

抗うつ薬治験データの人工知能解析に関する研究

研究分担者 丸尾和司 筑波大学医学医療系准教授  
研究分担者 古川壽亮 京都大学大学院医学研究科教授

研究要旨

日本で行われた7本、2399人分の、抗うつ剤のプラセボ対照治験データをプールして、QUINTという新しい機械学習手法を用いて治療効果の異なるサブグループを探索した。結果、うつ病初発から1年以上を経過しているかどうかと、性別とにより、質的な差異を示すサブグループが同定された。今年度は、このサブタイピングの外的妥当性を検討した。同一の臨床試験の後半データを利用した時間妥当性は概ね得られたが、まったく新しい臨床試験に対する外的妥当性は限定的であった。

A. 研究目的

患者特性に合わせて治療を選択する個別化医療が喧伝されるが、精神医療においてはこれまで実現されていない。

本研究では、日本で行われたプラセボ対照の抗うつ剤の治験の個人レベルのデータを用いて、抗うつ剤の選択の個別化（どういう患者ではとくに抗うつ剤が効きやすく、どういう患者ではプラセボとの差が付かないあるいはむしろプラセボの方が良いか）を明らかにする。

このように治療効果の大きさ（つまり、実薬とプラセボの間の差）に影響する因子を効果修飾因子と呼ぶ。効果修飾は、モデル的には因子と治療効果の間の交互作用によって検出されるが、交互作用には2種類ある。サブグループごとで治療効果（つまり実薬とプラセボの間の差）は異なるが、いずれのサブグループにおいても実薬がプラセボにまさっている場合と、サブグループによって優劣が逆転することがある場合である。前者を量的な交互作用、後者を質的な交互作用という。

個別化医療でとくに重要なのは、後者の、質的な交互作用である。本研究では、質的な交互作用に着目して、抗うつ薬の個別化医療を可能にする予測モデルを構築する。予測モデルの構築におい

ては、内的妥当性および外的妥当性の確保が重要になる。今年度は得られたモデルの妥当性検証に重点を置いて、本研究を行った。

B. 研究方法

本研究の対象とした7研究のデータ数の概要は以下の通りである。

試験	1週	2週	3週	4週	5週	6週	8週
試験1 <sup>1</sup>	438	423	413	405	0	391	0
試験2 <sup>2</sup>	296	285	0	277	0	265	259
試験3 <sup>3</sup>	483	468	455	451	0	437	423
試験4 <sup>4</sup>	262	255	243	228	218	216	0
試験5 <sup>5</sup>	399	395	393	379	0	367	360
試験6 <sup>6</sup>	367	345	0	327	0	318	312
試験7 <sup>7</sup>	521	516	502	499	0	489	530
計	2399	1874	1551	1788	218	1728	1572

試験5~7はポータルサイト上でのみ解析可能であるため、全試験の個人データの統合は困難であった。従って、試験1~4のデータを主要な試験とし、さらに試験1~4で被験者番号が前半の症例のデータを主要なデータと位置付け、主解析を実施した。試験1~4の被験者番号が後半の被験者のデータ及び試験5~7のデータは主解析の統計モデルの外部妥当性を検討するために用いた。

データ数も多く、治療効果も十分に検出される6週目を主要な時点として設定し、6週目のHAMDのベースラインからの変化量を主要なアウトカムとして設定した。

主解析では、Qualitative Interaction Trees (QUINT) 法<sup>8</sup>を適用した。この方法は、2つの治療の群間差が異なる2つの集団に逐次分割し、治療との交互作用に関して特徴的な部分集団を抽出できる決定木ベースのクラスタリング法である。

部分集団構成のための因子となり得る背景因子としては、少なくとも9割以上の症例で取られている以下の変数を利用した。

- i) 年齢
- ii) 性別
- iii) 体重
- iv) 大うつ病診断分（単一性 or 反復性）
- v) 身体合併症の有無
- vi) 最初の大うつ病発症時年齢
- vii) 現在の大うつ病エピソードの持続期間
- viii) 最初の大うつ病発症からの経過年数
- ix) HAMD ベースライン値
- x) HAMD サブスケール：anhedonia retardation (HAMD1+2+3+16)
- xi) HAMD サブスケール：body symptoms (HAMD9+13+14)
- xii) HAMD サブスケール：sleep problems (HAMD6+7+8)
- xiii) HAMD サブスケール：appetite (HAMD4+5)
- xiv) HAMD サブスケール：guilt & agitation (HAMD10+11+12+15+17)

欠測値の取扱いについては、欠測のある被験者を除外した完全症例解析を行った。QUINT法で抽出された部分集団(樹木の葉：リーフ)ごとにCohen's dとそのSEを推定した。

主解析の感度分析として、以下の解析を実施した。

- (i) 内部整合性と外部妥当性の検討のために、試験1~4のそれぞれで、主解析で抽出された各リーフ(以降、各リーフ)でのCohen's dとそのSEを推定した。

- (ii) 外部妥当性の検討のために、試験1~4の被験者番号が後半の症例データについて、各リーフでのCohen's dとそのSEを推定した。

- (iii) 外部妥当性の検討のために、試験5~7の各試験で、各リーフでのCohen's dとそのSEを推定した。

- (iv) 上記の全データセット(主解析のデータ、感度分析(i)~(iii)のデータ)で線形仮説 $L^T\beta = 0$ に関する推測を行った。ここで、 $L$ は対比ベクトルであり、 $\beta$ はリーフ、群とこれらの交互作用を因子とした線形モデルの回帰係数ベクトルである。治療群間差に関する各リーフの任意の順位に対応した複数の対比ベクトルを設定し、選択された対比の主解析との類似性を評価した。また、対比の検定から、リーフ間の治療効果の差が有意に異なるかを検討した。

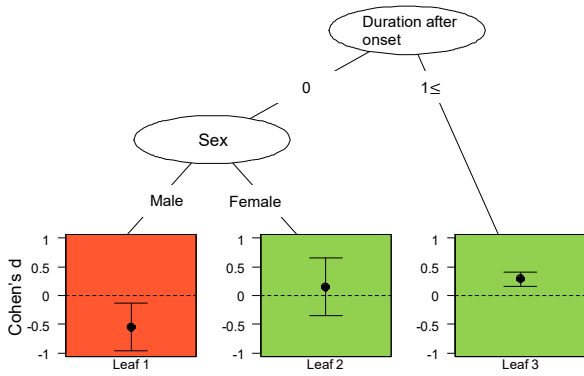
- (v) QUINT法は単一のRCTのデータを対象としており、試験間差をランダム効果として組み入れられない。この点でのQUINT法の妥当性確認のために、リーフ、群、リーフと群の交互作用を固定効果、試験IDを変量効果とした線形混合効果モデルを適用し、(iv)と同様の線形仮説の検定を実施した。そのうえで、主解析との類似性を検討した。

(倫理面への配慮)

本研究は、人を対象とする医学系研究の倫理指針の第1章第3の1「適用される研究」においていうところの、「既に連結不可能匿名化されている情報」を用いた研究であるので、倫理委員会の審査を要しない。

## C. 研究成果

QUINT法による最終的な樹状図および各部分集団でのCohen's dを以下に示す



リーブ		1	2	3
プラセボ	N	35	21	153
	Mean	-12.54	-10.19	-8.77
	SD	6.05	8.03	6.51
抗うつ剤	N	67	67	361
	Mean	-9.25	-11.13	-10.67
	SD	6.04	5.56	6.72
Cohen's d	d	-0.54	0.15	0.29
	SE	0.21	0.25	0.06

発症からの経過年数(duration after onset)と性別がモデルに含まれた。発症からの経過年数が1年以上のリーフ3では抗うつ剤の効果が有意にプラセボよりも高かった。経過年数が1年未満の場合において、男性(リーフ1)ではプラセボに有意に劣っており、女性(リーフ2)では治療群間差が小さかった。

以下の表は、感度分析(i)~(iii)の結果である。表内の数値はCohen's dでありカッコ内はそのSEである。

感度分析   Leaf	1	2	3
(i) 試験 1	-0.10(0.25)	-0.42(0.34)	0.54(0.13)
(i) 試験 2	-0.22(0.28)	-0.07(0.33)	0.13(0.16)
(i) 試験 3	-0.33(0.30)	0.88(0.33)	0.09(0.13)
(i) 試験 4	-0.79(0.68)	0.45(0.62)	0.50(0.17)
(ii)	-0.01(0.22)	0.03(0.26)	0.17(0.11)
(iii) 試験 5	0.52(0.25)	-0.03(0.23)	0.25(0.14)
(iii) 試験 6	-0.44(0.31)	-0.51(0.31)	0.18(0.13)
(iii) 試験 7	0.15(0.22)	0.02(0.22)	0.13(0.11)

(i), (ii)については、主解析とおおむね同様の結果が得られた。(iii)については、全ての試験においてリーフ3で抗うつ剤がプラセボと比較して効果がある傾向がみられたものの、その他のリーフでの効果は一定した傾向がみられなかった。

以下に、感度分析(iv), (v)の結果を示す。表中の数値は対比検定のp値であり、p値が最小の対比が選択された対比である。二重下線が最小のp値、下線が2番目に最小のp値である。主解析ではリーフ3, 2, 1の順に治療効果が高く、そのリーフ間差は有意水準0.05で有意ということが示された。

Order	主	(i) 試験 1	(i) 試験 2	(i) 試験 3	(i) 試験 4
3 2 1	<u>0.001</u>	<u>0.027</u>	<u>0.250</u>	0.250	<u>0.129</u>
2 3 1	0.581	<u>0.004</u>	<u>0.636</u>	0.076	0.969
2 1 3	0.581	0.004	0.636	<u>0.076</u>	0.969
3 1 2	<u>0.049</u>	0.330	0.655	<u>0.025</u>	<u>0.266</u>
1 2 3	0.001	0.027	0.250	0.250	0.129
1 3 2	0.049	0.330	0.655	0.025	0.266
Order	(ii)	(iii) 試験 5	(iii) 試験 6	(iii) 試験 7	(v)
3 2 1	<u>0.451</u>	0.322	<u>0.111</u>	0.920	<u>0.002</u>
2 3 1	<u>0.577</u>	<u>0.298</u>	<u>0.064</u>	<u>0.650</u>	0.465
2 1 3	0.577	0.298	0.064	0.650	0.465
3 1 2	0.931	0.098	0.846	0.657	<u>0.109</u>
1 2 3	0.451	0.322	0.111	0.920	0.002
1 3 2	0.931	<u>0.098</u>	0.846	<u>0.657</u>	0.109

感度分析(i), (iii)で選択された対比は概ね主解析と同様であったが, (iv)については, 一定の傾向がみられなかった。感度分析(v)では, 主解析とほとんど同等の結果が得られ, 変量効果を考慮しなかったことの影響は小さかったことが示唆された。

#### D. 考察

日本で行われた7本、2399人分の、抗うつ剤のプラセボ対照試験データをプールして、QUINTという新しい機械学習手法を用いて治療効果の異なるサブグループを探索したところ、うつ病初発から1年以上を経過しているかどうかと、性別とにより、質的な差異を示すサブグループが同定された。このサブグループは、内的妥当性および同じ試験の後半データセットにおける時間妥当性は概ね確認されたが、まったく別個の試験における外的妥当性は一定しなかった。

#### E. 結論

QUINTは、個別化医療においてとくに重要な質的な交互作用を検出するために有用な手法である。しかし、安定した外的妥当性を得るためにはさらなるデータと解析が必要である。

#### F. 健康危険情報

なし

#### G. 研究発表

該当なし

#### H. 知的財産権の出願・登録状況

該当なし

#### 参考文献

1. Higuchi T, Murasaki M, Kamijima K. Clinical evaluation of duloxetine in the treatment of major depressive disorder: placebo- and paroxetine-controlled double-blind comparative study. Japanese Journal of Clinical Psychopharmacology [in Japanese] 2009; 12: 1613-34.
2. Hirayasu Y. A dose-response study of escitalopram in patients with major depressive disorder: a placebo-controlled, double-blind study. Rinsho Seishin Yakuri

(Japanese Journal of Clinical Psychopharmacology) 2011; 14(5): 871-82.

3. Hirayasu Y. A dose-response and non-inferiority study evaluating the efficacy and safety of escitalopram in patients with major depressive disorder: a placebo- and paroxetine-controlled, double-blind, comparative study. Rinsho Seishin Yakuri (Japanese Journal of Clinical Psychopharmacology) 2011; 14(5): 883-99.
4. Kinoshita T. A double-blind, placebo-controlled study of a new antidepressant, mirtazapine, in depressed patients. Japanese Journal of Clinical Psychopharmacology [in Japanese] 2009; 12: 289-306.
5. Higuchi T, Hong JP, Jung HY, Watanabe Y, Kunitomi T, Kamijima K. Paroxetine controlled-release formulation in the treatment of major depressive disorder: a randomized, double-blind, placebo-controlled study in Japan and Korea. Psychiatry Clin Neurosci 2011;65:655-663.
6. Koshino Y, Bahk WM, Sakai H, Kobayashi T. The efficacy and safety of bupropion sustained-release formulation for the treatment of major depressive disorder: a multi-center, randomized, double-blind, placebo-controlled study in Asian patients. Neuropsychiatr Dis Treat 2013;9:1273-1280.
7. Higuchi, T., Kamijima, K., Nakagome, K., Itamura, R., Asami, Y., Kuribayashi, K., Imaeda, T., 2016. A randomized, double-blinded, placebo-controlled study to evaluate the efficacy and safety of venlafaxine extended release and a long-term extension study for patients with major depressive disorder in Japan. Int. Clin. Psychopharmacol. 31, 8-19.
8. Dusseldorp E, Van Mechelen I. Qualitative interaction trees: a tool to identify qualitative treatment-subgroup interactions. Stat Med 2014; 33(2): 219-37.