

附録 D ランダム誤差とバイアス

(1) ランダム誤差とバイアスの意味と相違

ランダムに発生するランダム誤差 (random error) には方向性がなく、バイアスとは明確に区別する必要がある。ランダム誤差の場合は、コイン投げを 100 回しても必ずしも 50 回表が出るとは限らないのと同じで、完全にバランスがとれたコインでも 10 回投げた場合に 8 回表が出てしまうこともあり、10 回中 9 回表が出ることもあるだろう。さらに、稀ではあるが、10 回とも表が出ることもある。コインを投げた結果の相違は偶然の作用によって左右されるもので、これをランダム誤差という。また、このような測定値のばらつきをランダム変動という。

一方、バイアスは系統誤差 (systematic error) であり、真の値 (true value) からの方向性のある逸脱である。真の介入効果が過小評価されるようなバイアスもあれば、過大評価されるバイアスもある。たとえば、資金源バイアスというものがある。これは、営利目的団体が資金を提供している臨床試験がそうでない臨床試験に比べて、当該団体の利益につながるような薬剤または患者管理に有利な結果を出すバイアスのことである。

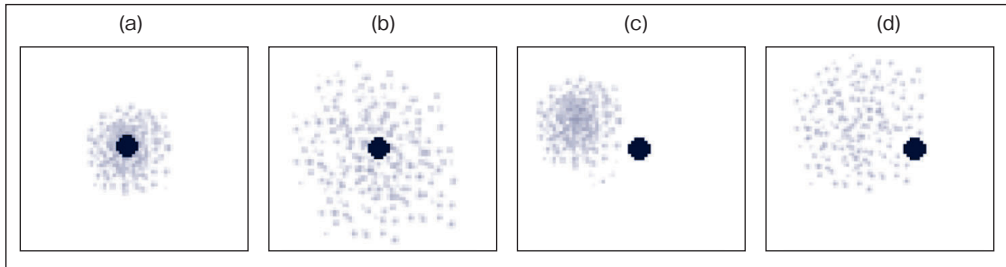
通常、ある特定の研究結果がどの程度バイアスの影響を受けているのかを把握することはできないが、RCT のデザインや実施・分析における限界がバイアスにつながるものが経験的に示されている。しかし、現実的には研究の方法論的欠陥があるにもかかわらず、研究結果にバイアスがないこともあるため、バイアスのリスク (risk of bias) を検討することが適切である。

risk of bias に着目すると、システマティック・レビューの潜在的な結果のばらつきの原因がみえてくる。厳格に実施された試験ほど、より真の値に近い結果を出す可能性が高いが、妥当性にばらつきがある一連の研究を統合すると (メタアナリシス)、バイアスによって介入の効果が過剰評価 (介入効果が有意なものと判断) されたり、過小評価 (介入効果が有意ではないと判断) されたりする。また、レビューに加えられた 1 つ 1 つの研究結果の妥当性やばらつきの有無にかかわらず、統合した研究における risk of bias を評価することが重要である。たとえば、一連の研究から一貫した結果が得られていても、すべての研究に限界が存在することもあり、このような場合、レビューの結果からの結論も偏ったものになる (第 1 章, 「コクラン risk of bias」, 48 頁)。

(2) ランダム誤差とバイアスの相違

附図 D-1 は、4 種類の研究が、小さなサンプルサイズと大きなサンプルサイズ、RCT と観察研究の組み合わせを表現している⁴⁷⁾。黒点は真の効果で、周辺の小さな点は個々の研究である。黒点と小さな点の距離が増加するほど、その研究結果と真の効果の差が増大する。

附図 D-1 ランダム誤差とバイアスの例（4 種類の研究）



(a) : サンプルサイズの大きな RCT, (b) : サンプルサイズの小さな RCT, (c) : サンプルサイズの大きな観察研究, (d) : サンプルサイズの小さな観察研究

図 a, b はランダム誤差の理解に役立つ

図 a は、サンプルサイズの大きな RCT 群で、研究デザインが厳格で各々の研究は妥当なことから、中心の黒点（真の効果）の周辺に各研究が分布している。当然ながら、偶然すなわち、ランダム誤差のため、各研究は正確に真の効果を示すわけではなく、黒点から遠ざかることもある。

図 b は、図 a と同様に各研究デザインは RCT で強固であるため真の治療効果の周囲に散らばっているが、サンプルサイズが小さいために、ランダム誤差が増大し、個々の研究は黒点から大きく遠ざかって広範囲に分布している。

この 2 つを比較することで、これらのデータ（研究 a, 研究 b）を統合すること（メタアナリシス）の理論的根拠が理解できる。すなわち、小さなサンプルサイズの研究を蓄積してサンプルサイズを大きくすると、ランダム誤差を減らし、推論を強固なものにすることができるということである。

図 c, d はバイアスとランダム誤差の関係の理解に役立つ

図 c は、観察研究であり、小さな点全体の中心は真の効果から遠くかけ離れている。これは研究デザインが観察研究のため、たとえサンプルサイズが大きくても、バイアスが生じやすいためである。各研究は同じ研究デザインであるために、各々の研究に同方向のバイアスが生じる。サンプルサイズが大きいので、ランダム誤差は小さくなるが、結果は誤っている。

図 d は、一連の観察研究がバイアスのために真の効果から方向性をもって離れていることを示している。さらに、各研究のサンプルサイズが小さいため、ランダム誤差も加わり、結果に大きなばらつきも加わっている。

これらのデータ（研究 c や研究 d）に対してメタアナリシスを試みるのは、これは誤った結果を導く可能性が高く非常に危険である。なぜならば、メタアナリシスを行うことでサンプルサイズは大きくなるが、ランダム誤差が大きく厳格ではない誤った推定を、ランダム誤差が小さく厳格な誤った推定で置き換えることになるため、バイアスをもった結果の推定をより強固にするだけで、バイアスの存在自体は不変である。

(3) バイアスを少なくする方法：治療と害に関する研究

研究者はバイアスを少なくするために、さまざまな対策を行う必要があり、RCT や観察研究においてバイアスが生じやすいポイントと、バイアスを減らす方法（附表 D-1）が報告されている⁴⁷⁾。これらのポイントは、GRADE におけるエビデンスの質（研究の限界など）の評価に参考となるものである。

附表 D-1 治療と害の研究においてバイアスを減らす方法

バイアスの生じる部位	治療： バイアスを減らす方法	害： バイアスを減らす方法
研究開始時に観察される相違点		
治療群と対照群の患者が予後の点で異なる	ランダム割り付け	データ解析における予後因子の統計学的補正
	層別化を伴うランダム割り付け	マッチング
研究が進むにつれて生じる相違点		
プラセボ効果	患者の盲検化	プラセボ効果を受けにくいアウトカム（例、死亡率）の選択
共介入	介護者の盲検化	治療の相違点の文書化と統計学的補正
アウトカム評価時におけるバイアス	アウトカム評価者の盲検化	観察者のバイアスが受けにくいアウトカム（例、死亡率）の選択
研究終了時の相違点		
追跡からの脱落	追跡の完了を確実にする	追跡の完了を確実にする
大きな効果が得られたための研究の早期終了	当初の計画通りに研究を完了させる	
割り付けられた治療を受けなかった患者を除外する	ITT 解析 ^e を遵守（*）	

(引用許可：JAMA Users' Guide 2008)

*** ITT (intention to treat) 解析**

臨床試験では、対象者の追跡からの脱落 (lost to follow-up) やプロトコルからの逸脱がしばしば生じるが、両者の意味は異なることに注意すべきである。追跡からの脱落とは、臨床試験などでエントリーされた患者が途中で脱落しアウトカムデータが欠損しているものである。脱落率について、経験的によく使われる脱落率 20%以内（追跡率 80%以下）などの閾値は無意味であり、どれくらいの脱落率なら「十分少ない」といえるかは、アウトカムのイベント発生率に依存している。

一方、プロトコルからの逸脱とは、研究開始時の割り付け後に全く治療を受けていない（治療拒否、服薬忘れ、手術待機など）という理由で、介入群から除外して扱うことをいう。逸脱例の中にはアウトカムデータが得られている場合があり、追跡からの脱落とは異なる。これは明らかに ITT 解析の原則を遵守していないことになる。追跡からの脱落やプロトコルからの逸脱は、研究の限界につながる大きな要因であり、通常いずれもすべて当初の割り付け群に含めて解析する方法が intention to treat (ITT) 解析である。

^e 比較試験において、割り付けられた介入を実際に受けたか（あるいは終了したか）に関わらず、あらかじめ割り付けられた群として解析する手法。対して、実際に治療介入を受けた者のみを対象として解析するものが、on-treatment analysis である。

RCTにおける脱落や逸脱の程度がどの程度われわれの確信性にとって重大なのかは、worst case scenario 法^fを使って判断できる(附表 D-2)。すなわち、治療群が対照群よりも有意に優れているという結論を出しているRCTの場合、治療群から脱落した患者は全て転帰が不良で、対照群から脱落した全ての患者は転帰が良好だったと想定して、両群の差を解析する手法である。この解析で、もし結果が不変ならば、脱落者の割合は多いものではないといえる。逆に、結論が変わるようならば、その比較試験の妥当性は低いものになる。

附表 D-2 追跡からの脱落

比較試験	試験 A		試験 B	
	介入群	対照群	介入群	対照群
ランダム割り付けされた患者数	1,000	1,000	1,000	1,000
追跡からの脱落 (%)	30 (3)	30 (3)	30 (3)	30 (3)
死亡例数 (%)	200 (20)	400 (40)	30 (3)	60 (6)
相対リスク減少 RRR	$(0.4-0.2) / 0.4 = 0.5$		$(0.06-0.03) / 0.06 = 0.5$	
RRR (worst case scenario 法)	$(0.4-0.23) / 0.4 = 0.43$		$0.00 / 0.06 = 0$	

試験 A と試験 B はいずれも、介入群 1,000 人、対照群 1,000 人の RCT で、いずれも追跡からの脱落が 30 人 (3%) である。試験 A では、介入群 200 人死亡、対照群 400 人死亡で、相対リスク減少 (RRR) は 50% である。ここで、脱落した 30 人がどうだったかにより、この RRR は大きく変化する。worst case scenario 法で考えると、つまり治療群での脱落例のすべてが死亡したとすると、230 例 (23%) である。もし、対照群での脱落例はすべて生存し、介入群での脱落例がすべて死亡とすると、死亡の相対リスク減少は $(0.4-0.2) / 0.4$ (50%) から、 $(0.4-0.23) / 0.4$ (43%) に低下する。このように、最悪のケースを考えても、介入効果には大きな変化はなく (50% 対 43%)、われわれの確信は変わらない。

試験 B では、同じく相対リスク減少は 50% であるが、死亡イベント数は少なく、介入群で 30 例、対照群で 60 例である。試験 A と同じように worst case scenario 法で検討してみる。すなわち、介入群での脱落例のすべてが死亡し、対照群のそれはゼロであったとすると、死亡例数は両群ともに 60 例となる。もし、この推定が正しければ統計学的に治療効果はないことになり、効果推定値の確信をおびやかすものである。つまり、脱落率 3% でも、信頼できないことになる。

^f 追跡からの脱落などで、データが欠損になっている場合の対応の一つ。