

Dead Heat

If you're walking down the right path and you're willing to keep walking, eventually you'll make progress.

--- Barack Obama

QUESTION

「統計学的有意差と臨床的意義の違いは何でしょう？」

THEORY

臨床試験の定義

「臨床試験とは予防や治療を評価する目的で人を対象に行なう科学的実験研究のことである。」

例えば車でオイルチェンジを 3000km 毎に行なうのと 5000km 毎に行なうのとどちらが良いかをみるとする。これも臨床試験に似てはいるが人を対象としていないので違う。車と違って人では個人差が大きい。臨床試験はこの個人差を乗り越えて、いかにノイズを減らすか工夫する。以下他の教科書の定義をみてみよう。

1. A clinical trial is a scientific research activity in human subjects undertaken to determine prospectively the effect and value of preventive, diagnostic, and therapeutic agents, devices, regimens, and procedures.

Hopwood MD, Mabry JC, Sibley WL; R-2635-NIH Rand Corporation, 1980

Prospectively とは、結果がでる前に研究を始めることを意味する。

2. A clinical trial is any form of planned experiment which involves patients and is designated to elucidate the most appropriate treatment of future patients... retrospective surveys are unplanned observational studies and hence do not fit the definition.

Pocock SJ. Clinical Trials: A practical approach. Wiley, 1983

the most appropriate treatment of future patients とは、その臨床試験によって、将来

の患者さん達が恩恵にあずかることを指している。臨床研究の意義はこの点にあるのではないだろうか？我々はサリドマイドで苦い経験をした。新しい治療は科学的に検証されてベネフィットがリスクに優ることが確認されてはじめて使われるようになるべきである。

3. A scientific experiment that generates clinical data for the purpose of evaluating one or more therapies on a patient population. Implicit in the trial is the fact that the clinical investigator has control of the process by which a treatment is assigned to a patient.

Gelber RD, Zelen M: Planning and reporting of clinical trials. In: Calabresi P (ed). Medical Oncology. Macmillan, 1985.

investigator has control とは、患者さんの治療を治療群にするか対象(コントロール群)にするか決定するのは研究者であって患者さんではないことを指している。他の臨床研究にはない特徴だ。臨床試験は、介入研究：intervention study とも呼ばれる。

倫理的問題

臨床試験は未来患者さんのために少しでも役立つように行なわれるわけだが、あくまで一種の人体実験であり、倫理的問題を避けて通るわけにはいかない。Biomedical Research の目的は 1964 年ヘルシンキ宣言の序文に規定されている。

The purpose of biomedical research involving human subjects must be to improve diagnostic, therapeutic, and prophylactic procedures and the understanding of the aetiology and pathogenesis of disease. (Declaration of Helsinki. Intro)

1. 治療は何らかの利益だけではなく副作用などのリスクをも伴う。よって有効性(ベネフィット)より危険性(リスク)の方が高いことが予測される臨床的特徴をもつ患者さんを除外する必要がある。

- a. 既に他の薬剤投与により被験者に腎機能障害があり、これから行なおうとしている治療薬の少なくとも1つが腎機能をさらに悪化させる可能性がある。
 - b. もしも薬剤が胎児に対して悪影響を及ぼしたり、結果が予測できない場合、妊娠中（および授乳中の）の女性は除外される。
 - c. 例えば 100%ベッド上に居る癌患者さん、うっ血性心不全の患者さんなど、状態が重症であり、治療効果が判る前に亡くなる可能性があるような場合は、しばしば臨床試験から除外される。
 - d. もしも薬剤による毒性がありえる場合で、患者が観察のみで治癒しえるような場合、例えば *in situ cancer* は *cancer trial* から除外される。
 - e. 理論的、経験的に有効性が期待できない状態にある患者さんに薬剤を投与する、例えば脳転移のある癌患者さん、ホルモン療法に無反応な乳癌に対する卵巣摘出術を施行するなど。
2. 以下のような特殊な亜集団を除外することも妥当とされている。
- a. クレアチニンクリアランスの低い患者さんを除外すること。
 - b. 高齢者に対する毒性が強い場合には、高齢者を除外できる。
 - c. 特に理由がなくても最小年齢を設定する。
 - d. 小児臨床試験において年齢制限を設ける。
 - e. 黒人しか入院していない南アフリカの病院において黒人を対象とした臨床試験を行なう。逆も真なり。

一方以下のような場合は許されない。

- a. 少年院のみから参加者を募る事。仮にその集団においてある病態、肝炎や薬物依存など、が多いからといっても許されない。
- b. 性行動に変化を与えるかどうかの試験で、あるマイノリティー集団を除外すること。

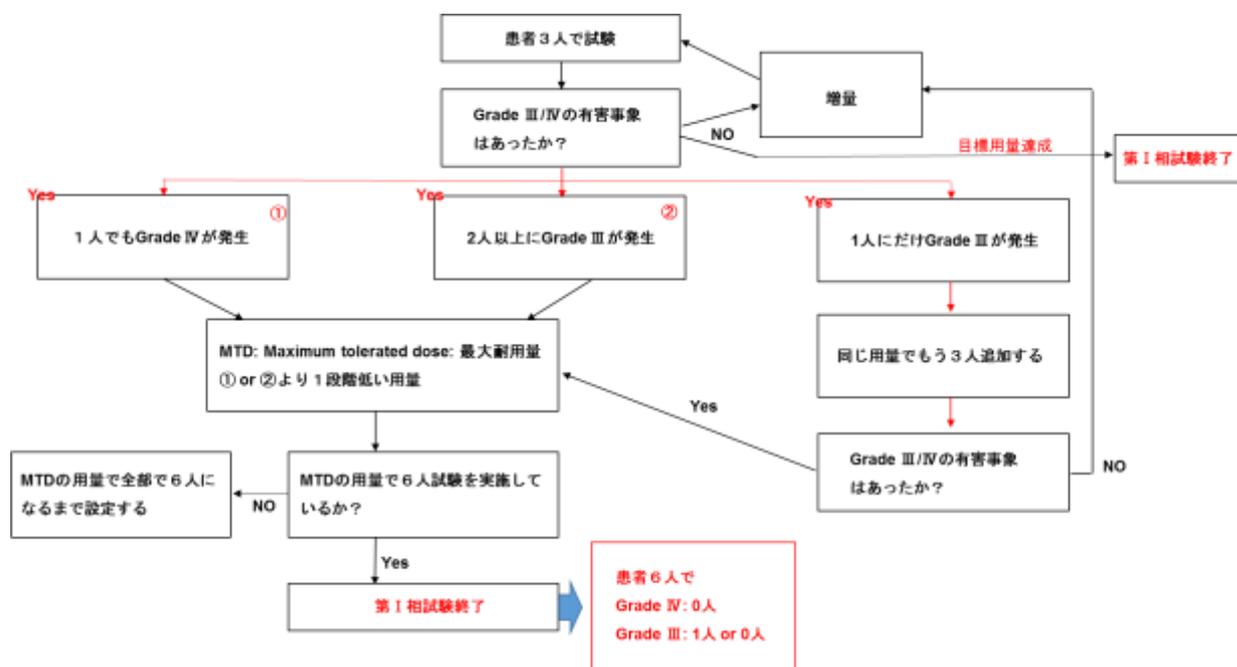
臨床試験の目的

製薬会社は数千以上ある物質の中から動物実験などを経て、ある疾患に効きそうなものをスクリーニングする。そして安全性を確認した上で、臨床試験に移行する。臨床試験は以下の4つの段階に分けられる。

第 I 相試験(Phase I):

動物実験が終わり、当局が臨床試験に移行することを許可したら、**phase I clinical trial** に入る。この相では、可能性のある副作用を探り、薬剤治験であれば最大耐用量 (**maximum tolerated dose: MTD**)を決める。デザインとしては、初期用量、1回増量それぞれの用量で何人検査するか、増量の基準、中止の基準、**MTD** の定義が重要となる。他の目的で同薬剤が人で使用されていれば、それも多いに参考になるが、無い場合には動物に投与して **10%**が死に至る用量の更に **1/10** から開始することもある。**MTD** を決定するための毒性レベルを予め決めておき、その毒性がみられるまで、あるいは目標値徐々に増量していく。1用量あたり **6**人、各レベル **3**人、全体で **6**人から **30**人程度が対象となる。通常は **3**人ずつ同じ用量で検討し、副作用が発現すると考えられる期間、十分観察した後、次のステップに進む。**3**人の協力者（がんなどでは患者さんが対象）に対して低用量から開始。**Grade III/IV**の毒性が発生しなければ、用量を増やす。**Grade III**の毒性が**3**人中**1**人にみられたら同量で別の対象**3**人に試みる。**Grade III**

/IVの毒性が発生しなければ、用量を増やす。Grade IIIの毒性が3人中2人、GradeIVの毒性が3人中少なくとも1人にみられたら、その用量より1段少ない用量を MTD とする。MTD の用量で既に6人で試験を行っていれば終了。行っていなければ、6人になるまで行う。3+3 design と呼ばれている。



第II相試験(Phase II):第II相試験は第III相試験に進むのに十分な効果があるか否かを見極めるために行われる。そのため、対象となる患者さんはその新薬が最も効きそうな人で構成される。倫理的に許される状況であれば、パフォーマンス・ステータスが良く、前治療も少ない患者さんが選ばれる。治療効果が判定できるように、癌であれば RECIST など測定可能な腫瘍が存在することが望まれる。

Phase II を early と late に分け、後者ではランダム化が行われることがある。この場合、Phase III のランダム化は本格的な比較目的なのに対して、late phase II では、

大人数をランダム化する意義があるかを見極めるために行われる。

第Ⅲ相試験(Phase III): 従来の治療あるいは placebo と新しい治療を比較する。

第Ⅳ相試験(Phase IV): 市販後調査であり、多くの人々に対する長期副作用、長期効果について検討する。

EXAMPLE AND EXERCISE

Example 1: Rociletinib by Clovis Oncology ⁽¹⁾

非小細胞肺癌の中でも上皮成長因子受容体(EGFR)に変異、特に exon 19 あるいは exon 21 における L858R に変異を持つ場合、EGFR 阻害薬 である gefitinib, erlotinib, afatinib が 50~70%の有効性を示す。しかし EGFR 阻害薬を使い続けると 1 年もしないうちに半数以上の患者で T790M 変異を来し、薬剤耐性を示す。この変異を示すと患者さんの余命は2年を切るようになる。最近開発された Rociletinibは exon 19 deletions, L858R, T790M に変異をもつ肺癌を縮小させる働きがあることが前臨床研究で明らかになっている。Rociletinib は従来の EGFR 阻害薬とは異なり、正常の EGFR に対する効果は少ないため、発疹や下痢といった副作用も少ないと思われる。また動物実験より 200ng/mL を維持すると抗腫瘍効果を発揮することも確認されている。そこで、既存の EGFR 阻害薬を投与しても腫瘍が増大してきた、且つ EGFR に変異をもつ進行非小細胞肺癌患者さんを対象に Rociletinib の第 I 相および第 II 相臨床試験を実施することにした。

1 日 150mg から開始し、900mg x 2/day まで増量したが、MTD を決めるには至らなかった。しかし、高血糖が高頻度でみられている。

Phase II: 73 人

T790M に遺伝子変異を持つ (46 人) : 有効率 59% (95%CI: 45 ~ 73%)。

T790M に遺伝子変異を持たない場合 (17 人) : 有効率 29% (95%CI: 8 ~ 51%)。

Example 2: AZD9291 by AstraZeneca ⁽²⁾

2015 年 4 月 30 日発行の New England Journal of Medicine に別の EGFR 阻害薬 AZD9291 の Phase I・II の結果が発表されている。

Phase I : 31 人

1 日 20mg から開始し、240mg /day まで増量したが、MTD を決めるには至らなかった。Grade III 以上では比較的下痢が多くみられている。

Phase II : 222 人

T790M に遺伝子変異を持つ (138 人中評価可能 127 人) : 有効率 61% (95%CI: 52 ~ 70%)。

Progression-free survival: 9.6 か月 (95%CI: 8.3 ~)

T790M に遺伝子変異を持たない場合 (62 人中評価可能 61 人) : 有効率 21% (95%CI: 12 ~ 34%)。Progression-free survival: 2.8 か月 (95%CI: 2.1 ~ 4.3)

もちろん二重盲検ランダム試験を実施しないと、この 2 つの新しい EGFR 阻害薬の優劣は不明であるが、数値上、ほとんど同等に見える。

Exercise 1

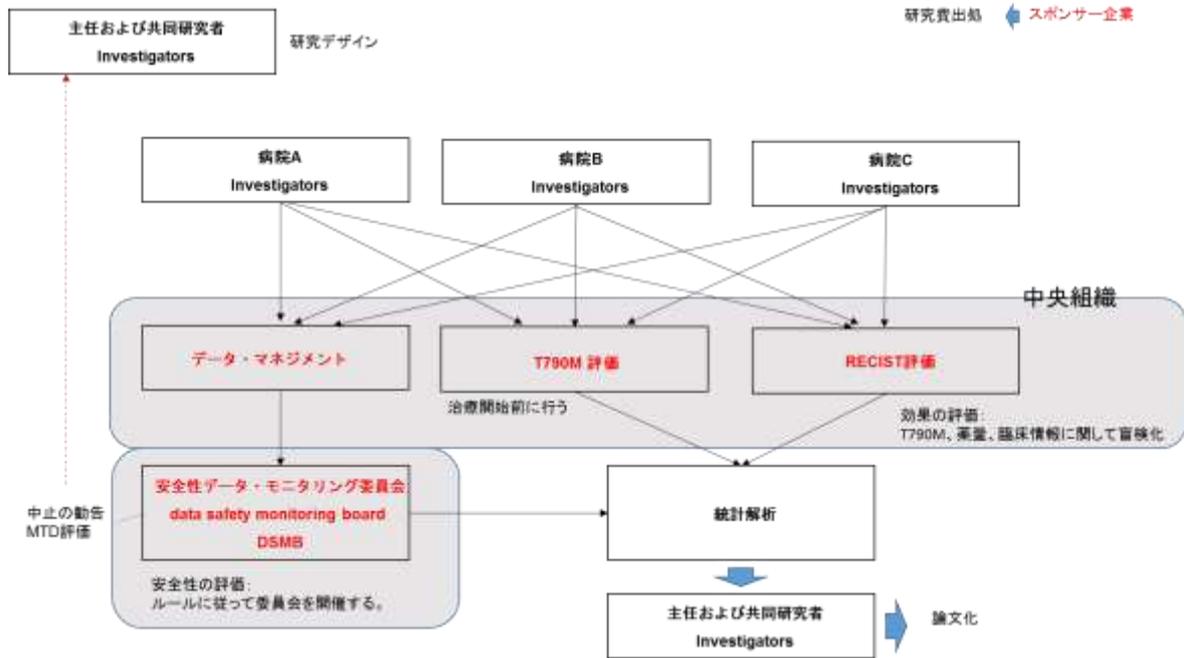
第 I 相試験で薬物を 10mg, 20mg, 30mg, 40mg, 50mg と増やしたとき、grade III 以上の有害事象が発生しないと仮定したとき、何人のエントリーが必要か？ MTD は何 mg か？

10mg: 3 人, 20mg: 3 人, 30mg: 3 人, 40mg: 3 人, 50mg: 6 人 ⇒ 合計 : 36 人

第 I 相試験ではだいたいこの程度の人数になることが多い。

Exercise 2

第II相試験で臨床試験を組むとき、どのような組織が理想的だと思うか？図示せよ。

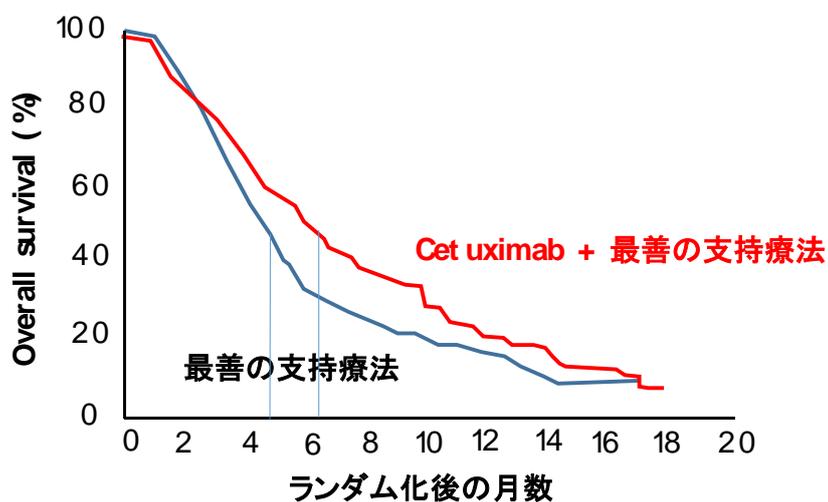


データマネジメントは臨床データなどを収集管理する。有害事象発生時は、DSMBにはかる。DSMBは想定外の有害事象が続くなどした場合には、主任研究者ないしは倫理委員会に試験の中止を勧告できる。T790Mの遺伝子解析は治療開始前に行うことが理想的であるが、治療後になってしまった場合には、RECISTの結果などは遺伝子解析者には伏せておく。RECIST評価者には画像のみの判断だけを依頼し、T790Mの有無、薬量、臨床情報に関しては盲検化する（知らせない）。全てのデータが固定された段階で、統計解析に回す。統計解析者と研究者らは共同で論文をまとめる。スポンサーは研究実施のための経済的支援を行い、試験に関する全てのプロセスに関与しない。赤字は研究デザイン、統計解析、論文執筆に加わらない人たち、黒字は試験の継続、

MTD の決定、T790M の評価、RECIST に従っての腫瘍の大きさの変化に関する評価には関与しない人たち、に分けてみてはどうだろう。スポンサー企業にも研究者にもよい結果を発表したいという意識が働いているはずである。これに対して、データマネジメント、T790M 評価、RECIST 評価、DSMB の委員は論文著者にはならないルールにしておけば、結果の評価にバイアスは入りにくいだろう。

Example 3. Cetuximab ⁽³⁾

Cetuximab は癌の成長因子である EGFR を阻害する最近開発された分子標的薬である。抗がん剤や放射線治療を試みたが十分な反応がなく、がん細胞が EGFR を強く発現している進行大腸癌患者 572 人を対象に Cetuximab 投与+最善の支持療法を施行する群と最善の支持療法のみで、生存率に差があるか否かを検討した。



Cetuximab 投与+最善の支持療法の患者さんの生存期間の中央値は 6.1 カ月であり、一方、最善の支持療法のみで患者さんの生存期間の中央値は 4.6 カ月であった。その差僅か 1.5 カ月である。しかし統計学的には P 値が 0.005 であり、「最善の支持療法のみではなく、これに Cetuximab を毎週点滴することにより有意に生存期間が

SUMMARY

第Ⅰ相試験とは、動物実験などで効果のあった薬物を人に投与して安全性や薬物動態などをみる試験である。少ない量から徐々に増量していき、最大耐用量を決めることが目的である。副作用が出にくい薬物では健康な人に投与されるが、抗癌剤など強い薬に関しては患者さんに投与する。

第Ⅱ相試験とは、第Ⅰ相試験で決めた用量で患者さんに投与し、効果を評価する。従来の治療より大きな効果が期待できそうであれば、第Ⅲ相試験につなげる。

MY THOUGHTS

分子標的薬開発競争の時代

従来の抗癌剤は、いくつかのメカニズムはあるものの単に細胞分裂を抑制するものが使われていた。癌細胞だけではなく正常細胞にも作用するため、大量に使うと、脱毛や骨髄抑制、吐き気、倦怠感が必発であった。しかし、癌は特殊な遺伝子変異をもっていることが多く、ここを分子標的とすることにより、正常細胞への影響、ひいては副作用を少なくすることが期待される。

この同時に *New England Journal of Medicine* に発表された臨床試験結果は、Clovis Oncology と AstraZeneca による開発競争に見える。どちらが先に着手したかは不明であるが、同じ着想で薬物を開発し、動物実験で、臨床で使えそうだという確信を得たに違いない。そこからは競争である。先に販売した方が市場の優位性を築けるからだ。そのような視点からこの論文をみてみたい。

スポンサーとの関係

Clovis Oncology が Rociletinib を開発し、この試験も支援している。一方、AZD9291 は AstraZeneca により開発され、同社により支援されている。Rociletinib 試験の論文共著者に 5 人の Clovis Oncology 社員が、AZD9291 試験でも同じく 5 人の AstraZeneca の社員が含まれていた。Rociletinib 試験の第一著者はマサチューセッツ総合病院、

AZD9291 試験のそれはダナ・ファーバー癌研究所の所属であり、両方ともボストンにあるハーバードの関連病院だ。この 2 人もライバルであったかもしれない。

企業と研究者が共同で研究をデザイン。企業がデータの回収。統計解析は企業と研究者の両者。研究費に政府の資金は入っていない。data and safety monitoring board (DSMB)を設置したという記載がない。しかし著者全員がデータにアクセスできたと、あるいはデータの正確性を確認したとしている。AZD9291 試験では第一著者が、Rociletinib 試験では第一著者とスポンサー社員が論文の最初の原稿を書いたとなっている。

研究組織についての疑問

大概の臨床試験では methods – statistics の段落に対象数計算が示されているが、この 2 つの試験においては記載がない。AZD9291 試験では「腫瘍の大きさ評価は研究者と independent central review が行った」と記されているが、「腫瘍の縮小を評価した者は薬剤量について盲検化した」とは書いていない。本来研究者とは別に（利害関係のない）複数の専門家よりなるエンドポイント評価委員会を構成し、T790M 遺伝子変異の有無や薬剤投与量などの情報は完全に伏せた形で腫瘍が縮小したのか、変わらないのか、増大したのかを評価してもらうべきである。AZD9291 試験において T790M 遺伝子変異ありの 138 人中評価可能であった 127 人について腫瘍縮小効果、患者さんの予後について評価されているが、11 人はどのような理由で解析に加えなかったのだろうか？

2 つの臨床試験の意義

T790M 変異を来した非小細胞肺癌患者さんに比較的安全に使える、腫瘍縮小について有効であることが示された。しかし、T790M 遺伝子変異が検出されてからの生存期間中央値が 2 年未満であり、この新規 2 剤が腫瘍を多少縮小させるだけでなく、本当に患者さんの余命を伸ばすことができるかは Phase III の結果を待たなくてはならな

い。

最近、この分子標的薬の第 III 相試験で、プラセボに比べ数か月 **progression-free survival** を、統計学的有意差をもって延長したとする論文が目立つ。Example 3 に至っては僅か 1.5 カ月の延命である。しかし研究者らは P が 0.005 で統計学的に有意であることを主張する。統計学的に有意かもしれないが、臨床的にはいかがなものだろうか？ 加えて、月に 80 万円かかる治療薬（ほとんどは国のお金で支払われ患者さんはあまり実感していない）で副作用もあり、病院には通い続けなくてはならず、はたしてそのような治療が本当に患者さんや社会のためになっているのだろうか？ 医療者側の自己満足であったり、製薬会社の利益のためになっていたりはないだろうか？

ANSWER TO THE QUESTION

統計学的に有意であったからといって、臨床的に意義があるとは限らない。前者では $P < 0.05$ など一定のカットオフがあるが、後者についてはそう簡単にラインを引けるものではない。

REFERENCES

1. Sequist LV, et al. Rociletinib in EGFR-mutated non-small-cell lung cancer. N Engl J Med. 2015;372:1700-9.
2. Janne PA, et al. AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. N Engl J Med. 2015;372:1689-99.
3. Jonker DJ, et al. Cetuximab for the treatment of colorectal cancer. N Engl J Med. 2007;357:2040-8.