

①研究の結果は妥当か？（内的妥当性の検証）

1. 臨床疑問を定式化できるか？	
何がベストか？	どこを読めばよいか？
研究で検証しようとする臨床疑問は、PICO（P：患者、I：介入、C：比較対象、O：アウトカム）の形に定式化できる必要がある。	「BACKGROUND」を読むと、その論文で検証しようとする臨床疑問が生まれるに至った経緯が記述されており、最後に検証すべき臨床疑問が簡潔に述べられている。
2. 治療の割り付けはランダム化されているか？	
何がベストか？	どこを読めばよいか？
中央でのコンピューターによるランダム化が理想である。多施設研究ではこの方法が多く用いられる。小さな研究では、自施設の第三者（薬剤師など）がランダム化をコントロールすることもある。	「METHOD」を読むと、患者がどのように割り付けられ、ランダム化が中央で行われたか、ローカルで行われたかが記述されているはずである。
3. 治療の割り付けは隠匿化されているか？	
何がベストか？	どこを読めばよいか？
現場で組み入れや除外を判断する研究者が、ランダム化の順序を予め知ることができると、選択バイアスが発生する。中央でランダム化されている場合では、隠匿化は適切である。	「METHOD」を読むと、ランダム化の方法が記述されており、その方法によって隠匿化が適切かどうかを知ることができる。
4. 評価者は治療内容について適切にマスキングされているか？	
何がベストか？	どこを読めばよいか？
「二重盲検化」されていること、すなわち患者自身も研究者も、患者がどちらの群に割り付けられているか分からない状態が理想である。アウトカムが客観的（死亡など）であれば、マスキングの重要性はやや低くなる。アウトカムが主観的（症状、機能など）であれば、評価を行う者のマスキングの重要性は極めて高くなる。	まず「METHODS」を読み、マスキングへの言及を探す。偽薬や sham therapy についての言及である。次に、「METHODS」ではどのような方法でアウトカムが評価されたかが述べられており、評価者が治療割り付けについて知ることができるのか否かについて述べられている。
5. 介入以外の治療内容に差はないか？	
何がベストか？	どこを読めばよいか？
検証の対象となっている治療以外は、検査や他の治療、フォローアップを含め、両群とも同じように治療されているのが理想である。	「METHODS」には、フォローアップの計画やその他の治療内容について言及されており、「RESULTS」には実際にどのような治療が行われたかが書かれている。
6. アウトカムは客観的で、かつ臨床的に意味があるか？	
何がベストか？	どこを読めばよいか？
「死亡」など、研究者の主観によって変化せず、かつ臨床的にも重要と思われるアウトカムが選択されていることが理想である。アウトカムの発生率が小さい場合は、複数のアウトカムを合わせて「Composite outcome」としている場合もある。研究の主要アウトカム=Primary outcome は、1つに限定されるのが理想である。	「METHODS」には、その研究の Primary outcome (Primary endpoint) や Secondary outcome (s) (Secondary endpoint (s)) が記述されている。RCT は一般的に、Primary (composite) outcome のみを正しく評価するために設計されており、Secondary outcome や Post-hoc 解析で得られた結果は、未来の研究のための仮説創出という意味しかない。
7. サンプルサイズは適切に設定されているか？	
何がベストか？	どこを読めばよいか？
類似の先行研究、あるいは自施設でのデータなどから、主要アウトカムのイベント発生率や検証しようとする介入による効果を推定し、それに基づいて必要なサンプルサイズを計算するの	「METHODS」には、統計解析についての項目があり、先行研究、イベント発生率、効果量について言及されている。その論文の $\alpha$ エラー、 $\beta$ エラーについてもここで記述されている。

が理想である。先行研究が無い場合は、サンプルサイズは根拠なく設定されることもある。	
8. 治療開始時に両群の背景因子に差はないか？	
何がベストか？	どこを読めばよいか？
ランダム化が上手くできていれば、両群の背景因子に大きな差は出ない。群間差は小さいほど望ましい。p 値など、群間差を客観的に示す何らかの指標が記述されることがある（最近では記述しないことが多くなった）。確実に調整したい背景因子がある場合は、その因子について層別化が行われる。	「RESULTS」には、結果に影響を与えると思われる患者背景を群間で比較した表が提示されているはずである（年齢、リスク因子など）。表がない場合は、両群で差がないことに本文中で言及されていることもある。
9. 組み入れられた全患者について結果が記述されているか？	
何がベストか？	どこを読めばよいか？
追跡不能例は少ないほどよく、20%未満が望ましい。アウトカムの発生頻度が少ない場合は、わずかな追跡不能例でも結果にバイアスを与える。	「RESULTS」には、何人の患者がランダム化され、何人の患者が解析に含まれるのかが「Flow diagram」の形で書いてあることが多い。追跡不能例については、その人数と理由が明確化されている必要がある。
10. 患者が組み入れられた群で最終的な解析が行われているか？	
何がベストか？	どこを読めばよいか？
患者は「組み入れられた群」で解析される必要があり、これを「Intention to treat (ITT) 解析」とよぶ。患者を「結果的に受けた治療」に基づいて解析する場合もあり、これを「Per-protocol 解析」とよぶ。	「METHODS」に、その研究で用いられた解析の方法が書かれている。

②結果はどのようなものか？

11. 治療効果はどの程度大きいか？	
多くの場合、研究結果は二値アウトカムで記述される（アウトカムが起こるか起こらないか）。癌の再発の有無、心筋梗塞発症の有無、死亡の有無、などである。例えば、コントロール群の15%が死亡し、介入群の10%が死亡した研究があったとする。この研究の結果は、以下に挙げるような効果量 (Effect size) として記述される。	
効果量の種類は？	その意味は？
<p><u>Relative Risk (RR)</u></p> <p>介入群のリスク (イベント発生率) / コントロール群のリスク (イベント発生率)</p> <p>上記の研究では、 RR = 0.10/0.15 = 0.67</p>	<p>相対リスク比 (RR) はコントロール群に比べて介入群でイベントが何倍発生し易いのかを表している RR が1であれば、両群に差はなく、したがってその介入は効果がないということになる。RR が1未満であれば、アウトカムのリスク (イベント発生率) は減少し、RR が1より大きければリスクは増大する。</p> <p>上記の研究では、治療により介入群での死亡リスクがコントロール群の67%に減少した。</p>
<p><u>Absolute Risk Reduction (ARR)</u></p> <p>{コントロール群のリスク (イベント発生率)} - {介入群のリスク (イベント発生率)}</p> <p>絶対リスク差とも言う。</p> <p>上記の研究では、</p>	<p>絶対リスク減少は、二群間でのイベント発生率の差の絶対値を表す。ARR が0であれば、両群間のイベント発生率に差はなく、したがってその介入は効果がないということになる。</p> <p>上記の研究では、治療により死亡率が5%減少した。</p>

ARR = 0.15 - 0.10 = 0.05 (5%)	
<p><u>Relative Risk Reduction (RRR)</u> 絶対リスク減少／コントロール群のリスク (イベント発生率)</p> <p>1 から RR を引くことでも計算できる。 (RRR = 1 - RR)</p> <p>上記の研究では、 RRR = 0.05/0.15 = 0.33 (33%) または RRR = 1 - 0.67 = 0.33 (33%)</p>	<p>相対リスク減少は、相対リスク比を補完する数字であり、最もよく報告される治療効果の指標である。コントロール群でのイベント発生率に比し、どのくらいのイベントが介入群では減少したかを示している。</p> <p>上記の研究では、治療により介入群でのイベント発生率がコントロール群の 33%減少した。</p>
<p><u>Number Needed to Treat (NNT)</u> 絶対リスク減少の逆数 (1/ARR)</p> <p>上記の研究では NNT = 1/ 0.05 = 20</p>	<p>NNT は、1 つの負のアウトカムを防ぐために、その治療を施さねばならない患者数であり、これには治療の期間も含まれる。アウトカムの臨床的な重要性は、NNT を見て判断されることが多いが、同じ治療の負の側面 (NNH : Number Needed to Harm、1 つの負のアウトカムを発生させるために必要な治療患者数) とのバランスでも判断される。</p> <p>上記の研究では、20 人を (例えば 2 年間) 治療して、1 人の患者の死亡を防ぐことができる。</p>

12. 得られた治療効果ほどの程度正確か？

対象とした患者母集団での「真理としての」アウトカムの発生率を知ることが不可能であり、それを類推する最も良い方法は、その研究に組み入れられた患者のデータから、母集団でのリスクを推定することである。これを「Point estimate=点推定」という。データからの推定値が真理にどの程度近いかは、推定値の「Confidence interval (CI)」から測ることができる。CI が非常に狭ければ、今回得られた推定値は母集団を反映していると言える。CI は、得られた結果の統計学的な有意性についても、「効果がない」とされる値 (RR なら 1、ARR や RRR なら 0) が CI に含まれていなければ、その結果は 0.05 のレベルで統計学的に有意だと言える。逆に「効果がない」とされる値が CI に含まれるならば、その結果は統計学的には有意ではないということになる。

13. 得られた研究結果ほどの程度脆弱か？

仮に、主要アウトカムが起こった患者のうち「何人に主要アウトカムが起こらなかつたら、その研究の統計学的有意差は消失していたか？」と考えるとき、この人数を「Fragility index」とよぶ。Fragility index は、その研究データの脆弱性を表す指標であり、Fragility index よりも追跡不能例が多い研究では、例え結果に有意差があったとしても、αエラーの可能性が高くなる。

③自分の患者の診療に役立つ結果か？ (外的妥当性、応用可能性)

自分の患者に結果を当てはめる前に、以下のことを考える

- 14. 研究対象の患者と自分の患者の背景、特徴が大きく異なっていないか？
- 15. 自分の診療の場において、実行可能性のある治療か？
- 16. 自分の患者において、その治療の益は害を上回っているか？

Produced from “Critical appraisal sheet” by CEBM

Translated and modified by Takashi Matsumoto.

Downloaded from <https://www.cebm.net/2014/06/critical-appraisal/>.

Last accessed on April 22<sup>nd</sup>, 2018.