

# 学位論文内容要旨

北里大学大学院 薬学研究科 臨床医学（医薬開発学）

氏名：野口 泰

## 【題目】

Scientific influence of drug-intervention clinical trials: a bibliometric analysis based on the Relative Citation Ratio

（薬剤介入臨床試験の科学的影響度に関する研究：相対引用率を指標とした計量書誌学的分析）

## 【背景・目的】

EBM (evidence based medicine) の進展には臨床試験が不可欠であり、毎年、膨大な数の臨床試験が実施され、その数は年々増加している。臨床試験のデザインは「エビデンスレベル」を定義する1つの要素であり、その試験結果の解釈や有用性に影響を与える。一方、臨床試験のデザインと試験結果の科学的影響度との関係を検討した研究はない。

計量書誌学は、公表された研究論文の情報を定量的に分析する学問である。これまで、研究者や研究機関の生産性を推定するために、Journal Impact Factor (JIF) や Hirsch-index などの指標が開発されてきたが、これらの指標は個々の論文に対する評価ではない。2015 年、米国 National Institutes of Health (NIH) は、論文毎の科学的影響度を計る指標として、相対引用率 (Relative Citation Ratio: RCR) を開発した。RCR は、論文の1年あたりの被引用数を、期待被引用数で除して求めることができ、以下の特徴を有する。

- i. 共引用ネットワークと呼ばれる、評価対象の論文を引用する論文が引用したすべての論文を用いて標的論文の学際領域を定義し、異なる研究領域を比較可能である。
- ii. 計算アルゴリズムが開示されており、無料でインデックスが利用可能である。
- iii. NIH-Research Project Grant Program R01 (NIH-R01) が資金提供した論文がベンチマークされており、RCR = 1.0 は、その年の当該学際領域における NIH-R01 が資金提供した出版物の中央値となる。

RCR は、当初 NIH が資金を提供する研究を評価するために開発されたが、NIH 以外でも文献分析に活用されつつある。

非小細胞肺癌 (non-small cell lung cancer: NSCLC) は、世界で最も患者数の多い癌の一つであり、癌による死亡の主因である。近年、免疫チェックポイント阻害薬といった新たな分子標的薬の開発により生存期間は延長されつつあるが、進行・転移性の NSCLC に対する治療選択肢はまだ限られている。

本研究では、臨床試験結果の科学的影響度に影響を及ぼす因子を特定するため、研究1及び研究2では、2007 年～2016 年に出版された NSCLC 患者を対象とした薬剤介入試験の論文を対象とし、試験デザインと RCR の関係を検討した。研究3では、NSCLC の治療を変革した3薬剤 (EGFR チロシンキナーゼ阻害薬: Erlotinib、ALK/ROS1 阻害薬: Crizotinib、抗 PD-1 抗体: Nivolumab) の臨床試験論文を対象とし、試験デザイン及び薬剤の開発経緯と RCR の関係を検討した。

## 【方法】

疾患領域の違いによる臨床試験デザインの差異を最小限にするため、アンメットメディカルニーズが高く臨床試験が盛んに実施され、EBM が定着しており、近年革新的な新薬が登場している NSCLC を研究対象疾患とした。

[研究1：全ての薬剤介入試験を対象とした研究]

研究開始時点で恒常的な RCR が算出されていた 2007 年～2016 年に公表された NSCLC の臨床試験論文を抽出した。PubMed をデータソースとし、検索用語は「non-small cell lung cancer」、対象種「human」及び言語区分「English」とし、論文種別は「Clinical trial」、「Phase 1~4」又は「randomized controlled trial」とした。RCR 及び NIH%ile (NIH 論文と比較した際の RCR の percentile 値) は、NIH が提供

する「iCite」よりダウンロードした。検索結果が 3100 件を超えたことを考慮し、(1) LOW50、(2) 50NIH%ile、(3) 95NIH%ile 及び (4) TOP50 の 4 カテゴリーを設定し、前向き薬剤介入臨床試験の主論文を上記カテゴリー毎に 50 報ずつ抽出した。NSCLC 以外の患者や薬剤以外の介入が含まれる論文は除外した。また、可能な場合、該当する臨床試験を臨床試験レジストリ（ClinicalTrials.gov 等）で特定し情報源とした。論文又はレジストリから、試験相、ランダム化、盲検化、営利企業のスポンサー又はサポートの有無、試験施設、実施国、登録被験者数、被験者の登録開始日、介入、責任著者等の情報を収集した。

#### [研究 2：ランダム化比較試験を対象とした研究]

ランダム化比較試験はエビデンスレベルピラミッドの上位に位置し、研究 1 の結果からも科学的影響度を高める因子の一つと示唆されたため、研究 1 で収集した 2007 年～2016 年の前向き薬剤介入試験論文から、ランダム化比較試験を固定因子とし RCR に影響する因子を調査した。

##### <研究 2-1：RCR カテゴリー間の比較>

研究 1 と同様の手法で、LOW50、50NIH%ile 及び TOP50 の 3 カテゴリーを設定し、各因子の割合を比較した。

##### <研究 2-2：Journal of Clinical Oncology (JCO) を対象とした研究>

研究 1 において RCR が JIF と相関する傾向が示唆されたため、論文掲載誌を固定因子とした検討を行った。研究 1 で最も論文数が多く 50NIH%ile から TOP50 の 3 カテゴリーに分布しており、研究 2-1 で最も論文数が多い雑誌の一つであった JCO を対象とした。研究 1 のデータベースより、JCO に掲載されたランダム化比較試験を全て抽出し、臨床試験デザインの因子毎に RCR の違いを比較した。

#### [研究 3：NSCLC の代表的治療薬に着目した事例調査]

薬剤の開発経緯と試験結果の科学的影響度の関係を検討するため、研究 1 で TOP50 に存在し、NSCLC の治療を変革した 3 薬剤（Erlotinib、Crizotinib、Nivolumab）を対象に、臨床試験結果と科学的影響度を調査した。研究 1 と同様の手法で、各薬剤の一般名、商品名及び開発コードを用いて 3 薬剤の臨床試験論文を抽出した。開発経緯に主眼を置くため、論文公表年に制限を設けず、RCR が算出されている限り最新の論文も調査対象とした。また、各薬剤の評価が主体となった臨床試験の結果を選択し、対照薬や新規薬剤を上乗せする標準治療として用いられた試験は除外した。さらに、論文から判明した臨床試験登録番号等を用い、主論文と副次論文（追加解析、フォローアップ、併合解析等）の紐付け及び追加検索を行った。

各研究における統計学的な解析として、カテゴリー間の各因子の割合の差を評価するために Fisher の正確確率検定を、カテゴリー間の症例数の差及び臨床試験デザイン因子毎の RCR の差を評価するために Mann-Whitney の U 検定を行った。

#### 【結果】

##### [研究 1：全ての薬剤介入試験を対象とした研究]

LOW50、50NIH%ile、95NIH%ile 及び TOP50 の RCR の中央値は、それぞれ 0.03、1.00、5.76 及び 26.89 であった。

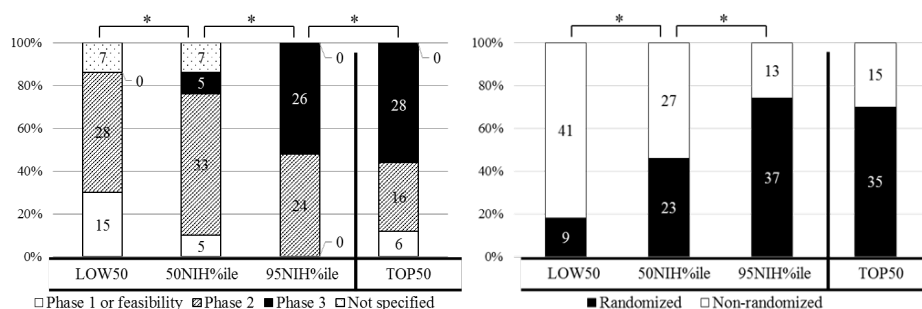


図 1. RCR カテゴリー毎の試験相（左）及びランダム化比較試験（右）の割合

4 カテゴリーに含まれる臨床試験デザインを比較した結果、第3相試験、無作為化、盲検化、複数施設／複数国での試験、営利企業スポンサー／サポート試験、及び被験者数の多い試験の結果は、より高い科学的影響を有することが示された。

また、EGFR チロシンキナーゼ阻害薬、ALK 阻害薬、免疫チェックポイント阻害薬など、新たに開発された分子標的薬の臨床試験の論文が、従来の化学療法よりも高い科学的影響度を有することが示唆された。

#### [研究2：ランダム化比較試験を対象とした研究]

##### <研究2-1：RCR カテゴリー間の比較>

LOW50、50NIH%ile 及び TOP50 の RCR の中央値は、それぞれ 0.14、1.01 及び 19.01 であった。3 カテゴリーに含まれる臨床試験デザインを比較した結果、研究1と同様の因子（第3相試験、盲検化、複数施設／複数国での試験、営利企業スポンサー／サポート試験、被験者数の多い試験、及び分子標的薬の介入）に加え、生存や疾患進行を主要評価項目に用いた試験、対照群との差の検出力が設定された試験（検出力を設定した試験）の結果が、より高い科学的影響を有する傾向が示された。TOP50 で 50NIH%ile と比較し特に多い因子（30%≦）は、第3相試験（84%）、複数国で実施した試験（78%）、営利企業スポンサー／サポート試験（92%）、及び検出力を設定した試験（98.0%）であり、LOW50 で 50NIH%ile と比較し特に少ない因子（≦30%）は、生存や疾患進行を主要評価項目に用いた試験（26%）、及び検出力を設定した試験（22%）であった。

##### <研究2-2：Journal of Clinical Oncology（JCO）を対象とした研究>

ランダム化比較試験は 89 試験存在し、RCR の中央値は 4.35（0.76－60.13）であった。試験デザインの因子毎に RCR を比較した結果、第3相試験、分子標的薬を介入に含む試験、及び検出力を設定した試験の RCR が高く（ $P<0.05$ ）、他の項目で有意な差はなかった。

#### [研究3：NSCLC の代表的治療薬に着目した事例調査]

Erlotinib 209 報/121 試験、Crizotinib 30 報/9 試験及び Nivolumab 32 報/17 試験を抽出した。十分なデータを収集できた Erlotinib の主論文を対象に臨床試験デザインと RCR の関係を解析した結果、第3相試験、ランダム化比較試験、盲検化、複数国での試験、営利企業スポンサー／サポート試験で RCR が高く、研究1及び2の結果を再現した。121 試験中 30 試験で副次論文が存在し（3.9 論文/試験）、それら主論文の RCR（中央値=2.88）は副次論文の無い試験の論文（中央値=0.63）と比較し高かった。主論文の RCR が高い（20≦）試験は 5 試験存在し、いずれも適応症や使用法の根拠データとして日米欧いずれかの添付文書又は治療ガイドラインで引用されていた。RCR の低い（≦0.2）試験は 12 試験存在し、探索的な第1相又は2相試験であった。また、12 試験全てが1カ国での実施であり、うち6試験は日本で実施されていた。121 試験中、単一国で実施した試験は日本が最多（28 試験、23.1%）である一方、日本の Weighted RCR（RCR の総和）は Global、中国、米国試験に次いで4番目であった。

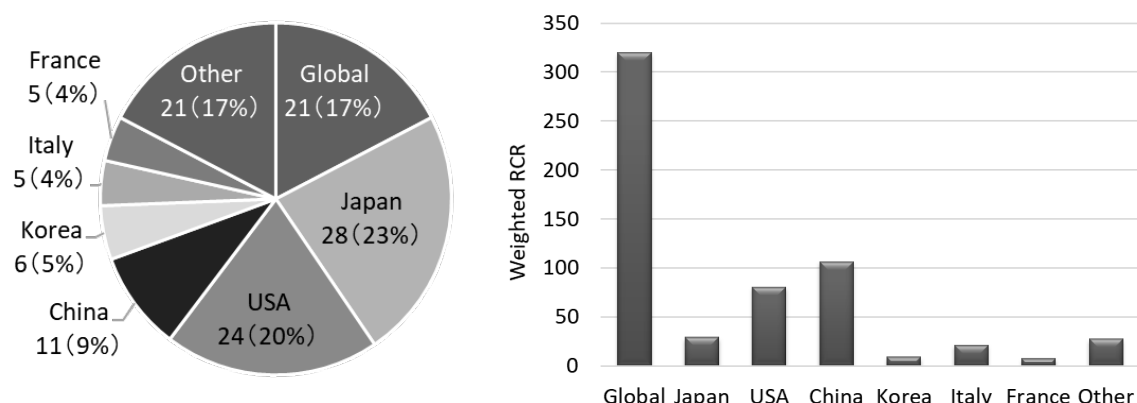


図 2. Erlotinib の臨床試験実施国の割合（左）及び実施国毎の Weighted RCR（右）

Crizotinib 及び Nivolumab でも同様に、RCR の高い試験の大部分（9 試験中 7 試験）は、添付文書等で適応症や使用法の根拠となる試験であった。

#### 【考察】

研究 1 及び 2 の結果から、第 3 相、ランダム化比較試験、盲検化試験、多施設及び国際共同試験、営利企業のスポンサー又はサポートのある試験、症例数の多い試験の結果が、科学的影響度が大きいことが示された。また、ランダム化比較試験のうち、対照群との差の検出力が設定された試験結果が高い科学的影響度を有する傾向が示された。これは、高いエビデンスレベルを有するようにデザインされ、統計学的に結論付けられる検証試験が、結果の科学的影響度を向上させることを意味するものである。さらに、新たに開発された免疫チェックポイント阻害薬や ALK 阻害薬などの分子標的薬の臨床試験の結果は、従来の化学療法よりも科学的影響が大きく、第 1 相試験や小規模な試験でも高い科学的影響度を有することが示された。この理由として、長く化学療法が主体であった NSCLC の治療において新規分子標的薬の注目度が高いこと、過去に例を見ない画期的有効性を示した薬剤であること、さらに抗がん剤開発の加速化に伴い開発初期の臨床データの重要性が増していること等が要因と考えられる。

研究 3 の事例調査は、研究 1 及び 2 の結果が支持されたことに加え、適応症取得の根拠となる検証的試験の科学的影響度が高いことが確認された。また、Crizotinib 及び Nivolumab において、薬剤の高い有効性と安全性を示す第 1 相試験の結果も高い RCR を有することが確認され、画期的新薬であれば探索的試験でも高い科学的影響度を有することが裏付けられた。

一方、研究 1 及び 2 の結果から、探索的で小規模な試験や、分子標的薬を含まない試験の科学的影響度が低いことが示された。これら試験は検証的試験の予備的検討や最適なレジメンの判断が主目的であることが多く、検証的試験と比較して結果の利用範囲が限定されることが要因の一つと考えられる。また、介入に分子標的薬を含まない場合、含む場合と比較して新規性が乏しいことも原因と考えられる。研究 3 の Erlotinib 事例調査においても上記を支持する結果が得られた。Erlotinib では、日本単一で実施した試験数が最多である一方、Weighted RCR は実施国別で 4 番目であることが判明した。研究 1 の LOW50 においても 22 試験（44%）が日本単一で実施されており、大部分は小規模な非ランダム化試験であった。ランダム化比較試験の検討では同様の傾向は見られていないことから（研究 2-1；日本単独試験：LOW50：7 試験、TOP50：4 試験）、小規模な非ランダム化比較試験を減らし、それら資源を大規模な検証試験の実施に振り分けることが、日本で実施する試験の科学的影響度を高める一助になると考えられる。

#### 【総括】

規模が大きく統計学的によくデザインされ、新規薬剤の介入を含む試験を実施することが結果の科学的影響度を高める。そのために、科学的影響度の低くなる小規模で探索的な臨床試験は最小限に留め、より大規模で検証的な臨床試験に研究資金や人的資源を振り分け、科学的影響度の高い試験結果を効率的に創出することが重要となる。製薬企業による新規薬剤の開発と大規模な臨床試験の主導やアカデミアのリソース（資金、薬剤等）面のサポート、及び癌臨床試験グループの活用等、臨床試験の実施環境の整備もその一助になると考える。

以上