フォーラム

# 根治療法は両刃の剣か？：患者・治療間の交互作用

医学統計解析グループ（代表：前谷俊三）[a]，小野寺　久[b]

[a] 天理よろづ相談所　医学研究所，天理よろづ相談所病院
[b] 京都大学大学院腫瘍外科

**背景：**二つの治療を比較するランダム化比較試験（RCT）において通常立てられる仮説は「一方の治療が他方より有効」かまたは「両方の治療効果に差がない」というものである．しかし臨床においては「ある患者では治療AがBより有効であるが，他の患者ではAがBに比べて効果が劣るか，むしろ有害である」という場合がある．これは患者―治療間の質的交互作用と呼ばれる．その2例を提示して現行のRCT（並行比較デザイン）の問題点を提起する．

**方法と結果：**第1例．直腸癌に対して腹会陰式直腸切断手術を行った後，会陰・骨盤腔に再発したため我々が1978年から1997年の間に骨盤臓器と骨盤壁を含めて根治的に一括切除を施行した36例の長期成績を後顧的に調査した．5年生存は10名，10年生存は6名と保存療法では達成できない長期生存が得られた．しかし他方では再々発例において血清CEA倍増時間が術後有意に短縮し，根治手術が癌増殖を促進し死期を速めたと考えられた．第2例．Emory Angioplasty versus Surgery Trialの報告によれば，冠状動脈多枝病変を有し適格条件を満たす842名の中で392名は経皮経管腔的冠状動脈形成術（PTCA）または冠状動脈バイパス吻合術（CABG）にランダムに割り付けられた．残りの450名はRCTへの参加を拒否したが，同じ処置のいずれかが主として医師の選択に基づき行われた．RCT参加者の中ではPTCAを受けた患者とCABGを受けた患者の成績には差がみられなかったが，RCT非参加者の生存率はRCT参加者に比べて有意に高かった（96.4% vs 93.4%）．RCT非参加者の中いずれの処置を受けた患者も，同じ処置を受けたRCT参加患者と比べて生存率が劣ることはなかった．

**結論：**現行のRCTで治療法の比較を行う場合，全般的によい治療法がどれかを決めることはできる．しかしもし患者・治療間の交互作用があれば，「よい治療」を選んだため反って損失を蒙る患者がどれだけいるかも，またどのような患者が損をするかも推定できない．その結果個々の患者にとって最適な治療選択を逸するかもしれない．医療資源と患者の利益の観点から現行のRCTを再評価する必要がある．

【別刷請求先】
〒632-8552 天理市三島町200
天理よろづ相談所　医学研究所
前谷俊三

**キーワード：**ランダム化比較試験（RCT），患者・治療間交互作用，直腸癌局所再発，冠状動脈疾患，仮説検定

## はじめに

放置すれば明らかに予後不良である進行癌に対して根治手術を行うことにより，全治または長期延命を達成できたと実感できるときがある．しかしその逆の経験も稀ではない．根治手術により腫瘍の増殖を加速し，患者の余命を短縮したのではないかという印象を拭い切れない例もある．この「根治手術は両刃の剣である」という考えを確かな「エビデンス」によって立証することは難しい．というのは，いわゆる証拠に基づく医療（evidence-based medicine, EBM）[1,2]において，強いエビデンスとは何かといえば，それは実験室で得られた遺伝子変異やサイトカインの異常など，疾患の原因や治療のメカニズムの研究から得られた知見ではない．患者自身が重視し，かつ実感できるアウトカム，つまり「死亡」や「再発」などが何時起きたか起きないかを評価基準として，根治療法によって達成したアウトカムが偶然には滅多と起きるものではないことを統計学的に証明しなければならない．このためには治療法以外の条件は同一にしてバイアスを排除する必要がある．そこで，できるだけ似通った患者同士で根治療法と保存療法のアウトカムを比較することが望ましい．もっと理想をいえば，各患者でどちらか一方の治療を行なって患者が死ぬまで追跡し，死後タイムマシンで患者を治療前の時点に引き戻し，今度は他方の治療を行なって，どちらでより長く生きたかを比較することが最も望ましい．勿論，実際には一人の患者ではどちらか一方の治療しかできないのが普通である．特に外科手術では左右に行う手術を除けば，二つの術式は別の患者でしか比較できない（並行比較デザイン[3]）．

本稿ではこのような制約を認めた上で，現実に可能な方法で，根治療法は両刃の剣である可能性をまず検証する．その理由は，この認識は根治療法の限界と有害性を自覚する上でも，また根治療法の真価を知る上でも重要だからである．もう一つの理由は，これは外科治療だけの問題ではなく，現在広く行われているランダム化臨床試験（randomized controlled trial, RCT）と深くかかわる問題であるからである．「両刃の剣」論をもっと一般化すれば，「ある患者では治療Aが治療Bに勝るが，別の患者では治療Bよりも劣る」ということになり，統計学ではこれを「患者と治療間の質的交互作用」という．臨床医はこの交互作用という言葉に馴染みがなくても，これが実際に起こることはむしろ当然と考えている．しかし一般に現行のRCTは暗黙の中にこの質的交互作用が起きないと想定している．[4] 例えば仮説検定では「治療A, Bの効果に差があるかないか」という仮説を検証しようとしている．しかし「どのような患者にAがよく，どのような患者にはBがよいか」という患者や医師が知りたい疑問には答えていない．用意している答えは高々，「A（B）が効く」か，または「差がない」という大雑把な解答である．この意味ではRCTにのみ基づいた診療をすれば，それは十把ひとからげの医療となり，個別の患者に対応した医療（tailor-made medicine）の対極にあるといえる．もし患者治療間の質的交互作用が起こり得るとすれば，現行のRCTは重大な弱点を抱え，仮説その他の見直しが必要かもしれない．以下質的交互作用の具体例を示し，それとRCTとの関りを論じる．

**直腸癌の会陰骨盤腔再発に対する拡大合併切除術**

**1．根治手術の対象**

1978年より我々は本疾患に対して，拡大根治手術を行ってきた.[5,6] これは再発病巣のみならず，隣接する骨盤臓器や骨盤壁の一括合併切除である．本治療が他の保存的治療と比べて本当に長期生存を達成することを立証するために，今回我々は対象を，直腸癌の初回手術で腹会陰式直腸切断術を施行した症例に限った．というのは長期生存の報告の中には，初回治療が放射線療法だけの症例[7]や，腫瘍切除のみを行った後に再発した症例[8]も含めている．これらは「最初に取ろうとすれば取れた筈の腫瘍を取り残したための再発」かもしれず，厳密には再発手術と呼ぶべきでないからである.[9] 同様に初回手術で自然肛門を温存した症例では，吻合部の縫合不全や排便障害を避けたいため，少しでも腸管を残そうとして，本来取れるべき病巣を取り残すおそれがある．こうして生じた再発に比べて，肛門と十分な長さの腸管を含めて直腸を切断した後に再発が起これば予後はもっと不良といわれる.[8,10] このような再発症例の予後を変えることができてこそ，根治手術の有効性がより確かなものとなる．そこで1997年までに京大とその関連病院で行った根治手術59例中上記の条件を満たす36例を対象としてその結果を調べた．

**2．根治手術の長期成績**

図1はそのOverall survivalであり，5年と10年の生存率はそれぞれ28%（95%信頼区間13~42 %）と19%（95%信頼区間6~32 %）であった．実際の5年生存者は10名，10年生存者は6名である．即ち，対象をこのように厳しい条件に絞っても，根治手術で長期生存（または全治）する直腸癌再発患者は一定の割合で存在する．

**3．保存療法の成績**

一方，本疾患に対する保存療法としては，放射線治療が主であり，補助療法として化学療法や温熱療法も含まれる．以下にはこれらを含めて放射線治療と総称する．施設によっては再手術ができないほど進行した症例ばかりが放射線治療に回されることも

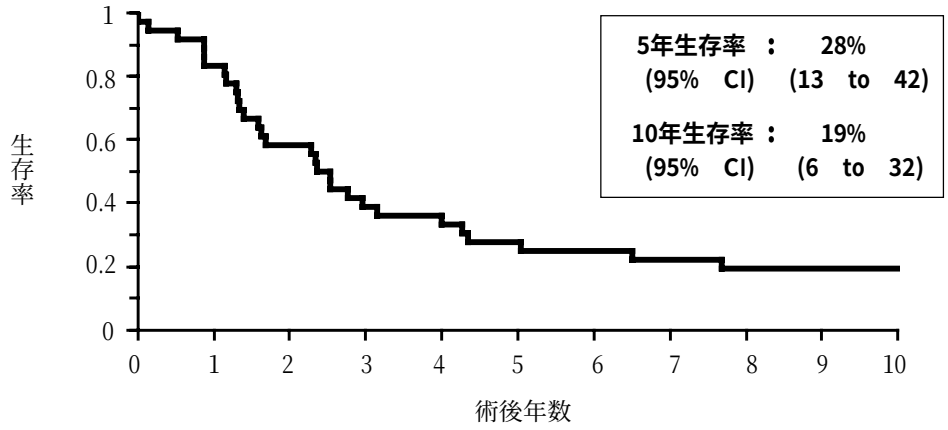


**図1．腹会陰式直腸切断術後の直腸癌局所再発に対する根治手術術の生存曲線**

ある．このようなバイアスのある場合は根治手術と放射線治療とを公正に比較できない．ただ欧米では放射線治療が圧倒的に多いにも拘わらず，長期生存例は僅かに過ぎない.[11] また5年生存率の報告も少なく，0%から高々17%である.[12-14] 10年生存に至っては我々の知る限り報告例はみられない．放射線の効果が不十分である一つの理由は，再発巣が腸管や泌尿生殖器と接するか，これを巻き込んでいるため，線量を上げるとこれらの管腔臓器に閉塞や潰瘍，出血，瘻孔を来すおそれがあるためである．以上からみれば，RCTをやらなくても，長期生存は根治手術に多く，他の治療ではごく少数に過ぎないといえる．

**4．根治手術の別の顔**

しかし根治手術はすべての患者で一律に生存時間を延長するかといえば，決してそうではない．一般論をいえば，根治手術では早期の治療関連死が増加する．我々の症例では入院死亡は2例（術後4日と48日）と手術侵襲の大きさに比べて少なかった．しかし早期死亡は治療関連死だけではなく，原病死も多い．術後1年内死亡は6例，2年内死亡は15例に達する．我々は根治手術を契機として病巣が急速に増大する症例が少なからずあるという印象をもっていた.[5,6] それを裏付ける客観的なデータが血清CEA値であり，これはほとんどの患者で正常値を超えていた．

さて患者の余命を決める因子を調べると興味ある結果が得られた．「再発病巣がどこまで広がっているか」，例えば「肝に転移しているかどうか」も有意な予後因子である．ところが変数選択法で予後因子をふるいにかけると，もっと強力な因子として残るのは，「現時点の病巣の広がり」ではなく，「病巣がどれだけの速さで増大するか」であった.[6] これはCEAの倍増時間で推定できる．即ち，CEA値そのものよりは，CEA上昇の勾配が急なほど余命は短い．CEA倍増時間と患者の余命は対数グラフ上では直

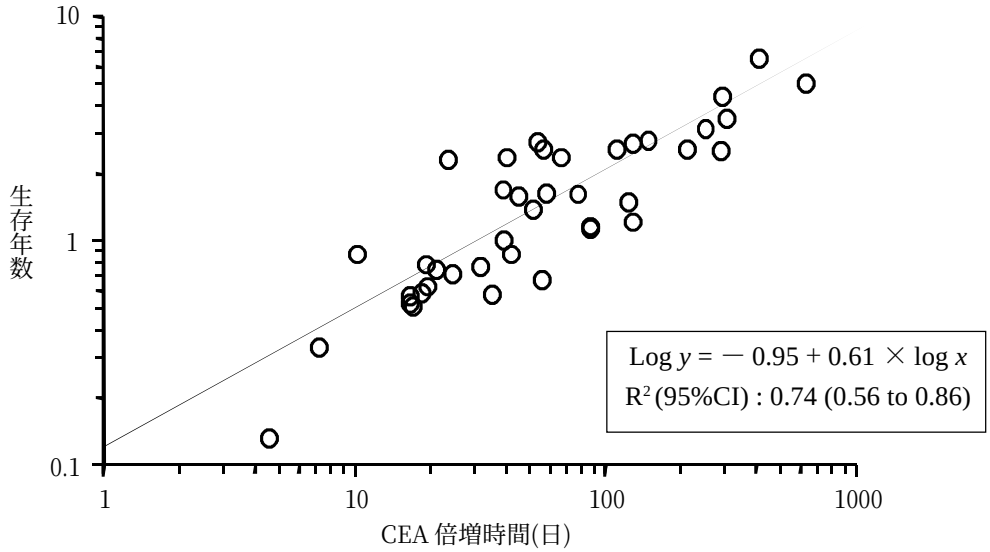


**図2．直腸癌局所再発に対する根治手術の血清CEA値に及ぼす影響**

線的な関係にあり（図2），前者からおよその余命を予測できる．注目すべきことに，早期再発患者では手術を境として，CEA倍増時間が短縮した例が有意に多かった．[6] 図3はその1例であり，その値は術前の92日から術後は36日に短縮した．CEAの変化を辿ると，根治手術直後CEAは正常値まで低下しているが，間もなく急勾配で再上昇し，約5か月で術前予測値（外挿値）を越えている．しかもその勾配はその後も衰えていない．これは術後の腫瘍の増殖は，手術侵襲に伴って遊離したサイトカインやchemical mediatorなど[15]とは別の永続的な機序によって加速したことを示唆している．

現在までに再々発をきたした患者は27名と全体の3/4に達した．この中の3名は5年以後に再発している．また手術創の感染，壊死，または哆開は12例（1/3）にみられた．その原因として術前の放射線治療や，血管，神経，骨組織を含む広範な体壁切除もある．これに加えて局所に再々発した腫瘍がadjuvantとして働き，感染を増悪，遷延させたためと考えられる．更に患者は人工肛門に加えて，回腸導管によるダブルストマの管理を強いられる．これらはいずれも患者のQOLを低下させる重要な因子となる．その結果患者一人一人を検討してみると，根治手術をやらないほうがよかったのではないかと感じる例が全体の1/3はあった．この割合がわれわれの言う$P'$(n=1)[16,17]であり，よいはずの治療を受けても，それが裏目に出る確率である．

### 5．RCTをやったとすれば

一方，もしより確かなエビデンスを得るために多施設共同でRCTを行い，根治療法と保存療法を比較したとすればその結果はどうだろうか．保存療法では長期生存例がほとんどないので，もし10年生存率を評価基準（エンドポイント）にすれば，我々の症例数と同程度のサンプルサイズで「根治手術が有意に生存率を上げる」という結果が得られると思われる．もし5年生存率を

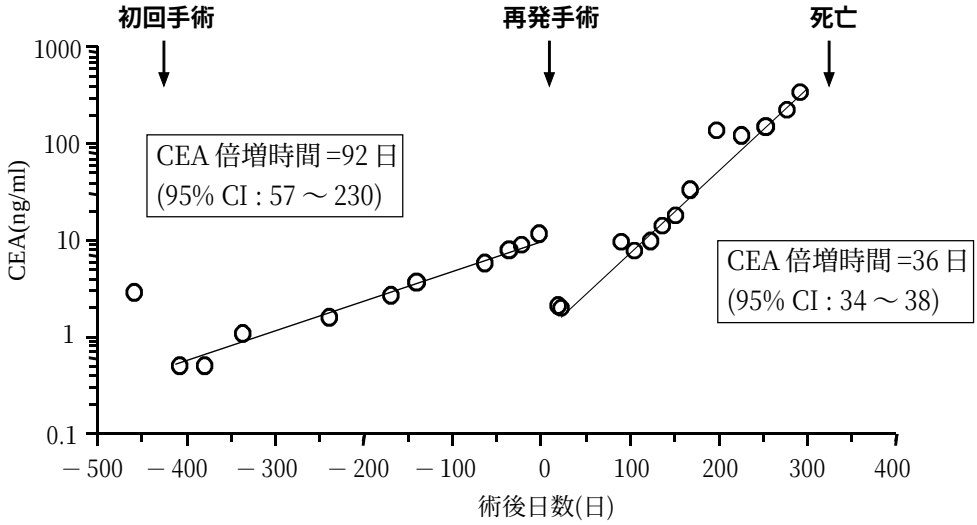


**図3．直腸癌局所再発に対する根治手術後の生存期間とCEA倍増時間の相関**

評価基準とすれば，もっと多数の症例数を要するかもしれないが，やはり有意差が出ることは確かであろう．しかもサンプルサイズを増やせば増やすほど生存率の差は高度に有意となり，[17] 強い「エビデンス」が得られる．もし同じ追跡期間のデータにハザード比やログランク検定を適用すれば，生存率に比べて有意性が得られにくい．[18] いずれを使用するにせよ2群の生存曲線が途中で交叉する場合は，追跡期間によって評価は逆転する．これを避け，一つの値で評価しようとすれば平均生存期間（生存曲線の下の面積）が望ましい．[19] しかし根治手術群は早期死亡が多くても，飛び抜けて長期の生存例が少数あれば，平均はこれに引かれて大きくなる．従って長期追跡をする限り「根治手術は有意に生存期間を延長する」ことになる．その結果RCTは現実の姿をありのままに反映せず，根治手術を過大評価する傾向がある．もし根治手術をその5年生存率が有意に高いという根拠から，標準手術とするならば，根治手術で被害を蒙る患者が増加するおそれがある．もし *P* 値が0.05に達しない場合，より確かなエビデンスを得るためという名のもとにサンプルサイズを増やしてRCTを続けば，それだけ多くの犠牲者を出すことになる．

### 6．医師は何をすべきか

それでは医師は何をするべきであろうか．第一に根治手術が有効な患者とそれが有害となる患者を選別し，不要な手術を避けることである．例えば術前既にCEA倍増時間が100日を切る症例は長期生存を期待し難く，おそらく根治手術は控えるべきであろう．[6] また，術中出血は手術野を覆い隠して正確な手術操作を阻むだけでなく，輸血を介して患者の免疫力を低下させる．制御工学により低血圧を維持することにより出血量も手術時間も節約できる．[20] 以上は我々が実際に行ったことであるが，そのほか，遺伝子変異などの解析により，何故手術後腫瘍の増殖が加速されるかを究明できれば，早期再発死を減らせるかもしれない．また隠れた病巣を描出する画像診断技術が進歩すれば，もっと的確な切除により，再々発や合併症が減少するかもしれない．さらに手術術式や周術期管理も改善の余地があり，これにより長期生存例は増える可能性がある．RCT以外にも臨床医は症例毎に学ぶべきことは多い．

### 冠状動脈多枝病変に対する血管形成術とバイパス手術のランダム化割付患者と非ランダム化割付患者の比較

もしある患者集団にRCTを行い，治療法A，Bを比較したところ，両者の効果に差がなかったとする．一方，同様の集団で医師が患者毎に本人に適すると思う治療をA，Bの中から選んで割り付けたところ，その結果はRCTよりもよかったとする．この差が本物ならば，事は重大である．治療の選択は最初から医師にまかせたほうがよく，RCTは無駄無用ということになる．実はこれは仮定の話ではなく，似たような結果が実際に報告されている．

### １．対象と治療の割り付けまたは選択

1987年7月から1990年4月までにEmory大学病院など3病院において冠状動脈疾患患者に対して血管形成術（PTCA）またはバイパス手術（CABG）を行った（EAST）．[20,21] 対象は冠状動脈の二枝または三枝病変を有し，これまでいずれの治療も受けたことの

ない842名の患者である．この中でRCTに同意したのは392名で，198名はPTCAに割り付けられ，194名はCABGに割り付けられた．一方，残りの450名は適格条件を満たしていたがRCTに同意せず，治療の選択には主として医師が関与した．その結果PTCAが選ばれたのは168名で，CABGが選択されたのは270名と多かった（残りの12名は内科療法）．

この試験の特徴はRCTで二つの治療を比較しただけでなく，同時にRCTに参加した患者と参加しなかった患者を比較した点にある．言い換えればランダムに治療を割り付けられた患者と医師が治療を選択した患者との比較を兼ねた注目すべき研究である．

**2．基礎因子の比較**

RCTに参加した患者の中でPTCA群とCABG群を比べると基礎因子に有意な差がみられなかった．

一方RCTに参加した患者と参加しなかった患者の基礎因子を比較すると，年齢，性別，血管病変の数，ejection fraction，糖尿病，高血圧，喫煙，総コレステロールなどで有意差がみられず，類似点が多かった．有意差がみられたのは，RCTに参加した患者にはクラスIII - IVの狭心症と，前行降枝の近位病変が多く，また薬物治療としてヘパリン静注，カルシウム拮抗剤，硝酸化合物の局所貼付が多かった．一方，RCTに参加しなかった患者には70%以上狭窄の血管数，完全閉塞の血管数，及び大学卒業生が多かった．これを治療別に比較しても両者の基礎的特性はよく一致した．

ところが血管造影所見でみると大きい偏りがみられた．医師がPTCAを選んだ患者は二枝一病変が多く，CABGは三枝多病変患者に多かった．その結果二枝が侵されている患者の54%はPTCAを受け，三枝が侵されている患者の84%はCABGを受けた．ちなみに医師がどのような所見のある患者にCABGを選ぶ傾向があるかをpropensity score[23,24]からみてみると，三枝病変と左前下行枝の近位病変があれば有意にCABGを選択することがわかった．

**3．アウトカム**

CABG対PTCA

術死は両者共1%で差がなかった．主エンドポイントを3年内の全死亡，またはQ波心筋梗塞，またはタリウムスキャン上での広範虚血病変の三つとすると，その発生率はCABGで27.3%，PTCAで28.8%と殆ど変わらなかった（$P$=0.81）．3年死亡率も6.2%と7.1%と似通った値であった($P$=0.73%)．ただし再度の血行再建術を要した患者はCABG後で13%，PTCA後で54%と後者で有意に多かった（$P$<0.001）．狭心症も後者に多くみられた．追跡期間を8年延長しても両者の生存率には有意な差がみられなかった．[25]

**4．RCT参加者対非参加者**

入院中死亡はRCTへの参加，不参加に拘わらずほぼ同じであった．3年間の追跡では医師が治療を選んだ方がランダムに治療を割り付けたよりも生存率は有意に高かった（96.4% vs 93.4%, $P$=0.044）．これを治療法別に比較すると有意差はなかったが，CABGにおいても，PTCAにおいても医師が治療を選んだほうの生存率が高かった．またその中で医師がPTCAを選んだ患者のほうが，RCTでPTCAに割り付けられた患者よりもPTCAによる再治療が少なかった．一方，Coxの層別多変量解析で基礎因

子を調整した場合も．医師が治療を選択したほうが有意に相対死亡率（ハザード比）が低かった．同様の有意差は医師がCABGを選んだ患者とRCTでCABGに割り付けられた患者の比較においても認められた．

**５．結果の解釈**

もし二つの治療をRCTで比較した結果，その効果に差がなければ，医師が同様の患者に自分で選んだ治療をしてもRCTの結果と変わらないと考えるのが普通である．まして両者の成績を治療別に比べた場合，いずれの治療においても，医師の選択のほうがRCTよりよい成績を挙げるとは考えにくい．もしそれが現実に起これば，その主な理由は「医師が治療を決めた患者の方が，RCTに参加した患者より予後がもともとよかったため」とみなされる．本臨床試験においてもその可能性は否定できない．例えば不安定狭心症や左前下行枝近位病変はRCT参加者に多く，大学卒は非参加者に多かった．ただこの試験ではRCTの参加者も非参加者も．共にRCT参加のための適格条件を満たした患者であり，違いよりは類似点が多いことは著者も認めている．この違いに比べると，医師が決めた治療群間の違いのほうが遥かに大きい．即ち，二枝病変患者にはPTCAを選び，それが難しいと思われる三枝病変患者はCABGに回したと推測している．事実，医師がPTCAと決めた患者ではPTCAを分割して行った例はRCT参加者の半分に過ぎない．

以上からみれば本試験の対象患者と治療法の間には交互作用があり，それがRCT参加と不参加との間で成績に差をもたらした重要な理由と考えるのが最も考えやすい．つまりCABGとPTCAにはそれぞれ得意と不得意の病変がある．それを医師が薄々気付き，意識的または無意識のうちに治療を使い分けたため医師のほうがよい成績を出したと考えるのが自然のように思われる．事実，propensity scoreから医師がどのようにCABGとPTCAを使い分けたかをみると，三枝病変と左前下行枝の近位病変があれば，CABGを選ぶ傾向があることは既に述べた．ところが追跡調査を8年に延長した結果の報告によれば，有意差こそ得られなかったが，CABGの生存率が時と共にPTCAを上回る傾向を示したのは，上記の二病変をもつ患者と糖尿病治療患者であった．[25] それ以外にも医師が患者と接して得る情報には，言葉や数字で客観的に表せないが治療の選択に重要なものがあったかもしれない．勿論医師が見逃した交互作用もあると思われる．例えば同様のRCTを行ったBARIグループはCABGがPTCAに有意に勝るのは糖尿病治療患者であることを報告している．[26]

意外なことに著者は最初の報告の結論で次のように述べている．「我々の研究で外科手術と血管形成術の間に差がないことが明らかになった．（中略）故に治療選択においては患者の希望を大いに尊重すべきである」．[21] これは医師が現行のRCTを正しく理解していないことの表れとみなされる．「外科手術と血管形成術の間に差がない」といえるのはあくまでも集団と集団の間で平均的結果を比較した場合に限られる．個人でみればCABGの方がよい患者も，PTCAの方がよい患者もいる．どちらを選ぶかを決めるためには医師の意見を尊重すべきことはその後の研究で示唆されている．

本研究の弱点を挙げるとすれば，RCTに参加せず医師が治療を選んだ患者の中の12

名は内科治療に回されている．その基礎所見やアウトカムが記載されていないのは片手落ちである．

## 考察

治療と患者間の交互作用の具体例を二つ挙げ，患者や医師がRCTから求めるものと，RCTが与えるものの間に乖離があることを示した．しかし臨床医学と統計学の間の行き違いはこれだけではない．我々は治療法の比較に仮説検定を導入したときから既に微妙な齟齬は始まっていたと考える．

### 1．統計学的仮説と臨床診断のための仮説

統計学的仮説といえば臨床家は難解な理論として敬遠しがちである．しかしこれは臨床診断で立てる仮説と本質的に変わらない．要は「自分が何を知りたいか」を仮説として明確に表すことである．例えば東南アジアの出張から帰国したばかりの男性が38℃を超える熱発と下痢を来たして来院したとする．医師は最初の問診や診察から可能性の高い疾患と，見逃してはならない重要な疾患を念頭におき，診断や治療を進めるのが普通である．この念頭に置く疾患が仮説であり，仮説の正しさ（または誤り）をその後の検査や経過で検証する．もし患者が実はSARSであったにも拘わらず，医師が最初の仮説として，細菌性腸炎を想定したとすればどうだろうか．この仮説を確かめるために便培養を行い，抗生物質を投与しながら検査結果を待っていたとする．その間に患者の病状は悪化し，医師や患者の周囲の人々がSARSに感染するおそれがある．この意味では適切な仮説を立てることは極めて重要である．

統計学的仮説も同様である．適切な仮説を立て，それが正しいかどうかをデータから検証する．非専門家はどのような仮説を選ぶかは，統計学者の仕事と思い込んでいるが，それをするのは本来医療を提供する医師か，あるいは医療を受ける患者かもしれない．勿論いかに知りたい疑問でもその答えが出せなければ意味がない．そのためには統計学者との助けが必要となる．

### 2．「A，Bに差があるかないか」という仮説

RCTで治療AとBの効果を比較する場合に通常立てる仮説は「A，Bに差があるか，ないか」である．即ち帰無仮説が「差がない」，対立仮説が「差がある」である．これは我々が最も知りたい疑問であろうか．いやしくもRCTで比較しようとする治療には，その効果に差があるのが普通である．「差がない」という帰無仮説は必要なのだろうか．またそれが否定されて，「差がある」という答えが出たとしても，それだけでは治療はできない．少なくとも医療では「どちらの方がよく効くか」という答えが必要である．それならばこのような回りくどい仮説を立てるよりは，「Aが効くか，Bが効くか」という仮説にしたほうが手っ取り早いのではないだろうか．この二つの仮説検定は似たようにみえるが，本質的な違いがあり，本来使い分けるべきものではいかと考える．これを以下の例で説明する．

### 3．モーツアルトの作品を聞かせたトマトは甘いか

モーツアルトが比類ない作曲家であることはアインシュタインを始め多くの天才も認めるところである．その作品を演奏すると，他の作曲家の作品には反応しないトマトまでが甘くなると言われた．しかしモーツアルトの心酔者でも大多数はこの伝説に

疑問をもち，「トマトが甘かったとすれば，それはたまたま甘味のあるトマトにモーツアルトの曲を聴かせたために過ぎない」と考えるかもしれない．この疑問を証明するため，トマトの苗をランダムに2群に分け，環境の等しい2箇所の畑に植え替えて実験をしたとする．一方にはモーツアルトの曲を，別の群には他の作曲家の音楽を聞かせて育て，トマトを収穫後，その糖分の濃度を測定した．

ここで知りたいのは「トマトが甘いのは偶然の結果か，それともモーツアルトの起こした奇跡のせいか」である．これを仮説検定で確かめるためには，「両群のトマトの糖濃度には差がない」という帰無仮説を立てる．ここで「差がない」というのは「測定値の差が0」という意味ではない．「測定値に多少の差があっても，その差は奇跡で生じたものではなく，個別のトマトの糖濃度にばらつきがある結果だ」という意味である．そこでもしモーツアルトの曲を聞かせたトマトの糖濃度が少し高くても，糖濃度の個体差の範囲内であれば，それは偶然の結果と考える．しかしし万一モーツアルトを聴かせたトマトの糖濃度が自然のばらつきで起こるとは考えられないような異常な値であれば，最初の帰無仮説を取り消して，奇跡（対立仮説）を認めねばならない．

この異常の程度を測る物差しが*P*値であり，*P*値が小さいほど偶然には起こりにくいことが実際に起きていることを示している．言いかえれば*P*値は「まぐれを否定する証拠の強さ」を表すことになる．慣習的には*P*値が0.05より小さければ，まぐれという帰無仮説を取り下げ，対立仮説を認める．ここで医師がよく犯す過ちは「*P*=0.05であれば，実際に観察されたことがまぐれに起きる確率は0.05に過ぎず，残りの0.95はまぐれ以外で起きる確率，つまり超能力のなせるわざである」という誤解である．これは丁度診断において「特異度」を誤解することと似ている．もし医師がある病気を疑い検査をしたところ結果が陽性に出たとする．そこで医師は患者に対して「この検査の特異度は95%（偽陽性率=5%）であるが，その検査でプラスとでたのだから病気である確率は特異度と同じ95%ある」といえば誤りである．病気の確率を知りたければ特異度ではなく陽性予測値[27]を求めなければならない．仮説検定でもこの陽性予測値に相当する確率を求めるのが本筋である．しかし実際にはそれが求められないので，やむを得ず*P*値で代用しなければならない．

### 4．*P*値の問題点

仮説は医師が知りたい疑問である以上，仮説検定においてはその仮説がどれだけ正しいかに答えなければならない．理想的には仮説の正しさは「確率」として表すとわかりやすいが，*P*値を使うと誤解の元になる．例えば「仮説は70%正しい」というところを，[*P*=0.3]といえばそれが何を意味するか理解しにくい．

ところが*P*値にはそれ以外にもいくつかの問題を抱えている．例え*P*値が同じ値（例えば0.05）だからといって証拠の強さが同じとは限らない.[28] 症例数が大きくなれば，わずかの差でも*P*値は小さい数値となり有意という結果が出る．一般には*P*値が小さいほどまぐれの可能性は小さくなるが，大きい（1に近い）からといって，まぐれである（帰無仮説が正しい）とは断言できない．

このため*P*値を使った有意性検定を「捉えどころのない（elusive）概念」とみなす識者は少なくない.[29] つまり*P*値を使う限り，不適切な物差しで仮説の真偽を判定しているという疑念が残るのである.

それはともかく「まぐれか奇跡か」を知りたい場合には「差がないか，あるか」という仮説を立てることは妥当である．それでは最初に帰って治療A，Bを選択する場合この仮説は適切であろうか.

**5．治療効果の差の区間推定**

医師や患者が2治療のいずれかを選択する場合,「効果の差がまぐれかどうか」という問題に強い関心をもっているだろうか．むしろ知りたいことの一つは「治療効果の差は平均してどのくらいの大きさ（effect size）か」であろう．これが治療選択の一つの基準であり，期待値と呼ばれて臨床決定分析に使用されてきた.[30] 例えば「治療により生存期間は有意に延長する」というよりは,「平均して3年延長する」というほうがわかりやすく，決心もつけやすい．1980年代後半から治療効果を比較したい場合には,「差があるかどうか」をみる「仮説検定」よりも「effect sizeの区間推定」をすることが推奨されるようになったのはこのような理由による．これに伴い「*P*値を記載するよりは95%信頼区間を示すべきである」という勧告がトップジャーナルに掲載され現在に至っている.[31-34] 例えば「Aの平均生存期間は*P*=0.03で有意にBより長い」と記載する代わりに,「Aの平均生存期間は，95%信頼区間では1~5 年Bより長い」と記載したほうが，効果の差が実感としてとらえやすい．ただしこれはあくまでも集団の平均の話であり，個人の生存期間と混同してはいけない.

1988年British Heart Journalは「*P*値の終焉？」という論説を掲載している.[34] にも拘わらず*P*値は今も医学誌に生き続けている．その一つの理由は，区間推定自体も医学統計学の根本的な問題を解決したとはいないためであろう.

**6．「Aが効くか，Bが効くか」という仮説**

治療法を選択するためには「A, Bの効果に差があるか，ないか」という仮説はなじまないことは既に述べた．それでは上述の「Aが効くか，Bが効くか」という仮説に切り変えればどうなるだろうか．注意すべきことは，両者では仮説の意味が違っていることである．前者で「効果に差がある」という場合は,「まぐれに生じた差ではなく，本物の差がある」という意味である．ところが後者で「Aが効く」というのは「まぐれであろうとなかろうと実際にAがBより効いた」ことを表す．その理由は臨床では効いたことが重要であり，それがまぐれかどうかは二の次だからである．この仮説ならば，それが肯定されても否定されても，治療に必要な最小の情報は得られる．何故ならば後者の利点として,「この仮説がどの程度正しいか」は*P*値ではなく「確率」で表されるからである．例えば「AがBより効く確率は80%である」と具体的な答えが得られるので，Aを選んだほうがよいことが即座にわかる.

ただここで断っておく必要があるのは,「80%はAが効く」というのは「80%の患者にはAのほうが効き，20%の患者にはBのほうが効く」という意味ではない．これはあくまでも集団と集団の比較，言い換えれば平均と平均を比較した場合のことであ

る．二つの集団の一方には治療Aを行い，別の集団にはBを行って集団の治療成績の平均同士を比較する．この試験を100回行えば，そのうち80回はAがBより効き，20回ではBが効くという意味である.[17] このような平均の比較は，集団を扱う医療行政や保険事業などでは有用であるが，患者や臨床医にとっては必ずしも十分な情報とはいえない．

## ７．臨床研究に課せられた課題（むすび）

日常の診療においては無数の岐路があり，そのいずれを選ぶかによって，患者の生死や余命，QOL，治療日数，或いは医療費が大きく左右される場合は少なくない．その岐路の選択において，今やEBMの基礎となるRCTの果たす役割は重要であり，医療界のみならず一般社会の熱い期待が寄せられている．しかしここに述べたように現行のRCT，特に並行比較デザインにはいくつかの問題点や限界があることを見過ごしてはいけない．例えば治療の選択において，平均的治療効果の差はRCTで推定できる．その結果，全体としてどちらの治療がよいかはわかるが，その治療が返って悪い結果を招く例がどれだけいるかはわからない．これは医師，患者にとって最も懸念すべき問題である.[35] われわれはそのリスクを$P'$(n=1)と呼び，特に注意を喚起した.[16,17] 敢えてその答えを求めようとすれば，タイムマシンに代わる方法を考案し，同じ患者で二つの治療を比較する必要がある．

更に質的交互作用がある場合には「平均的にはよいといわれる治療をすると，どのような患者では反って害になるか」という疑問に答えなければならない．それを知るために一律にRCTを行うと，その規模が大きいほど犠牲者の数も増加する．

今後医師の解きたい疑問は医療の選択肢の数と共に増加の一途を辿る．一方，臨床試験のための費用，労力，時間，被験者数には限りがある．現行RCTが医療資源と患者の利益という観点からみて最善の方法かどうか，もう一度問い直す必要がある．

本論文の一部は，2003年9月26日神戸で開催された1st US-Japan Biostatistics Workshopで発表した．

## 参考文献


1. Evidence-based Medicine Working Group. Evidence-based medicine. JAMA 1992;268:2420-2425.
2. 福井次矢．EBM. 日医 2001; 126: 1178-1180.
3. 折笠秀樹．臨床研究デザイン：医学研究における統計入門．真興交易，東京，1995.
4. Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? Stat Med 1984;3:409-402.
5. Maetani S, Nishikawa T, Iijima Y, et al. Extensive en bloc resection of regionally recurrent rectal carcinoma of the rectum. Cancer 1992; 69: 2876-2883.
6. Maetani S, Onodera H, Nishikawa T, et al. Significance of local recurrence of rectal cancer as a local or disseminated disease. Br J Surg 1998; 85: 521-525.
7. Wanebo HJ, Gaker DL, Whitehill R, et al. Pelvic recurrence of rectal cancer. Ann Surg 1987; 205: 482-495.
8. Hahnloser D, Nelson H, Gunderson LL, et al. Curative potential of multimodality therapy for locally recurrent rectal cancer. Ann Surg 2003; 237:502-508.
9. Hoffman JP, Riley L, Carp NZ, et al. Isolated locally recurrent rectal cancer: a review of inci-

dence, presentation, and management. Sem Oncol 1993;20:506-519.
10. Wanebo HJ, Koness RJ, Vezeridis MP. Pelvic resection of recurrent rectal cancer. Ann Surg 1994;220:586-597.
11. Vaughn DJ, Haller DG. Nonsurgical management of recurrent colorectal cancer. Cancer 1993;71 Sup:4278-4292.
12. Valentini V, MorgantiAG, DeFranco A. Chemoradiation with or without intraoperative radiation therapy in patients with locally recurrent rectal carcinoma: prognostic factors and long term coutcome. Cancer 1999; 86: 12-24.
13. Mohiuddin M, Marks G, Marks J. Long-term results of reirradiation for patients with recurrent rectal carcinoma. Cancer 2002; 95: 1144-1150.
14. Cummings BJ, Radiation treatment for rectal cancer 1995; 19: 275-281.
15. Hofer SOP, Molema G, Hermens RAEC, et al. The effect of surgical wounding on tumour development. Eur J Surg Oncol 1999; 25: 231-245.
16. 天理よろづ相談所医学統計解析グループ．Neyman-Pearson統計学から新しい臨床統計学へ．天理医学紀要 2002; 5: 100-116.(http: www. tenriyorozu- hp.or.jp/01hospital/06kenkyuu/igaku-kiyou/でダウンロード可能)
17. 天理よろづ相談所医学統計解析グループ．医療におけるエビデンスと *P* 値．天理医学紀要2003;6:54-79.(同上でダウンロード可能)
18. 前谷俊三．Boagモデルを拡張した新しい生存分析．天理時報社，天理，2002.
19. Maetani S, Nakajima T, Nishikawa T. Parametric mean survival time analysis in gastric cancer patients. Med Decis Making 2004;24:131-141.
20. Onodera H, Maetani S, Aung T, et al. Clinicl application of a blood pressure autoregulation system during hypotensive anesthesia. World J Surg 1999;23:1258-1263.
21. King SB III, Lembo NJ, Weintraub WS, et al. A randomized trial comparing angioplasty with coronary bypass surgery. N Engl J Med 1994; 331:1044-1055.
22. King SB III, Barnhart HX, Kosinski AS, et al. Angioplasty or surgery for multivessel coronary artery disease: comparison of eligible registry and randomized patients in the EAST Trial and influence of treatment selection on outcome. Am J Cardiol 1997;79:1453-1459.
23. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects Biometrika 1983;70:41-55.
24. Rubin DB. Estimating causal effects from large data sets using propensity score. Ann Intern Med 1997;127:757-763.
25. King SB III, Kosinski AS, Guyton RA, et al. Eight-year mortality in the Emory angioplasty versus surgery trial. J Am Coll Cardiol 2000; 35:1116-1121.
26. The BARI investigators. Seven-year outcome in the Bypass angioplasty revascularization investigation (BARI) by treatment and diabetic status. J AmColl Cardiol 2000; 35:1122-1133.
27. 福井次矢監訳．Fletcher RH, Fletcher SW, Wagner EH 著 . 臨床疫学．第3版，東京：メディカルサイエンスインターナショナル；1999.
28. Freeman PR. The role of p-values in analyzing trial results. Stat Med 1993; 12: 1443-1452
29. Royall RM. The effect of sample size on the meaning of significance tests. Am Stat 1986; 40: 313-315.
30. Weinstein MC, Fineberg HV. Clinical decision analysis Philadelphia, W.B. Saunders, 1980
31. Gardner MJ, Altman DG. Confidence interval rather than P values: Estimation rather than hypothesis testing. Br Med J 1986; 292: 746-750.
32. Simon R. Confidence intervals for reporting results of clinical trials. Ann Int Med 1986; 105: 429-435.
33. Bulpit CI. Confidence intervals Lancet 1987; 494-497.
34. Evans SJ, Mills P, Dawson J. The end of p value? Br Heart J 1988; 60: 177-180.
35. Maetani S. Another approach to clinical trial numbers. Lancet 1990; 335:114.

# Are radical treatments double-edged? Qualitative interaction between patients and treatments

Statistical Analysis Group (representative: Maetani S)[a], Onodera H[b]

[a] Tenri Institute of Medical Research

[b] Department of Surgery and Surgical Basic Science, Kyoto University Graduate School of Medicine

***Background*** : In randomized controlled trials (RCT) for comparing two treatments, it is generally assumed either that one treatment is more effective than the other, or that both treatments are equivalent. However, there may be cases in which treatment A is more effective than treatment B in some patients, but is less effective or even harmful in other patients. Two examples of such interaction are presented, to highlight the inherent limitations of current RCTs ( parallel comparison).

***Methods and Results***: Example 1. We studied retrospectively the long-term outcomes of 36 patients, who underwent radical en bloc resection of pelvic organs and walls between 1978 and 1997 for perineopelvic recurrence of rectal cancer after abdominoperineal excision. Although radical surgery achieved 10-year survival in 6 patients and 5-year survival in 10 patients, the postoperative CEA doubling time was significantly reduced in patients with second recurrence, suggesting that surgery accelerated their deaths. Example 2. We analyzed the reports of Emory Angioplasty versus Surgery, in which 392 of 842 eligible patients with coronary multivessel disease were randomized either to percutaneous transluminal coronary angioplasty (PTCA) or to coronary-artery bypass grafting (CABG); the remaining 450 patients, who met the eligibility criteria but refused the RCT, received one or other procedure according to their physicians' choice. Although the outcomes for patients randomized to PTCA are very similar to those randomized to CABG, the non-randomized group achieved a significantly higher 3-year survival than the randomized group (96.4% vs 93.4%); neither PTCA patients nor CABG patients in the non-randomized group displayed lower survival than patients randomized to the same procedure.

***Conclusion***: Current RCTs for comparing two treatments can determine the generally better treatment, but fail to show how many and which patients suffer from the "better" treatment when there is qualitative interaction between patient and treatment. This may cause suboptimal decision-making in the care of individual patients. Current RCTs have yet to be re-evaluated from the perspective of medical resources and benefit to patients.

**Keywords**: randomized controlled trial (RCT), patient-treatment interaction, locally recurrent rectal cancer, coronary artery disease, hypothesis testing

**依頼コメント**

森川　敏彦
武田薬品　統計解析部

## はじめに

本論文は天理よろづ相談所医学統計解析グループが継続してランダム化比較試験(RCT)を告発し続けている一連の論文の延長上にある．本論文で扱っているのは患者・交互作用の問題であるが，議論の対象としているRCTは主として並行群間比較デザインによるものを指すと考えられるので，本コメントにおいてもその前提で議論する．

本論文の内容を，私なりに整理して示すと以下のようになるかと思われる．

① 患者・治療交互作用は患者にとって極めて重要な情報である．
② しかるにRCTでは患者・治療交互作用が評価できない．
③ したがってこのような重要な情報を提供できないRCTは極めて不完全な試験デザインであり，再考が必要である．またこのようなデザインの結果に基づき治療方針を立てるのもEBMとはいえず誤りである．
④ 上記以外にも現在用いられている臨床統計学には多くの欠陥があり，それを是正すべきである．

このうち$についての議論は興味深いものの文献17でも議論されているので，ここでは本論文の主題である患者・治療間の交互作用に絞って検討したい．まず患者・治療間交互作用とはなにかについて触れた上で上記三点に絞って議論しよう．

## 患者・治療間交互作用

患者iの治療X,Yに対する反応をそれぞれXi,Yiとすると患者iに関する治療差はDi=Xi-Yiにより定義される．もし患者によってDiの値が異なるならば患者・治療間交互作用が存在するという．反応は大きな値の方がよりよいものと仮定すると，患者によらず常にDi>0ならば常にXi>Yiであり，たとえ交互作用が存在しても，治療Xは常に治療Yに優ることになる（逆に常にDi<0ならば治療Yは常に治療Xに優る）．このように差の大きさが患者により異なっていても差の向きが同じ場合に，量的交互作用と呼ぶ．また患者によってDiの不等号の向きが異なる場合，すなわちある患者ではDi>0，別の患者ではDi<0となる場合は質的交互作用と呼ぶ．患者・治療間に質的交互作用が存在する場合は患者によって治療を使い分けなければならない．本論文で特に問題としているのは質的交互作用である．因みにDi=D（一定）ならば，どの患者についても治療効果は同じである．すなわち患者・治療間交互作用は存在しない．このときD>0なら常に治療XがYに優り，D<0ならば常に治療YがXに優ることになる．D=0ならば常に両治療は同じ効果を示し，

正確な意味で同等であるといえる．

患者集団上でのXi, Yiの平均をそれぞれμ(X)，μ(Y)で表したとき，μ(X)> μ(Y)ならば，平均的に（あるいは周辺的な意味で）XがYに優るといい，μ(X)< μ(Y)ならば，平均的に（あるいは周辺的な意味で）YがXに優るという．患者・治療間に質的交互作用が存在する場合は，たとえμ(X)> μ(Y)となっても常にXがYに優るとは限らず，患者によってはXを適用することにより反って不利益を受けることになる．その意味でμ(X)> μ(Y)の場合は“周辺的な意味で”XがYに優ると呼ぶのである．通常用いられている並行群間比較デザインのRCTでは，同じ患者に同時に複数の治療を施して比較することができないから，ここで定義した患者･治療間交互作用が評価できず，したがって周辺的な意味での治療間の比較，すなわちμ(X)，μ(Y)間の比較しかできない．本論文の著者らはこの点を問題にしRCTを告発しているのである．

**①RCTの位置づけと患者･治療交互作用の意味**

比較臨床試験としてのRCTの位置づけは，対象とする母集団に当該治療を適用した場合の治療効果に関する全般的な結論付けを行なうものであると考えられる．そしてその結論は患者情報を十分持たない初期治療の指針となり得るようなものでなければならない．従ってRCTの結論は必ずしも特定の患者に対する最適治療の指針を与えるものではない．また技術の進展に伴い治療法間の優劣関係が逆転することもあり得る．promisingではあるがimmatureであり未だ確立していない治療法については特にそうである．RCTの結論は，あくまでも対象母集団全体に対する現時点でのオーバーオールとしての結論であることを忘れてはならない．

しかしこのような情報は，医師が患者と治療との関係に関する十分な情報を持たない場合の治療方針を決める際に役立つ．たとえば大規模RCTによりA薬がB薬よりも延命効果があることがわかったとしよう．医者は目の前の患者さんに対してA薬が不適応であるとか，特別の理由のない限りまずA薬を患者に投与しようとするだろう．より現実的にはA薬がB薬よりも延命効果において優れるといってもそれはほんの僅かであり，実質的に差がないと判断すれば，自分が使い慣れたB薬の方を選択するかもしれない．あるいはA薬が高価なものであれば医療経済学的効果も考慮するかもしれない．薬に限らず手術療法など他の治療法についても同じような側面があるだろう．

もし医師が患者・治療間交互作用について十分な情報をもっていれば，平均的にある治療が他の治療よりも優れるというだけでその治療を選択するということはしないであろう．その結果患者は不要なリスクを避けながら有効な治療を受けることが可能となる．

**②RCTでは患者・治療間交互作用が評価できないか？**

本来患者・治療間交互作用の有無を知るためには各患者に治療A,Bを共に施し，その結果を比較評価しなければならない．このような試験（matched pair design）を実際に実施することができたとし，結果として表１のような仮想的な結果を得たとしよう．ここに＋は正反応（有効），－は負反応（無効）を表す．実際にはサンプリングによる

ばらつきが存在し，試験毎にデータはばらつくが，議論を簡単にするためサンプルサイズは十分に大きく，データは十分に患者母集団を反映しているものとしよう．

**表1．患者に治療A,Bを共に施した場合の仮想的な結果**

| 両治療と反応 | | 治療B | | |
|---|---|---|---|---|
| | | ＋ | － | 計 |
| 治療A | ＋ | a | b | $p_A$ |
| | － | c | d | $1-p_A$ |
| | 計 | $p_B$ | $1-p_B$ | 1 |

表１で$p_A$は治療Ａの有効率，$p_B$は治療Ｂの有効率を表す．表中のセルｂは治療Ａで有効，治療Ｂで無効となる患者の割合，逆にセルｃは治療Ｂで有効，治療Ａで無効な患者の割合を表す．表中のb, c の存在が，患者・治療間の交互作用を生み出すことになる．ここでもしｂ＞0，かつｃ＝0なら，治療Ｂの方が治療Ａよりもよい患者層がないので，治療Ａは治療Ｂに対し常に優る(dominant)ことになる．またｂ＝ｃ＝0なら，本来の意味で治療法ＡとＢは同等ということになる．通常実施される並行群間比較試験としてのRCTからはこのような情報は得られず，単に周辺割合$p_A$, $p_B$に関する情報だけが得られるのみである．従ってこのような意味から本論文で「RCTからは患者・治療間交互作用に関する情報は得られない」とする主張は正しい．

通常のRCTでは，$p_A > p_B$なら“周辺的な意味で”治療Ａが優り，$p_A = p_B$なら治療ＡとＢは“周辺的な意味で”同等ということになる．しかし患者・治療間交互作用まで考えると，たとえ，$p_A > p_B$という結果になったとしても無条件に治療Ａが治療Ｂに優るとはいえない．これは著者らが主張しているように集団評価と個別治療との違いである．

上記のように，RCTは明らかに周辺優越性，ないし周辺同等性を調べるためのデザインであり，a, b, c, d セルの中身は判らないから，著者らのいうように患者・治療間交互作用に関する情報は与えない．その意味で個別治療の立場からは不完全で不十分なデザインであるといえよう．著者らのこの見解には筆者も同意する．

領域によっては両側試験（皮膚病など）やクロスオーバー（CO）試験（慢性疼痛や不眠など）を実施することが可能で，その場合は完全でないにしても表１に示したような情報を得ることが可能であろう．しかし重大な臨床アウトカムをエンドポイントとする場合，このようなデザインが組める場合は少ない．従って多くの場合は並行群間比較試験にならざるを得ず，その場合に論文で述べられているような欠陥を免れることはできない．特に治療方法が大きく異なる場合にそのような懸念が大きくなる．

**③このような重要な情報を提供できないRCTというデザインは極めて不完全な試験デザインか？**

もし両側試験やCO試験の適切な実施が可能だったとして，その結果表１に示すような患者・治療間交互作用が判明したとしても，それだけでは十分ではない．どのような患者で治療Ａの方がよく，またどのような患者で治療Ｂの方がよいのか，更にはどのような患者でA,B共に効果が発揮され，どのような患者でA,B共に結果が思わしくないのかを分析しなければならない．

そうでなければ著者らが主張するような個別治療に結びつかない．

遺伝子検査も含め，交互作用に影響を与える共変量の探索が行なわれることによって交互作用を示唆する要因が見つかるかもしれない．その場合は，その要因で層別した上で再度RCTを実施することにより（患者・治療間交互作用が存在するという）作業仮説が検証されることになるだろう．

一方もしこのような場合にはRCTにおいても同様にサブグループ解析あるいは交互作用解析により，交互作用を与える要因が探索されることになるであろう．並行群間比較デザインは交互作用を見つけるにはCOデザインよりもはるかに効率が悪いが，もし結果に大きな影響を与える因子が存在すれば，遺伝子，性別，年齢，重症度，罹病期間，マーカーなど内的及び外的諸因子の探索により見つかるであろう．表2はそのような典型的な場合を表示しており，この例ではRCTによる層別（サブグループ解析）の結果は対応のあるデータの場合と基本的に同じ情報を与える．本論文における直腸癌の例も基本的に似たような状況になっていると考えられる．そもそも有効無効というように同一治療においても反応を示す患者と示さない患者がいるということ自体が患者・治療間交互作用の可能性を示唆するものと考えることもできよう．

従って結論として並行群間比較デザインは交互作用検出の立場から不利で効率も悪いかもしれないが全く不可能というわけでもなく，依然として治療効果を評価する有効なデザインであると考える．

**③’ CABGとPTCAの比較結果（EAST）の解釈**

やはり表１の上で考えよう．医師はこれまでの知識と経験により当該患者が表a, b, c, dのいずれに属するかがある程度予測できるものとしよう．たとえば三枝多病変等病巣が広がっている場合は局所的な対応では収拾が付かずPTCAよりはCABGが選ばれるかもしれない．いずれにしても医師は可能な療法のリスクとベネフィットを天秤に掛けながら治療法を選択するのだろう．このような観点からEASTにおいては，医師が対象患者に対してPTCA，CABGのいずれかが明らかによいと思った場合にその治療を選択し，判定が難しい場合にランダム化を選んだのかもしれない．病態そのものが致死的で重大なものであるだけに，そのような判断はあり得たであろう．その結果ランダマイズを拒否した患者では表１でのb, cの部分が相対的に多くなった可能性がある．

このような推論はまさに憶測に過ぎない

**表２．質的交互作用に強い効果を与える共変量で層別した場合の典型的な結果**

| 層別要因 | | 層１ | | | 層２ | | | 全体 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 両治療と反応 | | 治療B | | | 治療B | | | 治療B | | |
| | | ＋ | － | 計 | ＋ | － | 計 | ＋ | － | 計 |
| 治療A | ＋ | 0 | b | b | 0 | 0 | 0 | 0 | b | b |
| | － | 0 | 0 | 0 | c | 0 | c | c | 0 | c |
| | 計 | 0 | b | b | c | 0 | c | c | b | b+c |

が, もしそのようなことが生じているなら, 結果としてランダム化を拒否した患者群で相対的に良好なレスポンスが得られたことの説明がつく．それは患者選択の問題（患者から見れば治療選択）であり，結果として同一の患者層上での比較にはなっていない．必然的に非盲検となることから管理バイアスや測定バイアスなどが入り込み比較可能性を保証し得ない問題もある．

以上述べた意味だけではなく，このようなバイアスを伴った条件下の比較結果を鵜呑みにしてその後の治療方針を決定すれば, それこそ著者らが指摘するように重大な誤りを招く恐れがあるのではないかということを逆に危惧する. 本例の対象疾患は致死的な性質をもっているから尚更のことである．

私は著者らの述べているようにEASTの著者達の結論に同意しないが，だから本論文の著者らの主張に全面的に賛成するわけでもない．一般に医師の選択する治療法の方がランダムに治療法を選択するよりも優るという蓋然性はあるにしてもその保証はどこにもない．十分な情報がないことによる誤解や間違った信念から誤った治療を選択する可能性もある．

## おわりに

著者らが主張するように，倫理的な観点から患者がRCTに晒される危険性は最小限にすべきであり，致死性を持っているような疾患の場合は特に慎重にならなければならない．その領域でどこまでわかっているか，どこから先がわかっていないかを慎重に見極め，患者リスクと試験から得られる社会的利益を十分に吟味したうえで，実施の可否も含め計画を立てなければならない．

RCT，並びに臨床統計学に関して著者らの批判するところはよく理解できる．いずれも方法論上の不完全性に基づいており，またその臨床応用上の不適切性に基づいている．本論文においても個別治療の立場から患者・治療間の交互作用の評価の必要性と重要性が説かれ，そしてその重要性にも関わらず，現実に実施されている臨床試験で交互作用解析の実践が十分になされていないことに対する批判がなされている．これらを臨床試験の実施者は謙虚に受け止め，今後の計画と実施に生かしていかなければならないと考える．

依頼コメント

# 冠状動脈バイパス術とランダム化比較試験，そして患者－治療間交互作用
## 討論:　心臓外科医の立場から

阿部知伸　，　　上田裕一

名古屋大学医学部附属病院　心臓外科
名古屋大学大学院医学研究科　病態外科学講座胸部機能外科

### はじめに

安定狭心症に対する冠状動脈バイパス術（CABG）というのは，著者らが提示された患者・治療間交互作用，そして手術適応を決めるstudyとしてのRandomized Controlled Trial (RCT)の役割を考える上で，示唆するところの多い一例であると思われる．このコメントでは，著者らが述べた点に心臓外科医の立場から若干の意見を付け加えて責を果たしたい．

“ CABGは両刃の刃 ”ということならば，心臓外科医はそれを痛みとともに知っている．我々は手術死を経験するからである．「こんなことなら手術なんか受けさせなければ良かった」という御遺族の率直な感想に対して，我々は自己防衛的に反論したい誘惑に駆られるけれども，正しいのは御遺族である．その患者さんに限ってはおそらく手術はしなかった方が良かった．なぜなら左主幹部病変での不安定狭心症など特殊な場合を除けば，狭心症の予後は普通，数日以内に死ぬというほど極端に悪いものではない．その患者さんが手術後数日で死んでしまったのは，CABGが害をなした，“ harm であった ”可能性が極めて高いのである．

### Emory Angioplasty versus Surgery Trial (EAST)の背景

患者・治療間の交互作用ということを，安定狭心症の患者全体とCABGについて考えると，その存在は明らかであるといえる．すなわちCABGは重症の，high riskの狭心症に対して生命予後・機能予後を大きく改善するが，軽症の，low riskの狭心症に対しては行う意味が余り無いといえるからである．

1960年代後半にCABGは登場した．それまでになかった冠状動脈に直接血行再建するというこの治療は，理論的に虚血性心疾患の予後を劇的に変えるかも知れないと考えられ，1970年代，この手術について，内服のみの内科的治療をcontrolとして，安定狭心症の患者を対象に一連のRCTが行われた．ただし，当時行われた七つのRCTのうちで，overallのrandomized cohortの生存率で有意差を出したのは，実に一つしかない.[1] その結果とそれに引き続いた当時の循環器内科医，心臓外科医の反応は興味深いもの

であり，結果の誤った解釈が多く発表され，ことに「Coronary Artery Surgery Survey (CASS )などは知見よりも混乱をもたらした」などと言われた.[2] しかし，今日の知識を以ってこれらのstudyを振り返るとき，私たちは「何故多くのstudyで差が出なかったのか」容易に説明出来る．前述したように，CABGは安定狭心症の中でも，内科的予後の悪いhigh riskなsubgroupにしか延命効果がないのである．low riskの患者を多く含み，high riskの患者は倫理的配慮からむしろ厳しく除外されていた背景により，これらのRCTでは差は非常に出にくくなった．いま一つの理由は，結果がintention-to-treat basisで解析されたことで，内科的治療からCABGにcross-overした症例（これはCASSで10年間に約35%に起こった）はそのまま内科的治療群としてfollow-upされ，両群の生存率の差はさらに薄められた．

しかし同時に，今日まで用いられている教科書的なCABGの手術適応は，これらのstudyのsubset解析から決められたといってよい．すなわち，冠動脈造影上，CABGは左主幹部病変の症例に対して極めて強い延命効果があった．三枝病変や左前下行枝近位部病変を含む二枝病変でもやはり生命予後の改善が認められた.[3] これらの生命予後の改善は，狭心症状の強い患者群で強く，また左心収縮能低下の強い症例で強かった.[4]

1970年代後半，CABGに約10年遅れてPTCA（経皮的冠状動脈形成術．バルーン拡張術で始まったが，ステント留置など各種手法を含めてPCI：経皮的冠状動脈インターヴェンションと最近は総称される）が登場した．ここに於いて，冠状動脈の血行再建にはPTCAとCABGの2種類の治療戦略が可能となった．PTCAは主にlow riskの患者でmedical therapy（いわゆる薬物療法）との間でRCTが組まれ，主に狭心痛の緩和をendpointに有用性が証明された.[1] なお，安定狭心症では，PTCAの延命効果は今日までRCTで証明されてはいない．これは主に軽症安定狭心症の内科的な生命予後が，差を証明するには良すぎることによると考えられる．その後，PTCAは徐々に多枝病変に適用されるようになり，多枝病変安定狭心症の標準治療であるCABGに対してRCTがデザインされた．今回引用されたEASTはその文脈で行われたRCTの一つと位置づけられる．

### EASTの低いrandomization率について

EASTのdesignについては著者らが概説されたとおりである．CABGに対して新しい治療であるPTCAの相対的効果を見るこのstudyは，CABGのmedical therapyに対する優位性が証明されている多枝病変を対象とした．

心臓外科医の立場から付け加えて置きたいことは，このstudyではごく少数の患者しかrandomizeされていないという点である．5118人の多枝病変をもつ患者のうち，3371人，65.9%が冠動脈所見から除外されている．除外の理由は，左主幹部病変，径1.5mm以上のviableな心筋を栄養している血管の8週間以上の完全閉塞，2本以上の冠状動脈の完全閉塞，長さ20mm以上の狭窄病変などであり，これらはいいかえればPTCAに不適当でCABGが適応と考えられる病変である．これらの冠動脈造影上のexclusion criteriaで除外されなかった患者は，さらにPTCA, CABGそれぞれの術者から直接不適

当として除外できるという二重の逃げ道が用意されている．この段階で除外された患者が191名，そのうち169名はPTCA不適当と考えられた症例で，多くはCABGにまわった．こうして結局trialにeligibleと考えられたのは最初にscreenされた多枝病変の患者の16.5%に過ぎない．[5]

これは「このstudyの結果を臨床判断に用いたい」と臨床医が考えるとき非常に大きな欠点となる．単純に考えて循環器内科医が日々の診療で診る多枝病変の症例の80%以上がこの研究の対象から除外されているのである．むしろ，例外的ともいえる症例しかrandomizeしていないこのstudyの結果を，日常診療で診る多枝病変の患者全体にあてはめるのは問題である．これはBypass Angioplasty Revascularization Investigation (BARI)も含めたこの時期の一連のCABG versus PTCAのRCTの共通の欠点であった．

## Registry GroupとRandomized Groupの差はなぜ生じたのか

Registry GroupとRandomized Groupの差がなぜ生じたのか，ということについて，著者らはRegistry GroupでCABGとPTCAがうまく使い分けられた影響が大きいのではないかという推測にそって論を進めた．述べたところは理にかなっており，原著で示されたEmory大学のデータからこれ以上議論するのは蛇足となろう．差はbaselineのわずかな違いから生じた可能性が否定できないし，著者らの言う通り治療の選択によって生じたのかも知れない．EASTの他の文献，そしてBARIの論文から数点補足するに留めたい．

まず一つ確認しておきたいのは，EASTでは冠動脈造影上の重症度によるsubset解析で，三枝病変や左前下行枝近位部病変のsubgroupでもCABG群とPTCA群との間に生存率で統計学的有意差は示されなかった，ということである．[6] 8年間のフォローアップで，三枝病変では生存曲線はCABG群で良いように見えるが，log-rank testによるp値は0.35，左前下行枝近位部病変を含む患者群でもCABG後の生存率が良さそうに見えるがp値0.16であった．[6] これは循環器内科医が三枝病変，左前下行枝近位部病変をCABGに回したからregistry群がよかった，という単純な説明を少し難しくすると考える．

ただし，上の結果はEmoryの著者達が自ら述べているように，studyがunderpowerであったことによる可能性が高い．[6] 問題のRegistry GroupとRandomized Groupの比較では僅かな差が相加的に働いて有意差に至った可能性は否定できないと考える．またこれら単純な分類以上の微妙な冠動脈造影上の差で循環器内科医が治療法を分けた可能性もある．後段でDuke大学の研究結果を示すので読んで頂きたい．そして勿論，冠動脈造影所見以外のところで循環器内科医が治療を使い分けた可能性があるのは著者らが述べた如くである．

原著でEmoryの著者達は，PTCA後の生存率がRegistry群よりRandomized群で悪くなるのは，Randomized群でより重症な症例をPTCAで治療することになるから分かるが，CABGでRandomized群の方が生存率が悪くなるはよく分からない，と述べている．[7] これについて少し解説すると，CABGが行われた患者では，Randomized群の方が冠動脈造影上の二枝病変などより軽い症例が多

くなったはずである.“軽い症例”が“CABGの良い適応”ではない，と私たちが考えるのはCABGによる死亡率が高かったり長期予後が悪かったりするためではない.“軽い症例”が“CABGの良い適応”でないのは，内科的な予後が良いのでCABGによるさらなる予後の改善が少ないからである.絶対的な数字としては,“軽い症例”の方が“重い症例”より当然ながらCABGによる周術期の死亡率は低いし，長期生存率も良い．したがって，軽い症例のより多いはずのRandomized群の方が生存率が悪いのはよく分からない，ということになる.

ここに，この現象を説明するかも知れないデータがあるので紹介しておく.BARIでは糖尿病の患者においてCABG群の方がPTCA群より生存率が良かったという結果が出ており，これはこのstudyの最も重要な結果でよく引用されている．この糖尿病群の患者に関してやはりRegistry群とRandomized群を比較した研究がある．ここで興味深いのは，CABGではRegistry群の方がRandomized群より予定通り血行再建できた率が高い（96%対87%, p<0.05），ということである.[8] 完全血行再建の成否はCABG後の長期生命予後に影響することが知られている．標的血管のgraftability(グラフトを外科的に建てられるかどうかということ)は，血管の径と性状で決まる．余りに細い冠状動脈の枝，余りに動脈硬化による石灰化が強い血管は，グラフトを技術的に吻合できないか，あるいは吻合できても閉塞する．殊に私たち心臓外科医は冠動脈造影からこのgraftabilityを読もうとするのであり，これはこの二枝病変，三枝病変，といった分類には組み込まれていない変数である．BARIでRegistry群の方が予定通りの血行再建の率が高いのは，循環器内科医がgraftabilityが高い症例をよりCABGに回し，graftabilityが低い，血行再建が難しい症例は多分内科的治療に回していた（一般にグラフトが建てられないような血管はPTCAにも不適である）可能性がある．すなわちRandomized群ではgraftabilityが無いような症例まで強制的にCABGに割り付けられていたのではないか，という推測が成り立つと考える.

### Duke大学のDatabase　Studyで示される患者ー治療間の質的交互作用

最後にDuke大学の大きなdatabase studyの結果を紹介したい．虚血性心疾患に於けるPTCAとCABGの患者ー治療間の質的交互作用の存在を強く示唆する研究である．かつRCT以外のdesignのstudyからも素晴らしく有用な情報が得られるという当然のことを示す好例であると考える.

研究はDuke大学で7年間に心臓カテーテルで虚血性心疾患が診断された9263例，内科的治療2449例, PTCA 2924例, CABG 3890例の長期生存率の解析である.[9] この研究で出色なのは, 独自のCoronary Anatomy Scoreによる虚血性心疾患のきめ細かい冠動脈造影上の分類である．一般に虚血性心疾患の冠動脈造影上の分類は, 左主幹部病変, 一枝病変, 二枝病変, 三枝病変に分類するものが多く，左前下行枝（LAD）の近位部病変は，生命予後に重大な影響を及ぼすので，これを分類に加えることもある.紹介するDukeの分類はさらに細かく,冠動脈病変を12に分類している．分類に用いたのは75%以上の狭窄の数，95%以上の狭窄の数，そして

左前下行枝の病変の位置と狭窄度である．そしてDukeの6034例の内科的に治療された虚血性心疾患（別のData set）の予後の解析に拠って，冠状動脈病変のない患者を0，95%以上の左主幹部病変の患者を100として，Relative Prognostic Weightが並ぶように上記の変数の組み合わせで1から12に分類した．この解析では絶対的にCABGが良いと考えられる左主幹部病変の症例を除外しているので，12のうち9段階のscoreでそれぞれの治療は比較された．endpointは心血管関連死で，hazard ratioで治療間の相対的な利益は検討された．この解析で，虚血性心疾患が重症度に応じて治療への反応が変わるのが，ほとんど連続的に（9段階の分類は私たちにむしろ連続的な変化と認識させる）示された．CABGとmedical treatmentの比較では，やはり重症なほどCABGの治療効果が大きくなり，この効果はScore5，95%以上のLAD病変がある一枝病変ないしは甘いLAD狭窄がある二枝病変，以上で明らかである．そして，注目すべきCABGとPTCAの比較ではScore 6以上でCABGがPTCAより生存を改善することが明らかで，Score 5でほぼ同等，Score 3-4でPTCAの方が良い傾向が認められ，Score 1-2でPTCAの方が明らかに良かった．つまり，虚血性心疾患全体でPTCAとCABGを比較して，ついに質的交互作用，ある症例に対してはPTCAの方がよいし，ある症例に対してはCABGの方がよい，ということが示された研究といえる．

ただし付け加えて置くと，EASTではそもそも多枝病変しか対象にしていないので，Dukeの Score 3未満の症例は含まれていなかったし，さらに多枝病変の中でも多くはExclusion criteriaで除外されており，この二つのstudyで検討しているのは虚血性心疾患の中でも大きく異なる患者群である．EASTのRandomization Eligibleの患者群の中で質的交互作用があったかどうかは，このDukeの研究の結果だけからは分からない．

## おわりに

以上，Emory大学のPTCA対CABGのRCTのRegistry群とRandomized群での比較に対する著者の統計学的見地からの論文に，心臓外科医の臨床的立場からの若干の考察を加えた．

既に述べたように，狭心症とCABGというのは，重症度に応じて，強い患者―治療間交互作用があるペアーではないかと考える．1960年代に生まれたこの手術は，これまで常にRCTで検証され続けているのではあるが，RCTと臨床の関係が単純ではなかったのはこの患者―治療間交互作用のためだった，といういい方も出来るかもしれない．ここで単純ではない，といったのは，殆どのstudyで，核となる臨床的な疑問に最も明確に答えを出さなければならないはずの，overallのrandomized cohortでの有意差検定の結果が，そのまま日常の診療に生かせる答えになって来なかった，ということである．

PCIには巨額の研究費がつぎ込まれ，次々と新しいdeviceが開発されている．一方，CABGも周術期死亡率は近年は著減し，長期予後をよくするための工夫も多く見られている．これからもCABGは多枝病変治療のdefending championとして新しいdeviceと比較され続けなければならない

運命にあるといえる．かくも患者―治療間の交互作用がある，あるいはheterogeneousな虚血性心疾患では，著者が述べるようにRCTの用い方に工夫が必要であるのかも知れない．

**謝辞**

大変おもしろい着眼から議論を展開された前谷先生に改めて敬意を表するとともに，CABGとRCTというテーマについて整理させて頂く機会を与えられたことに感謝いたします．

**参考文献**

1. Rihal CS, Raco DL, Gersh BJ, et al. Indications for Coronary Artery Bypass Surgery and Percutaneous Coronary Intervention in Chronic Stable Angina. Review of the Evidence and Methodological Considerations. Circulation. 2003;108: 2349-2445.
2. King SB 3rd. The coronary artery surgery study (CASS)--widely misinterpreted Med Assoc Ga 1984;73(4):251-2.
3. Varnauskas E. Twelve-year follow-up of survival in the randomized European Coronary Surgery Study. N Engl J Med 1988; 319: 332-337.
4. Alderman EL, Bourassa MG, Cohen LS, et al. Ten-year follow-up of survival and myocardial infarction in the Randomized Coronary Artery Surgery study. Circulation 1990;82:1629-1646.
5. Spencer B, King 3rd , Nicholas JL, et al. Emory Angioplasty Versus Surgery Trial (EAST): Design, Recruitment, and Baseline Description of Patients. Am J Cardiol.1995;75:42C-59C.
6. Spencer B, King 3rd, Kosinski AS, et al. Eight-Year Mortality in the Emory Angioplasty Versus Surgery Trial (EAST): JACC.2000;35(5): 1116-1121.
7. Spencer B, King 3rd, Barnhart HX, et al. Angioplasty or Surgery for Multivessel Coronary Artery Disease: Compatison of Eligible Registry and Randomized Patients in the EAST Trial and Influence of Treatment Selection on Outcomes: Am J Cardiol. 1997;79:1453-1459.
8. Katherine MD, Guo P, Holubkov R, et al. Coronary Revascularization in Diabetic Patients A Comparison of the Randomized and Observational Components of the Bypass Angioplasty Revascularization Investigation (BARI). Circulation. 1999; 99:633-640.
9. Jones RH, Kesler K, Phillips 3rd HR et al. Long-Term Survival Benefit of Coronary Artery Bypass Grafting and Percutaneous Transluminal Angioplasty in Patients with Coronary Artery Disease. J Thorac Cardiovasc Surg. 1996; 111 (5):1013-1025.

## 依頼コメント

斉藤成達，木村　剛
京都大学医学部大学院　循環病態学講座

EBM(Evidence Based Medicine，証拠に基づく医療)の実践は今や，臨床医の新たな行動指針となったといっても過言ではない．EBMの推進により，医療の質と向上と標準化が図られ，地域や医師による診療内容のばらつきを極力減らすことが可能となる．また根拠のある臨床判断を患者に明示することによってインフォームド・コンセントも得やすく，医師と患者のよりよい信頼関係の構築にも役立つ．EBMにもいろいろなレベルがあり，米国予防医療サービス特別研究班は証拠の質の程度によって5段階に分類している．これによると，「某先生の理論によれば」というのは最低レベルのエビデンスであり，ランダム化比較試験(RCT)が最善のエビデンスだとされている．また米国保健政策研究局(AHCPR)は，複数のRCTを統合したメタ・アナリシスをさらに上位のエビデンス研究として位置づけている．

しかしながらランダム化比較試験の欠点も指摘され始めている．Concatoらは，大型観察研究は必ずしもRCTに劣るものではないことを2000年にN Engl Medに発表している．RCTにおいては観察研究で問題とされる不測の予後因子を均等に振り分けられると考えられているが，実際には経済的，倫理的，リクルート等々の問題があり，完全に予後因子を排除できるわけではなく，偏った亜集団を対象にしている可能性が指摘される．これが“RCTは内的妥当性においては優れているが，外的妥当性においては劣る”とされる根拠である．また最上位のエビデンスとされるメタ・アナリシスに関してもpublication biasが存在することを考えると，やはり外的妥当性には問題があるものと考えられている．本稿では患者－治療間の質的交互作用が問題とされており，RCTに対する新たな疑問の提起がなされている．この問題については，サブグループ解析やRCTのentry criteriaの設定によってある程度，解決することができると考えられるが，著者らの指摘するように臨床試験の費用，労力，時間，被験者数には限りがあるのも事実である．

本邦においては臨床試験を行う基盤が十分に整備されておらず，RCTについては皆無といっていい状況である．本稿において指摘された問題点などを参考にし，本邦独自のEBMを構築することは急務と考えられる．