

博士論文（要約）

断片的な観測データから
慢性疾患の長期的な病態進行を推定する
新規統計手法の構築

石田 崇朗

【序論】

一般に、長期間かけて徐々に進行する慢性疾患においては、病態進行のステージごとに病態の増悪に寄与する分子メカニズムが切り替わるケースが多い。そのため、慢性疾患患者に対して最適な治療戦略を立案するためには、患者の病態ステージを正確に推定して薬剤を選択することが重要となる。また、医薬品開発においても同様に、候補化合物の作用メカニズムに基づいて、最も薬効が発揮されると想定される病態ステージに位置する被験者集団を選択して臨床試験を行うことが必要となる。このような、患者ごとの正確な病態ステージの推定には、病態進行の全体像が定量的に理解されていることが必須である。一方で、慢性疾患を対象としたコホート研究の観察期間は、長期的な病態進行に比べて短く設定せざるを得ないことが多く、これが病態進行の全体像を俯瞰することを困難にする原因の一つとなっている。そこで本研究では、多数の患者から取得された、時間的に断片化された観測データをもとに、数理モデルを用いてデータを繋ぎ合わせることで、長期的な病態進行全体像を再構築する新規方法論を構築することを目的とした。なお、本方法論の有用性の検証には、アルツハイマー病 (AD) を選択することとした。AD は、アミロイド β ($A\beta$) の脳内沈着に始まり、最終的な認知機能低下を発現するまでに 10 年以上かかると考えられている。 $A\beta$ を標的とした治療薬は、投与開始がより早期であるほど有効である可能性が過去の臨床試験結果より示唆されており、一方で既に認知症状低下が生じた患者の場合は、神経活動を賦活化させる症候改善薬が有効とされている。従って、AD においても患者個々の病態ステージの推定が臨床上有用となる可能性が十分に考えられる。加えて、病態初期から後期までの幅広い被験者から測定された多種類のバイオマーカー情報が豊富な、米国 ADNI のデータベースが利用可能であることも、AD を解析対象とした理由である。

【方法・結果】

1. 構築した新規方法論 (SReFT) のコンセプト

本研究では、長期的な病態進行全体像を再構築する方法として、“SReFT (Statistical Restoration of Fragmented Time-course)” を新規に構築した。母集団薬物動態解析では、投与時点から測定までの期間が明確であるため、測定時点が既知の時系列データから母集団分布の推定を行うために一般的に利用されている、非線形混合効果モデルの適用が可能である。一方で、本研究で用いるようなコホート研究データでは、病態の発症から観測が行われるまでの時間経過を知る術がない。そこで本研究では、従来の非線形混合効果モデルを拡張し、各患者の「病態時刻」(すなわち、病態の発症から観測が行われるまでの

時間経過)と母集団分布を同時に推定する SReFT を構築した。SReFT を利用することで、測定時点が未知の患者データベースからでも、病態進行全体を再構築することが可能である(図 1a)。さらに母集団分布を推定していることから、新規に患者データが得られた際に、測定時点数が少ないデータであっても、その患者の病

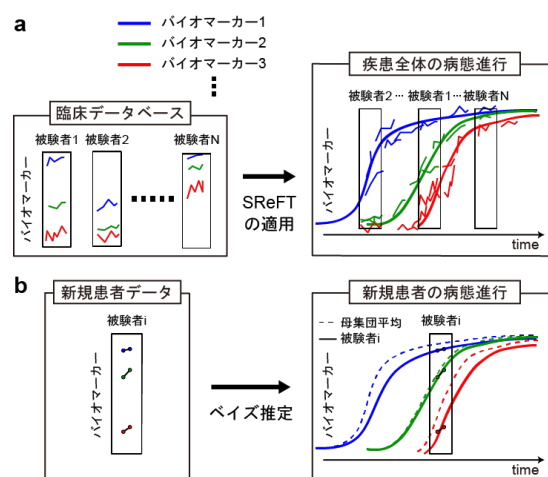


図 1 SReFT のコンセプト (a) 疾患全体の病態進行の推定、(b) ベイズ推定による新規患者の病態進行及び病態時刻の推定

態時刻、及びその後の病態変化の様子を、ベイズ推定手法により精度良く予測することが可能である(図 1b)。また SReFT は、統計学的には最尤推定法として定式化できることを確認しており、性別や遺伝子型などの共変量が病態進行に影響を与える効果の有意性を統計学的に判断することが可能となっている。なお、今回は AD の病態進行に沿ったバイオマーカーの変化を記述する関数として、3 種類のみパラメータ(a, b, c)で構成され、かつ線形的な変化や指数関数的な変化、さらにはシグモイド関数的な変化などの、様々な変化を柔軟に表現可能な、二重指数関数($f = a \exp(b \exp(c t))$)を使用した。実際に解析を行う際は、データの偏りを補正するため、データを対数変換している。そのため、関数もそれに伴い必要な変換をしていることから、解析ではパラメータ α, β, γ からなる関数 $f' = \alpha + \beta(\exp(\gamma t) - 1)/\gamma$ を使用した。

2. 仮想データセットを用いた SReFT の検証

SReFT では、母集団分布の推定に用いる非線形混合効果モデルを拡張し、各患者データの病態時刻に関する推定も同時に行っている。そこでまず、SReFT による病態時刻の推定精度に関して検証を行うこととした。病態時刻の推定精度を見積もるためには、病態時刻が既知のデータから時刻情報を取り除いたデータを用い、時刻情報の復元が可能であるかを検証する必要がある。そのため本解析では、以下の手順で生成したデータセットを使用した。①まず、3 種類の仮想的な病態バイオマーカーを仮定し、それらの経時的な変化を二重指数関数で表現するための各パラメータ値に関し、母集団平均・分散の値を設定した。②次いで、被験者ごとのパラメータ値を①の分布に従ってランダムに発生させ、被験者ごとに各バイオマーカーの経時変化を記述した。③さらに被験者ごとの観測開始時点(発症からの時間経過)をランダムに設定し、②で記述した被験者ごとのバイオマーカー経時変化曲線を元に、一定間隔で複数時点の観測データを生成した。各観測データには、一定のランダムな観測ノイズを加えた上で、最終的に病態時刻の情報を取り除き、擬似的な疾患コホートデータセットを作成した(図 2a)。このデータセットに対して SReFT 解析を適用することで推定された各被験者の病態時刻は、データ生成時に設定した各被験者の病態時刻と良く一致しており、SReFT が病態時刻を精度良く復元できることが確認された(図 2b, c)。これらの結果は、非線形混合効果モデルからの拡張部分である病態時刻推定に関して十分な精度が得られることを示している。

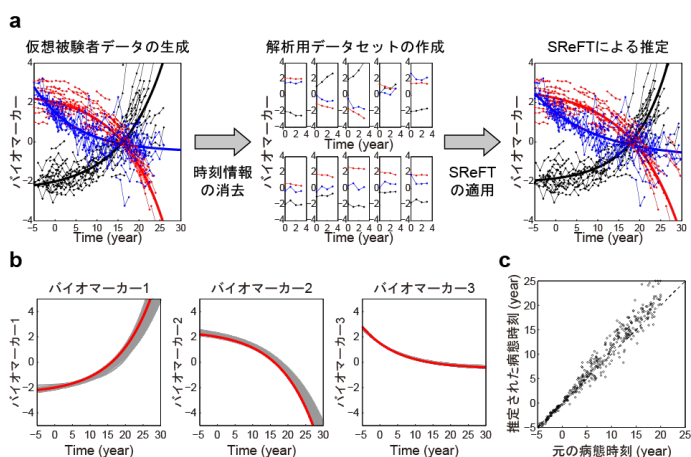


図 2 仮想データセットによる SReFT 解析

(a) データ生成から解析までの流れ

(b) 推定された母集団平均の病態進行(黒:元の病態進行、グレー:仮想データから推定された病態進行($n = 100$))

(c) 推定された病態時刻

的変化を二重指数関数で表現するための各パラメータ値に関し、母集団平均・分散の値を設定した。②次いで、被験者ごとのパラメータ値を①の分布に従ってランダムに発生させ、被験者ごとに各バイオマーカーの経時変化を記述した。③さらに被験者ごとの観測開始時点(発症からの時間経過)をランダムに設定し、②で記述した被験者ごとのバイオマーカー経時変化曲線を元に、一定間隔で複数時点の観測データを生成した。各観測データには、一定のランダムな観測ノイズを加えた上で、最終的に病態時刻の情報を取り除き、擬似的な疾患コホートデータセットを作成した(図 2a)。このデータセットに対して SReFT 解析を適用することで推定された各被験者の病態時刻は、データ生成時に設定した各被験者の病態時刻と良く一致しており、SReFT が病態時刻を精度良く復元できることが確認された(図 2b, c)。これらの結果は、非線形混合効果モデルからの拡張部分である病態時刻推定に関して十分な精度が得られることを示している。

3. ADNI データセットへの SReFT の適用

続いて ADNI データセットを用いた SReFT 解析を実施した。本解析では、脳脊髄液 (CSF) 中の $A\beta(1-42)$ 濃度がカットオフ値 (192 pg/mL) を下回る被験者、計 437 名 (健常:83 名、軽度認知障害 (MCI):242 名、AD:112 名) を解析対象とした。また、解析には 7 種のバイオマーカー時系列データを使用した。すなわち、脳内の $A\beta$ 蓄積を反映する CSF $A\beta$ 濃度及びアミロイド PET スコア、神経変性の程度を反映する CSF 中のタウタンパク質 (Tau) 濃度、MRI で測定された海馬及び脳室の容積、認知機能及び日常生活動作のレベルを評価する CDR-SB スコア、そして脳のブドウ糖消費量を反映する FDG-PET スコアである。これらの

データを用いた SReFT 解析の結果、推定された病態時刻は、健常-MCI-AD の順にほぼ並んでおり、SReFT により復元された AD 病態の進行過程は、従来の診断基準と合致することが示唆された。また、CSF $A\beta$ とアミロイド PET スコアの値は病態初期に、CDR-SB スコアと FDG-PET スコアの値は病態後期に、それぞれ急激な変化が観察された。海馬及び脳室の容積は、病態が進行するに従って値の変化が徐々に大きくなった。CSF Tau に関しては、病態進行に伴い緩やかに値が大きくなる様子が観察されたものの、特に病態後期において、値の変化の大きさに対して被験者間での値のばらつきが大きくなった。

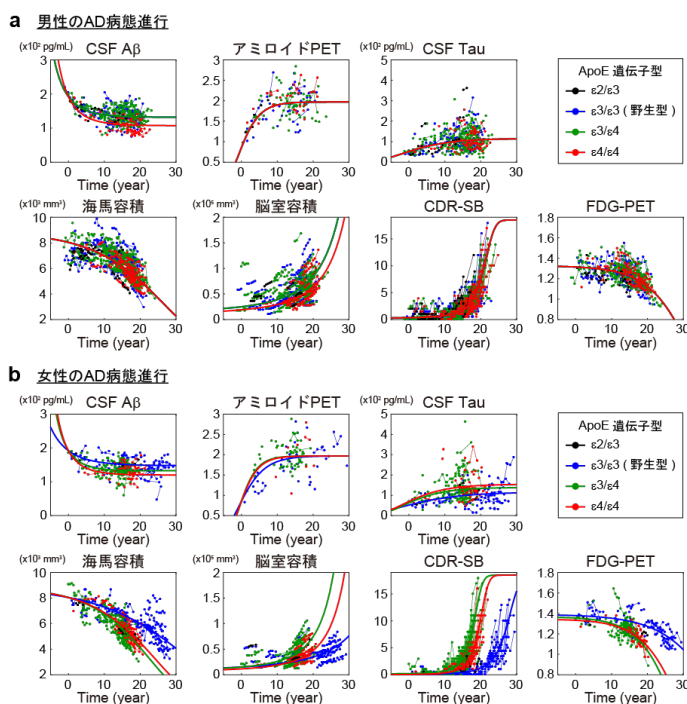


図 3 ADNI データをもとにした長期的なバイオマーカー変化の推定 (a:男性の AD 病態進行、b:女性の AD 病態進行) (※病態時刻ゼロ:CSF $A\beta$ が 192 pg/mL を下回る時点 (Shaw et al, 2009))

次に、AD 発症の危険因子として示唆されている、性別や ApoE 遺伝子型、及びそれらの交互作用を共変量として取扱い、AD の病態進行に与える影響を定量的に評価した。その結果、男性の病態進行速度は、野生型 (ApoE $\epsilon 3$ ホモ) と ApoE $\epsilon 4$ 保有群 ($\epsilon 4$ ヘテロ及びホモ) との間で有意な差は認められなかった一方で、女性では、野生型の病態進行速度は、 $\epsilon 4$ 保有群と比べて 40% 程度遅いことが明らかとなり、男性の野生型と比べても 21% 遅かった (図 3)。この結果は、 $\epsilon 4$ アリルが AD 発症に与える影響が、男性に比べて女性の方が大きい、という過去の知見と矛盾しないものと考えられた。なお、本推定結果の再現性に関しては、ADNI データセットから重複を許したランダム再抽出を行うことで、ブートストラップデータセットを 100 パターン生成し、同様に SReFT 解析を行った結果、各パラメータの推定精度が高いことを確認している。

4. SReFT を応用した臨床試験のシミュレーション

SReFT により病態進行の母集団分布が得られれば、ベイズ推定を用いて各被験者の病態時刻を推定することができる。この病態時刻の情報は、臨床試験における、被験者の組入れ基準として利用可能である。そこで、仮想的に薬物の臨床試験のシミュレーション解析を行い、病態時刻を組入れ基準とした場合の必要被験者数

を、特定のバイオマーカー値(AD では認知機能テストの値)を基準として組入れる従来の方法と比較することで、臨床試験における SReFT の有用性を検証した(図 4a)。本解析で用いる仮想薬物は、AD のその後の病態進行速度を抑える効果があり、かつ投与後の相対進行速度比 R は、治療開始の時期によって変化すると仮定した(図 4b)。なおこの仮定に関しては、抗 $A\beta$ 薬の過去の臨床試験において、投与開始時期がより早期であるほど、治療効果が高まる可能性が示唆されていることを反映している。シミュレーションの結果、認知機能テストの値に基づく試験組入れを行う従来の手法では、病態初期と後期での R の値の差が大きくなるほど、必要被験者数が指数関数的に上昇することが明らかとなった(図 4d)。一方で、病態時刻を組入れ基準とした方法では、組入れ時の被験者の病態時刻を適切に調節することで、必要被験者数を現実的な値にまで抑えられることが示唆された(図 4c, d)。従って、病態時刻を組入れ基準として利用することで、臨床試験の成功率が高まる可能性が示唆された。

【まとめ・考察】

本研究により、病態時刻が不明な短期間の患者データから、慢性疾患の長期的な病態進行を再構築する新規方法論 SReFT を構築し、これを米国 ADNI データに適用した結果、AD の長期的な病態変化を推定することができた。また共変量解析からは、性別や ApoE 遺伝子型といった要因が病態進行に与える影響を定量的に評価することができた。今後は様々な因子を共変量とした解析を行うことで、慢性疾患の増悪に関与する因子の探索に繋げることができる。さらに本手法は、各患者の病態時刻を正確に推定することができ、この推定された病態時刻の情報を臨床試験に応用することで、サンプルサイズを縮小できる可能性が示唆された。今後は我が国における J-ADNI コホートのデータ、さらには別の慢性疾患にも解析対象を拡張する予定である。

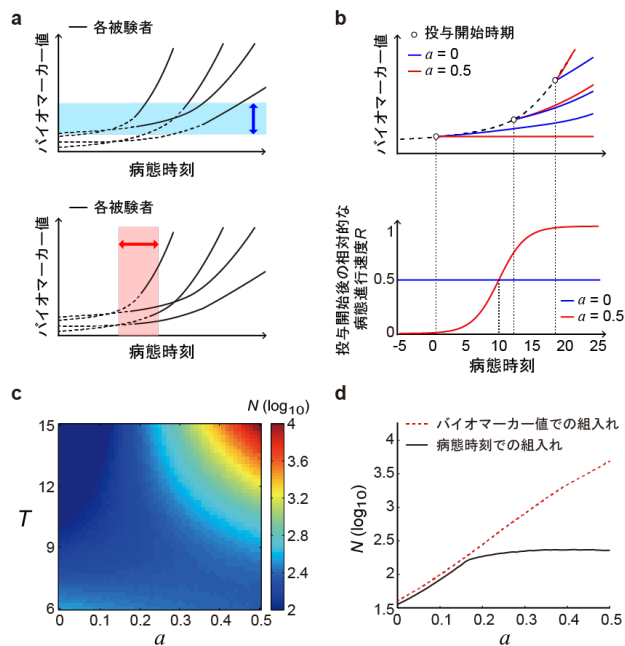


図 4 SReFT 解析を応用した臨床試験のシミュレーション (a) 上:バイオマーカーの値による従来の組入れ基準(従来条件)、下:病態時刻による組入れ基準(SReFT 条件)。(b) 仮想薬物による病態進行速度の抑制効果(上)とその大きさ(下)。薬物投与後の相対進行速度比 $R = 1/(1+\exp(-a(t-10)))$ 。(c) 投与開始時期 T 及び R の傾き a と、必要被験者数 N との関係性。(d) R の傾き a と、必要被験者数 N の最小値(従来条件と SReFT 条件との比較)