

向精神薬の新規出口戦略の妥当性検証：観察研究を使った方法論

研究分担者 渡辺 範雄 医療法人社団蘇生会 蘇生会総合病院

研究要旨

向精神薬の一定時間継続投与後に中止するのが妥当かどうか検討するためには、中止か継続かをランダム化比較試験にて行うことが望ましい。しかし実際は長期的かつ大規模な試験が必要になることから、これを実施できないことも多い。妥当な方法論により、観察研究や実際の診療データからこれを見積もることができれば、無作為割り付け対象試験のない介入領域における診療に資するものは大きい。本分担研究では、これを行う方法論の一つとして傾向スコアマッチングに焦点を当てて情報収集することで精神医学領域における研究例をまとめ、本研究での利用方法について考察した。

A. 研究目的

無作為割り付け対照試験、あるいはランダム化比較試験（randomized controlled study: RCT）は治療や予防の介入の効果を見るにあたり、もっとも確からしさが強いエビデンスを提供するとされる。RCTは、治療や予防介入の選択に関して研究者が人為的に介入して無作為の割り付けを行うため、「介入研究」の一部とされる。介入という言葉が、治療や予防のための医療行為の、選択と実施の両方に使われているために紛らわしいが「介入研究」の「介入」はこの前者を指す。無作為割り付けあるいはランダム化を伴わなくても、研究者の意思で治療・予防行為の選択が行われれば「介入研究」と呼ぶ。

一方、通常の治療の中で医療者・患者で話し合っただけで治療・予防のための医療的介入を選択し、それを実施してその結果を見る研究であれば介入研究とは言わず、「観察研究」と言う。似たような言葉で「疫学研究」というものがある。疫学研究とは「地域社会や特定の間人集団を対象として、健康に関する事象の頻度や分布を調査し、その要因を明らかにする医学研究」と定義される。疫学研究というのは大きな概念で、その中に介入研究と観察研究を包含するというのが一般的な定義である。

介入研究、特にRCTは既知・未知を問わず交絡因子を無作為割り付けにより（これがうまくいけば）治療・予防介入群間に均等に分布させることができるため、エビデンスの強度としては高いとされる。逆に観察研究はこれができないため、強度は低いとされる。

向精神薬を一定時間継続投与後に中止するという行為は、治療介入にあたる。そのため、中止によって起こりうる再発や離脱といった問題が、継続した場合と比較してどのようになるのか、ランダム化比較試験にて該当患者を中止か継続かの群で検証することが望ましい。しかし実際は、再発のアウトカムを見るためには長期的な経過観察が必要であり、また急性期治療と比較すると相対的に両群間の差がつきにくいことから、症例数の大きい試験が必要になると考えられる。そのため、十分な科学的妥当性をそなえたRCTは実現可能性に問題がある。

しかし、妥当な方法論により、観察研究や実際の診療データからこれを見積もることができれば、無作為割り付け対象試験のない介入領域における診療に資するものは大きい。本分担研究では、これを行う方法論の一つとして傾向スコアマッチングに焦点を当ててその方法論を情報収集することで精神医学領域における研究例をまとめ、本研究での利用方法について考察した

B. 研究方法

1. 情報収集

本分担研究では、傾向スコアマッチングに関する情報源として、国内外の代表的参考書をもとに下記をまとめた。

- a. RCTの強みおよび弱点
- b. 観察研究における疑似ランダム化
- c. 傾向スコアマッチング
- d. 傾向スコア分析の方法と問題点

またMedlineにより精神医学領域における研究例を収集した。

2. 本総合研究への適用についての検討

本総合研究では、精神科領域、また薬物治療における中止がテーマであり、他領域、また薬物療法以外の治療とは異なる配慮が予想される。幸い、薬物療法はレセプトデータに反映されやすいため、レセプトデータを利用した国内のいわゆるビッグデータが利用可能な可能性があるため、これらを加味して導入可能性を検討した。

倫理面への配慮

本分担研究は先行研究や公表されている方法論のレビューであり、新たに個人データを扱ったものではないため、倫理的問題は発生しない。

C. 研究結果

参考文献1-6をもとにまとめた。

a. RCTの強みおよび弱点

一般診療や観察研究では、年齢や性別、心理社会的因子などの患者の背景因子や、医療者の経験や医療設備など治療・予防を施す施設の要因は、治療・予防効果に影響を与えるだけでなく、治療・予防介入の選択そのものにも影響を及ぼす。そのため、群別の治療予防効果は、治療・予防介入そのものの結果なのか、患者の背景因子や施設要因によるものであるか分からない。この場合、患者の背景因子や施設要因を交絡因子と呼ぶ。RCTの最大の強みは、交絡因子を治療・予防介入群間

に均等に分布させることができるため、交絡を分析や研究結果の解釈において無視できることである。

新薬や新しい医療技術であれば、RCTを行って標準治療もしくは既存の治療と比較することは、比較的容易である。しかしすでに普及している薬や医療技術に関して、RCTを実施することは必要であるにも関わらず、研究資金や人的資源の点で実現性は低いと言わざるを得ない。特に介入群間の効果の差が少ない場合は差を検出するために多くの参加者が必要であるし、長期的な主作用および副作用を見るためには長時間の追跡を要するため、実現性が下がる。また治療・予防介入とはいえ妊婦・若年者・高齢者における向精神薬の投与結果をRCTで見るとは、倫理的に難しい。飲酒や喫煙・生活習慣などの暴露要因がその後の健康アウトカムに及ぼす結果も同様である。

また介入研究では、患者を選択したのち研究の説明をして同意を得られたもののみデータを取得でき、特にRCTではランダム割り付けに同意した患者のみ研究参加者となりうる。これでは研究結果を一般臨床に適用するにあたり、研究参加者が一般臨床の代表性を担保しているのか疑問に思う考え方もあるだろう。

b. 観察研究における疑似ランダム化

観察研究でも、RCTと同じ水準とは言わなくとも、複数の治療・予防介入群の背景がなるべく均質な状況を作りだすことができれば、一般臨床でも役に立つ結果が得られそうである。観察研究のデータからこれを行うことを、疑似ランダム化という。

疑似ランダム化の中でも代表的なものが、マッチングである。観察研究では条件を満たせば研究の事前説明や参加者同意もなくオプト・アウトでデータを取得することができるし、また説明や同意が必要な場合でもRCTよりも比較的容易であるため、多くの参加者のデータが得られることができる。もしも十分多くの参加者のデータが得られれば、各治療・予防介入を受けたの群の中から一対ずつ、背景因子が「ほどほど」に似た参加者の組み合わせを選ぶことが可能となる。ただし一対の全く同じ背景の参加者二人を選ぶことは不可能

なので、「ほどほど」加減を研究者が選択することになる。研究で見られるのが、既にあるデータから年齢や性別、重症度が似通ったペアを集めてくるものである。年齢はプラスマイナス何歳まで、などと許容範囲を決めておけば、なんとか研究者が自分の目視で行うことが可能であろう。これをマッチド・ペア・コホート研究という。

しかし、治療・予防介入の群間差に影響を及ぼしそうな背景要因は、これだけではない。精神科では身体合併症、現在までの疾患有病期間や治療歴、学歴、家族・社会的サポート、家族歴、医療機関の種類、特に薬物療法以外の複雑介入では介入実施者の経験など、多くの考慮されるべき背景要因があるが、これらが無視されることになる。統計解析の段階である程度調整はできるにしても、多くの要因を同時に考慮した疑似ランダム化を行うことができれば、RCT に近づくことが可能となる。

c. 傾向スコアマッチング

多くの背景要因を考慮したうえで疑似ランダム化を行うことを可能にする方法として、傾向スコアマッチングがある。傾向スコアとは、二つの群間比較において、各参加者がいる一群に割り当てられる確率を示す。この数字は各参加者の背景因子を独立変数、実際に受けた治療・予防介入の群を目的変数として、ロジスティック回帰分析(他にプロビット回帰分析などで行うこともある)のモデルに投入することで計算する。確率なので、傾向スコアは0から1の間のだこかの数字を取るようになる。そして治療・予防介入の異なる群において一対ずつ、傾向スコアがほぼ等しい患者のペアを順に選んでくる。これを行えば、両群間での患者のほとんどの投入した背景因子はほぼ均質化されるはずである。これを傾向スコアマッチングと言い、疑似ランダム化を行う一つの手段である。そして両群間で、各治療・予防介入を行ったあとでのアウトカムを比較する。

ただし、通常の変数解析にベースラインの交絡因子、特に傾向スコアの計算に用いた変数を直接投入して調整することのほうが、モデルに投入する人数が増えるのでよいのではないかと、という疑問が生じるかもしれない。しかし、傾向スコアを用いるほうがベースラインの違いをよりよく

調整できる場合が多い。多変数解析では同時に多くの背景因子を投入することができないし、また特にアウトカムのイベントが稀である場合には、サンプルサイズが足りなくてモデルに投入できる因子数は非常に限られる。傾向スコアを用いれば、多くの背景因子を投入できる。また個々のベースラインの交絡因子とアウトカムの間は何の仮定も持ち込む必要がないため、アウトカムの統計解析もカイ2乗検定やt検定などの単純な解析を行えばよいことになる。

d. 傾向スコア分析の方法と問題点

各参加者に傾向スコアを計算したのちには、計算したスコアが2つの治療をうまく識別できているか、マッチングを行う前に確認する。治療の割り当て変数を状態変数(1 または 0)として、傾向スコアを検定変数とする ROC (receiver operating characteristic) 曲線を描き、曲線化面積を求める。この面積を c 統計量と呼ぶ。c 統計量は 0.5 以上 1.0 以下となるが、これが 0.5 のときは識別力なし、1.0 の時は完全識別と呼ぶ。

c 統計量は低すぎると、マッチングできる参加者数が減ってしまって通常の変数解析とあまり変わらない結果になる。逆に高すぎると治療群間のオーバーラップが少なく、やはり参加者数が減ってしまって統計的な検出力が低下する。これらの場合、このモデルでの傾向スコアマッチングは向かないということになる。

次の過程は、マッチングである。2つの治療群から傾向スコアが近接しているペアを抽出していくが、これを最近傍マッチングという。最近傍マッチングでは、ペアとして抽出される参加者二人の傾向スコアの差の絶対値が一定の閾値以内に収まるようにする。この閾値をキャリパーと言い、全対象患者の傾向スコアの標準偏差の 0.2 倍に設定されることが多い。

その後、マッチング後の集団で両群の患者背景がうまくバランスが取れているか確認する。この目的のためには、標準化差 (standardized difference) というものを計算する。標準化差 ≥ 0.1 の場合、バランスが取れていないと判断され、傾向スコアを計算するロジスティック回帰モデルを修正する必要がある。

そして最後に、両群間のアウトカムを比較し、

二値変数であればカイ 2 乗検定、連続変数であれば t 検定などを行う。また生存分析を行う場合もある。

ただし、傾向スコアは既知の背景因子のみをモデルに投入した分析であるため、未測定の変数となりうる背景因子が存在する可能性がある。そのため、どうやっても RCT の結果と同様の頑健性は得られない。これを対処するために、未測定の変数因子の存在を仮定してその影響量の大きさを変えていってモデルに投入して、感度分析を行う。これによって、結果の信頼区間の変化を見ることで、主要分析の結果の頑健性を検証する。

傾向スコアを使った群間のアウトカム比較には、マッチングの他にも傾向スコアの逆確率による重みづけ、傾向スコアによる調整がある。ただしマッチングが一番理解しやすく、汎用されているためこの項ではマッチングに関してのみ言及した。

傾向スコアを使った観察研究の実例

研究例 1

統合失調症治療をクロザピンで開始するか他の通常抗精神病薬で開始するか(参考文献 4)

通常臨床現場において、統合失調症の初期薬物治療をクロザピンで開始するか他の通常抗精神病薬で開始するか、その後のアウトカムを比較するために 2001 年から 2009 年の米国 Medicaid のデータが用いられた。クロザピンで開始した 3123 人と、傾向スコアマッチングで抽出した他の通常抗精神病薬で開始した 3123 人を比較し、精神科入院を主要アウトカム、開始抗精神病薬の中止、抗精神病薬の追加、重要な医学的状態の発生、死亡を副次アウトカムとして比較した。

結果、クロザピンで開始することは他の通常抗精神病薬よりも精神科入院が統計学的有意に少なく(ハザード比 0.78, 95% CI=0.69-0.88), 副次アウトカムでも開始抗精神病薬の中止(0.60, 95% CI=0.55-0.65) 抗精神病薬の追加(0.76, 95% CI=0.70-0.82)において有意に少なかった。しかしクロザピンは糖尿病の発生(クロザピン 2.8% vs. 通常抗精神病薬 1.4%; ハザード比 1.63, 95% CI=0.98-2.70)をはじめ、高脂血症、小腸閉塞が有意に多かった。

研究例 2

妊娠中の抗精神病薬使用と母体・周産期アウトカム(参考文献 5)

妊娠中に抗精神病薬が安全に使用できるかどうかを見るためには、倫理的に RCT を行うことはできない。そこで高次(high dimensional)傾向スコアマッチングを行い、カナダのオンタリオ州の健康管理データベースにおいて妊娠初期または中期に抗精神病薬を 2 回以上処方された妊婦 1021 人と、そうでない妊婦 1021 人を抽出し、妊娠糖尿病、高血圧、静脈血栓症といった母体側アウトカムと、早産、体重が 3 パーセント未満または 97 パーセントの出生体重といった児側アウトカムを条件付き Poisson 回帰分析で比較した。

結果として、抗精神病薬を使用した妊婦はそうでない妊婦と比較して、妊娠糖尿病(7.0% vs 6.1%, rate ratio 1.10, 95%CI 0.77 to 1.57)、早産(14.5% vs 14.3%, 0.99, 0.78 to 1.26)を含む、母体・児側アウトカムのいずれにおいても高いリスクとはならないようであった。

ここで用いられた高次傾向スコアマッチングとは、通常傾向スコアが測定された変数因子によって治療選択におけるバイアスの解決を試みるのに加え、他の質的診療情報を結合して他の重要未測定変数全体の代替である残差変数としてモデルに入れることによって、よりバイアスの少ない結果を得ようというものである。この研究では、ヘルスケアにおける診断・手順・薬剤の保険請求情報がこの目的で用いられた。

研究例 3

気分障害患者の電気けいれん療法と認知症のリスク(参考文献 6)

電気けいれん療法では、短期的な記憶障害がよくある有害事象として知られているが、長期的な認知機能のアウトカムについては結論が出ていない。そこで 2005 年から 2015 年のデンマーク国立患者登録システムのデータを用いて、ICD-10 において F30.0 から F39.9 の範囲内で初めて気分障害と診断された患者情報から、電気けいれん療法を受けた患者と、社会人口学のおよび臨床

特性における傾向スコアマッチングを用いて電気けいれん療法を受けなかった患者とを選択し、2016年までに認知症の診断が発生したかどうかをCox回帰分析によって比較した。

結果、およそ16万人の患者データが得られ、5901人が少なくとも1回のECTを受けていた。追跡観察機関の中央値は4.9年で、827284人年のデータが得られ、5901人のECTを受けた患者のうちの217人(3.6%)、受けてない患者162114人中の4987人(3.1%)が認知症の診断が発生した。人年法ではECT群で10000人年あたり70.4例、非ECT群では59.2例であった。50歳未満と50-69歳の比較的若い年代において、年齢をマッチングしてECTを受けた5901人を受けていない5901人と比較した結果ではECTは認知症の発生リスクとの関連はなかったが(調整ハザード比 [HR] 1.51, 95% CI 0.67-3.46, p=0.32 および 1.15, 0.91-1.47, p=0.22)、70歳以上ではECTは認知症のリスク減少と関連したものの(0.68, 95% CI 0.58-0.80; p<0.0001)、傾向スコアマッチングではこの結果は減弱されていた(0.77, 0.59-1.00; p=0.062)。また追加分析で認知症のリスクと合わせて死亡(追跡期間中に17.6%)も含めて見たところ、HR 0.98, 95% CI 0.76-1.26; p=0.24とECTとこれらのアウトカムの相関の有意性は消失した。

この研究では、追加分析として傾向スコアがCox回帰分析のベースラインの共変量に関する補正として用いられた。傾向スコアが最も近い患者をECT群と非ECT群から1対1で選択するマッチングを行ってハザード比を再計算している。

本総合研究への適用についての検討

わが国では、米国におけるMedicaid、また北欧諸国における医療ナショナル・データベースはないが、主にレセプト情報を集めたデータベースは存在する。代表的なものは下記である。レセプト情報・特定健診等情報データベース(NDB):厚生労働大臣が保有し、日本全国のレセプトデータ・特定健診等データを収集している。レセプトデータは2020年6月時点で193億件、特定健診・保健指導データは2.9億件と非常に

大きなものである。

JMDC claims database:複数の健康保険組合より寄せられたレセプトデータ(入院、外来、調剤)および健診データを保持している。2005年よりデータを蓄積し、累積母集団数は1000万人を超えている。

両データベースともレセプトを主としているため、処方中止・継続についての情報取得は容易である。しかし、精神症状の寛解・再発、離脱症状の出現などの患者の中止スタート地点・アウトカムは向精神薬の再開や他剤併用、レセプト病名から定義せざるを得ず、それが一番の課題になるだろう。

D. 考察

RCTが困難な臨床疑問に対しても、傾向スコアマッチングを観察研究データに適用することで臨床に必要なエビデンスを創出できる可能性がある。本研究でも、積極的に利活用することが望まれる。

E. 結論

本研究でも、傾向スコアマッチングを観察研究データに適用してあらたな知見を得る試みが望まれる。

参考文献

1. 康永秀生. できる!臨床研究 最短攻略 50の鉄則. 東京: 金原出版株式会社; 2017.
2. 康永秀生. できる!傾向スコア分析: SPSS・Stata・Rを用いた必勝マニュアル. 東京: 金原出版株式会社; 2018.
3. 木原正博, 木原雅子. 医学的研究のための多変量解析 第2版 標準一般化線形モデルから一般化推定方程式まで:最適モデルの選択、構築、検証の実践ガイド 第2版. 東京: メディカルサイエンスインターナショナル; 2020.
4. Stroup TS, Gerhard T, Crystal S, Huang C, Olfson M. Comparative Effectiveness of Clozapine and Standard Antipsychotic

- Treatment in Adults With Schizophrenia. *Am J Psychiatry*. 2016;173(2):166-173.
5. Vigod SN, Gomes T, Wilton AS, Taylor VH, Ray JG. Antipsychotic drug use in pregnancy: high dimensional, propensity matched, population based cohort study. *BMJ*. 2015;350:h2298.
 6. Osler M, Rosing MP, Christensen GT, Andersen PK, Jorgensen MB. Electroconvulsive therapy and risk of dementia in patients with affective disorders: a cohort study. *Lancet Psychiatry*. 2018;5(4):348-356.
- F. 研究発表
1. 論文発表
 1. Funada S, Watanabe N, Goto T, Negoro H, Akamatsu S, Uozumi R, Kishimoto S, Ichioka K, Segawa T, Furukawa TA, Ogawa O. Clinical feasibility and acceptability of adding cognitive behavioral therapy to pharmacotherapy for drug-resistant overactive bladder in women: A single-arm pilot study. *Low Urin Tract Symptoms*. 2021;13(1):69-78.
 2. Takeshima M, Utsumi T, Aoki Y, Wang Z, Suzuki M, Okajima I, Watanabe N, Watanabe K, Takaesu Y. Response to 'Efficacy and safety of bright light therapy for bipolar depression'. *Psychiatry Clin Neurosci*. 2020;74(7):410-411.
 3. Sahker E, Sakata M, Toyomoto R, Hwang C, Yoshida K, Luo Y, Watanabe N, Furukawa TA. Efficacy of brief intervention for drug misuse in primary care facilities: systematic review and meta-analysis protocol. *BMJ Open*. 2020;10(9):e036633.
 4. Leerssen J, Blanken TF, Pozzi E, Jahanshad N, Aftanas L, Andreassen OA, Baune BT, Brack I, Carballido A, Ching CRK, Dannlowski U, Dohm K, Enneking V, Filimonova E, Fingas SM, Frodl T, Godlewska BR, Goltermann J, Gotlib IH, Grotegerd D, Gruber O, Harris MA, Hatton SN, Hawkins E, Hickie IB, Jaworska N, Kircher T, Krug A, Lagopoulos J, Lemke H, Li M, MacMaster FP, McIntosh AM, McLellan Q, Meinert S, Mwangi B, Nenadic I, Osipov E, Portella MJ, Redlich R, Repple J, Sacchet MD, Samann PG, Simulionyte E, Soares JC, Walter M, Watanabe N, Whalley HC, Yuksel D, Veltman DJ, Thompson PM, Schmaal L, Van Someren EJW. Brain structural correlates of insomnia severity in 1053 individuals with major depressive disorder: results from the ENIGMA MDD Working Group. *Transl Psychiatry*. 2020;10(1):425.
 5. Kikuchi S, Imai H, Tani Y, Tajiri T, Watanabe N. Proton pump inhibitors for chronic obstructive pulmonary disease. *Cochrane Database Syst Rev*. 2020;8:CD013113.