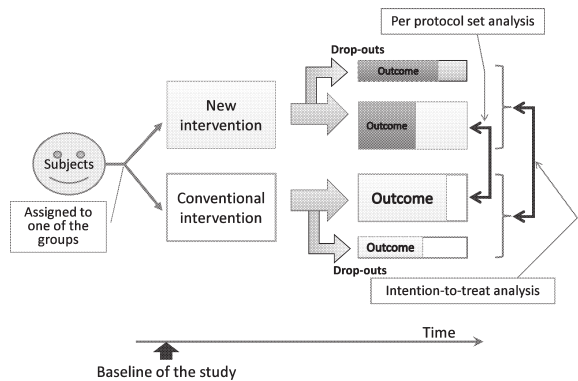


図3 プロトコルからの逸脱例の解析

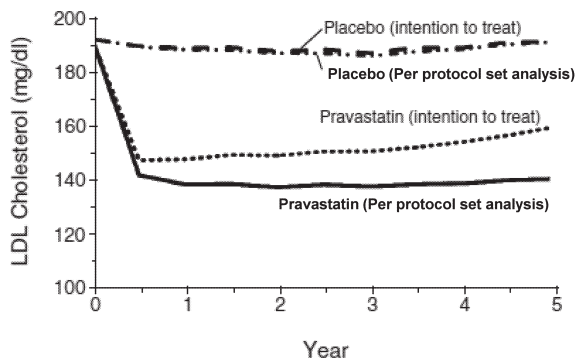


Figure 1. Effects of Pravastatin Therapy on Plasma LDL Cholesterol Levels.

(Shepherd J, et al. N Engl J Med. 1995;333:1301-7)

図4 割付重視の解析とプロトコル重視の解析

ness) を評価するものと言える。実際に、ITT 解析解析では PPS 解析よりも効果の大きさが小さくなる。図 4 は中等度の高コレステロール血症はあるが、心筋梗塞の既往のない45-64歳の男性6595人で、pravastatin の LDL コレステロールへの影響をみた West of Scotland Coronary Prevention (WOSCOP) Study という RCT の結果である¹。pravastatin 群では placebo 群に比べて LDL は大きく低下しているが、ITT 解析では PPS 解析より低下の大きさが縮小している。ITT は試験治療には不利な解析法で、不利な解析をしても、なお有効であることを示すものと言える。

[盲検試験]

非無作為割付デザインでは多くの場合、医療内容が患者に知られているので、偽薬バイアスが生じうる。無作為割付デザインでは患者への治療の割付は通常、患者登録センターで行われるので、割付の段階では患者にも主治医にも分からない。そこで、いずれの群に割り付けられたかを知らさないまま試験に入るのが盲検試験である。患者にのみ割付を伏せるのが単純盲検試験、主治医にも伏せるのが二重盲検試験、データ解析者にも伏せるのを三重盲検試験という。単純盲検によって偽薬バイアスをなくし、二重盲検によって観察者バイアスと試験治療以外のケアの意図的な差もなくし、三重盲検により意図的な解析も無くすることができる。ただし、薬剤の場合は試験薬と対照薬の区別ができない特別な製剤が必要だし、手術やりハビリのように違いを無くすことが難しい場合は盲検試験にならないこともある。

4. 診断的検査の有効性を確認するためのデザイン

理想的な診断的検査では、患者はすべて陽性となり、非患者はすべて陰性となる、あるいは、陽性者はすべて患者で、陰性者はすべて非患者である。それを目指して、新たな検査の開発とその有効性を評価する臨床研究が行われている。このためのデザインは症例対照デザイン Case-control design で、後ろ向きではなく、時間の流れを想定しない一時点での断面研究 Cross-sectional study となる。

[実施手順]

症例対照デザインの実施手順を表3に示した。ここでのポイントは当該疾患の診断である。診断は現状でもっとも確実な方法 (Gold standard) でなければならない。したがって、試験検査は Gold standard を超える検査を目指すのではなく、より簡便、より安全、より安価に Gold standard 検査に近づくことを目指している。切除された組織の病理診断で確定された腫瘍について血液マーカーの診断能を検討す

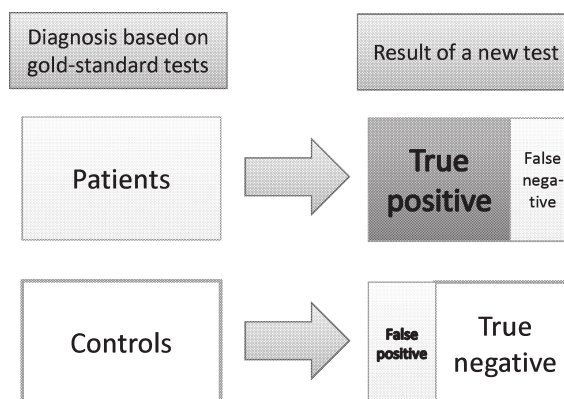


図5 診断的検査の有効性を評価する症例対照デザイン

表3 症例対照デザインの実施手順

- ①当該疾患の診断基準と確実な診断方法 (Gold standard) を確定する。
- ②研究参加をを求める患者と非患者の組み入れ基準と除外基準を作成する。
- ③試験検査の実施方法と判定方法を標準化する。
- ④予定症例数を決定する。
- ⑤該当患者を登録し、Gold standard 検査とは全く独立して試験検査を行う。
- ⑥該当対照者を登録し、Gold standard 検査とは全く独立して試験検査を行う。
- ⑦予定した症例数が登録されたら、登録を終了する。
- ⑧感度、特異度を求め、ROC 解析を行う。

	Test results		Total
	Positive	Negative	
Cases	TP	FN	TP+FN
Controls	FP	TN	FP+TN
Total	TP+FP	TN+FN	

図6 診断的検査の有効性を評価する症例対照デザインの結果をまとめた四分表

TP : true positive FP : False positive
TN : True negative FN : False negative

るといったものだ。もう一つのポイントは、試験検査は Gold standard 検査と独立して実施されなければならないことだ。試験検査が Gold standard 検査の一部を含んでいてはいけないのは当然だし、検体検査であれば、患者からの検体か対照からの検体かは盲検にする、身体診察が必要であれば、診断医に患者か対照かは伏せることも必要である。

[解析]

症例群、対照群それぞれについて検査陽性と検査陰性があるので、結果は図6の四分表にまとめることができる。この結果から、

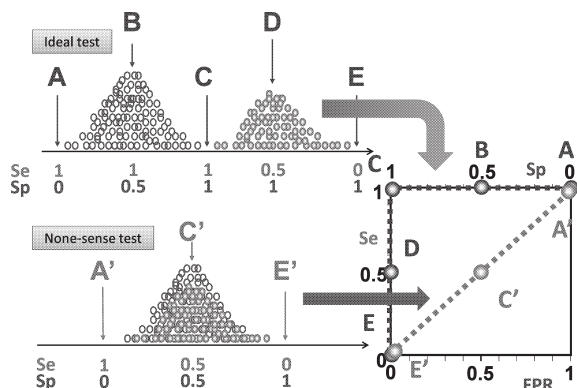


図7 ROC (Receiver operating characteristics curve) 曲線

$$\text{感度 sensitivity (Se)} = \frac{\text{真陽性数 True positive}}{\text{症例数}}$$

$$= \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{特異度 Specificity (Sp)} = \frac{\text{真陰性数 True negative}}{\text{対照数}}$$

$$= \frac{\text{TN}}{\text{TN} + \text{FP}}$$

及び,

$$\begin{aligned} \text{偽陰性率 False negative rate (FNR)} \\ = \frac{\text{偽陽性数 False negative}}{\text{症例数}} = \frac{\text{FN}}{\text{TP} + \text{FN}} \end{aligned}$$

$$\begin{aligned} \text{偽陽性率 False positive rate (FPR)} \\ = \frac{\text{偽陽性数 False positive}}{\text{対照数}} = \frac{\text{FP}}{\text{TN} + \text{FP}} \end{aligned}$$

が計算できる。いずれも0から1の値をとり、SeとSpは大きいほど、FNRとFPRは小さいほど有効な検査である。なお、これらには、

$$\text{FNR} = 1 - \text{Se}$$

$$\text{FPR} = 1 - \text{Sp}$$

の関係がある。

試験検査が連続変量として得られる場合は陽性と陰性の境界値 cutoff も連続的に変化させることができる。そこで、横軸にFPR、縦軸にSeをとって、cutoffを連続的に変化させ、そのcutoff毎にSeとFPRを求めてプロットし、Receiver operating characteristics curve (ROC 曲線) を描く。図7に示したように、カットオフをAからEへと動かし、それぞれのSeとFPRからROC曲線を描くと、完全な検査の場合は左上角を通る二辺となり、まったく無効な検査の場合は左下がりの対角線となる。通常の検査のROC曲線は両者の中間に位置し、左上角に近いものほど有効な検査と判定される。これを数的に表した検査の全般的有効性の指標として、ROC曲線下面積 Area under the curve (AUC) が

ある。これは0.5から1の値をとり、大きいほど有効である。

cutoffは目的に応じて決めるが、通常はSe, Spのいずれも高くしたいので、ROC曲線が左上角に最も近づく点に設定される。

5. 予後予測因子を検索するデザイン

診断ができれば、次なる関心は治療法の選択である。どんなに有効な治療法にも完璧はありえず、必ず有効性には個人差が発生する。個々の患者にとってもっとも有効な治療法を選択して実施することをTaylor-made medicine, 近年では個別化医療 personalized medicine と呼び、冒頭で述べたEGFR遺伝子変異陽性非小細胞肺癌やHER2陽性転移性乳癌等のように実用化されているものもある。多くは遺伝子の変異や発現から予後予測をしようという試みだが、遺伝子に限る必要はない。身体特性や代謝指標はもとより喫煙、飲酒など行動要因も含めて、検討が進んでいる。

[実施手順]

この目的で用いられるコホートデザイン Cohort design を図8に示した。対照となる一群の患者に予後予測因子となる可能性のある様々な特性を評価しておき、その後、治療経過を観察して、アウトカム

表4 コホートデザインの実施手順

- ①当該疾患の診断基準を確定する。
- ②研究参加を求める患者の組み入れ基準と除外基準を作成する。
- ③アウトカムを決定し、その把握方法と追跡期間を定める。
- ④予後予測因子候補を決定し、測定法を定める。
- ⑤予定症例数を決定する。
- ⑥該当患者を登録し、予後予測因子の測定を行う。
- ⑦予定した症例数が登録されたら、登録を終了する。
- ⑧予後予測因子候補の有無別にアウトカムの発生率を比較し、予後予測因子を抽出する。

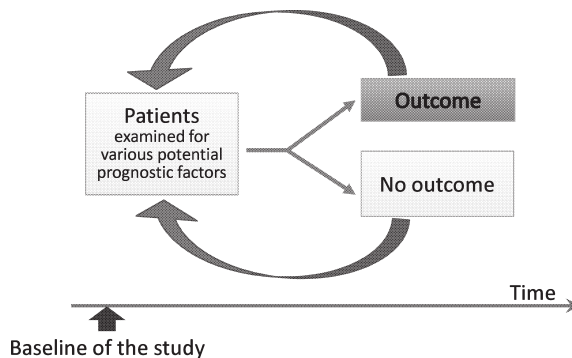


図8 予後予測因子を検索するコホートデザイン

を起こした患者と起こさなかった患者を特定する。そこからベースラインを振り返って患者の特性値とアウトカムの生じ方の関係を解析し、アウトカム発生を予測できる要因を検索、抽出する。

ベースラインにおける患者特性はすべてが測定できるわけでない。そこで、患者の血清や場合によってはDNAを保存しておき、アウトカムが確定後に測定することも想定しておく。

[解析]

予後予測候補因子の有無別にアウトカムの発生率を比較するのが基本になる。アウトカムが何らかの検査値である場合は平均値や中央値を比較する。これらに差が無いような因子に予測力は期待できないので、候補から外す。EGFR 遺伝子変異陽性非小細胞肺癌や HER2 陽性転移性乳癌の例のように、単一要因で大きな予後予測能が見つければ、臨床的に利用価値が大きい。しかし、複数の要因の組み合わせた予測モデルを作成し、予測能を高める試みもある。

たとえば、向こう10年間の骨粗鬆症性骨折発生確率を予測する FRAX[®] モデル² は骨強度指標である骨密度に加えて、年齢、性別、骨折既往、骨折の家族歴、ステロイド使用歴など10の要因を組み合わせている。このようなモデルの作成には多重ロジスティック回帰分析が使われることが一般的で、アウトカムが検査値のような連続変数である場合には多変量線形回帰分析が用いられる。

[予後予測モデルと因果モデル]

予後予測モデルの有効性はいかに高い確率でアウトカムの発生を予測できるかということで、それを高めるために必要な要因をモデルに含めていく。その多くはアウトカムの発生要因になるが、アウトカムと因果関係を想定する必要はない。極論すると、まったく無関係と見える要因や明らかな交絡要因であっても予測能が高まるものならば採用して良い。

一方、個々の要因とアウトカムの間の因果関係を推定しようとする因果モデルは、形式的には予測モデルと同じであるが、まったく無関係の要因をモデルに入れておくわけにはいかない。そのような要因はその研究では測定されていない要因や未知の要因と関連があって、それらの要因の影響を表現している代理要因と考えられるからだ。因果モデルでは真の発生要因を特定しなければならない。したがって、年齢、性別、他疾患の既往や現病歴、あるいは喫煙、飲酒といった生活習慣要因を含む同じモデルであっても、予測モデルではモデル全体の予測能を上げることに関心があり、因果モデルではそれぞれの要因の影響の大きさに重み付けすることに関心があるという違いがある。

[多重仮説の問題]

コホートデザインでは多くの候補要因の中から予後予測に有用な要因を抽出することになるが、たとえば、100の要因をテストして、5要因が有意水準5%でアウトカムと有意な関連を示したとしよう。果たして、これに意味があるか。有意水準5%とは、統計的仮説検定で帰無仮説を棄却した場合に誤る確率が5%ということ、ここでの意味は、真実には予後予測に役に立たない要因を、誤って役に立つと判断してしまう確率だ。そのような確率が5%ということは、100個の役に立たない要因をテストして偶然役に立つという結果がでる要因の数の期待値は5個となるので、上記の結果は偶然の所産かもしれない。このような多重仮説検定の弊害を避けるためには、100回の検定全体で、偶然有意な結果を得てしまう確率を5%未満とする調整が必要になる。Bonferroni法では、100回の検定をする場合には5%の100分の1、即ち0.05%を5%と扱うことにする。しかし、これは中々厳しい検定法だ。何でもかんでも検定するのではなく、仮説を絞って検定するのが賢明である。

[モデル生成コホートとモデル検証コホート]

以上のようにしてアウトカムをたいへんよく予測できるモデルが得られたとする。それで十分だろうか。予後予測モデルは新しい患者に適用して予後を予測できねばならない。確かにそのモデルは、そのモデルを生成したコホート (Development cohort) ではアウトカムをよく予測できたかもしれないが、そのコホートに属さない個人に適用して同様に有効かどうかはわからない。そこで、モデル生成コホートとはまったく別のコホート (モデル検証コホート Validation cohort) に適用した場合に、同様にアウトカムを予測できるかどうかを確かめる必要がある。予後予測モデルは、予測に用いられる要因の値が、生成コホートで存在する値の範囲を超える場合には適用できない。さらに予後予測モデルはその生成コホートに最適化されたモデルとなっているので、他のコホートに適用すると、たとえ似たようなコホートであっても、予測性は生成コホートにおけるよりも低下する。低下してもなお有効なレベルかどうかを確認する必要があるわけだ。

そこで、モデル確認コホートを立ち上げ、予測に必要な要因を測定し、アウトカムを予測し、実際のアウトカムを確認して予測と実測が期待する程度に一致していることを確認する。しかし、このようなことをすると研究期間がモデル生成と同程度かかってしまう。これを回避するために、コホートの追跡が完了した段階で対象者を無作為に2群にわけ、一

方でモデルを生成し、他方でモデルを検証すれば、研究期間は短くて済む。ただし、必要な症例数は2倍になる。

6. 結 語

本稿の表題にある「人を対象にする研究」とは、筆者の考えでは疫学研究で、臨床研究は広義の疫学研究に含まれる。本講ではあえて臨床研究を取り上げ、その目的でよく見られるものに必要な研究デザインに絞って解説をした。それは近畿大学大学院医学研究科では臨床医学系の学生が多く、臨床研究のニーズが高いからだ。そのため、疾患のリスク要因やベネフィシアル要因を特定することを目的とする疫学研究については述べていない。また、膨大なカルテというデータベースを用いて実施可能な回顧的コホート研究やコホート内症例対照研究も紙面の都合で解説しなかった。また、Cochrane library や systematic review で用いられるメタ分析も重要なデザインである。これらについては参考文献を参照されたい。

7. 参考文献

本稿では十分に解説できなかったデザインや取り上げなかったデザインの中にも重要なものはある。研究を企画する上で、参考になる書籍を紹介する。

①木原正博訳：医学研究のデザイン。医学書院
人を対象にした研究デザインについて平易にまと

めた良書。初学者に好適。

②中村好一：基礎から学ぶ楽しい疫学。医学書院
臨床研究だけでなく、疫学研究一般を初学者が学ぶには良い書籍。

③山口拓洋：サンプルサイズ的设计。健康医療評価研究機構

様々な必要症例数の計算方法をマニュアル化した役に立つ本。

④丹後俊郎：無作為化比較試験。朝倉書店
RCTについて詳しく解説した良書。

⑤浜島信之：多変量解析による臨床研究。名古屋大学出版会

多重ロジスティック回帰モデルとCox比例ハザードモデルを中心に、あまり数式をつかわず、考え方と応用を解説したもの。

⑥丹後俊郎：メタアナリシス入門。朝倉書店
メタ分析の種々の方法を解説したもの。

引用文献

1. Shepherd J, Cobbe SM, Ford I, Isles CG, Lorimer AR, MacFarlane PW, McKillop JH, Packard CJ (1995) Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. West of Scotland Coronary Prevention Study Group. N Engl J Med 333: 1301-1307
2. WHO fracture risk assessment tool. <http://www.shef.ac.uk/FRAX/tool.jsp?country=3>