

- Diabetic Patients. *Diabetes Care* 2014;37:217-225.
7. Goto A, Arah OA, Goto M, Terauchi Y, Noda M; Severe Hypoglycaemia and Cardiovascular Disease: Systematic Review and Meta-Analysis with Bias Analysis *BMJ*. 2013 Jul 29;347:f4533. doi: 10.1136/bmj.f4533
 8. Noto H, Goto A, Tsujimoto T, Noda M: Effect of calcium channel blockers on incidence of diabetes: a meta-analysis *Diabetes Metab Syndr Obes* 6:257–261, 2013
 9. 「糖尿病と癌に関する委員会報告」
糖尿病と癌に関する委員会 ; 春日 雅人、植木浩二郎、田嶋 尚子、野田 光彦、大橋 健、能登 洋、後藤 温、小川 渉、堺 隆一、津金昌一郎、浜島 信之、中釜 斉、田嶋 和雄、宮園 浩平、今井 浩三
糖尿病 56 (6) : 374-390, 2013
 10. Kishimoto M, Noda M: A report on diabetes seminars for medical staff: findings from a questionnaire survey. *Diabetol Int* 4(4):261-265, 2013
 11. Goto M, Goto A, Morita A, Deura K, Sasaki S, Aiba N, Shimbo T, Terauchi Y, Miyachi M, Noda M, Watanabe S; Saku Cohort Study Group. Low-molecular-weight adiponectin and high-molecular-weight adiponectin levels in relation to diabetes. *Obesity* 2013 Jul 1. doi: 10.1002/oby.20553. [Epub ahead of print]
 12. Goto A, Morita A, Goto M, Sasaki S, Miyachi M, Aiba N, Kato M, Terauchi Y, Noda M, Watanabe S; Saku Cohort Study Group. Validity of diabetes self-reports in the saku diabetes study. *J Epidemiol.* 23(4):295-300, 2013
 13. Noto H, Goto A, Tsujimoto T, Osame K, Noda M. Latest insights into the risk of cancer in diabetes. *J Diabetes Invest* 4:225-232, 2013
 14. Goto M, Morita A, Goto A, Deura K, Sasaki S, Aiba N, Shimbo T, Terauchi Y, Miyachi M, Noda M, Watanabe S, for the SCOP Study Group: Reduction in Adiposity, β -Cell Function, Insulin Sensitivity, and Cardiovascular Risk Factors: A Prospective Study among Japanese with Obesity *PLoS ONE* 8(3): e57964, 2013
 15. Kishimoto M, Noda M: Diabetes care: After the Great East Japan Earthquake *J Diabetes Invest* 4(1): 97-102, 2013
 16. Noto H, Goto A, Tsujimoto T, Noda M: Low-Carbohydrate Diets and All-Cause Mortality: A Systematic Review and Meta-Analysis of Observational Studies *PLoS ONE* 8(1): e55030, 2013
 17. Kasuga M • Ueki K • Tajima N • Noda M • Ohashi K • Noto H • Goto A • Ogawa W • Sakai R • Tsugane S • Hamajima N • Nakagama H • Tajima K • Miyazono K • Imai K. Report of the Japan Diabetes Society/Japanese Cancer Association joint committee on diabetes and cancer. *Cancer Sci* 2013; 104: 965–976 / *Diabetol Int* (2013) 4:81–96

1 8 . Ina K, Hayashi T, Araki A, Kawashima S, Sone H, Watanabe H, Ohru T, Yokote K, Takemoto M, Kubota K, Noda M, Noto H, Ding QF, Zhang J, Yu ZY, Yoon BK, Nomura H, Kuzuya M; Japan CDM Group. Importance of high-density lipoprotein cholesterol levels in elderly diabetic individuals with type IIb dyslipidemia: A 2-year survey of cardiovascular events. *Geriatr Gerontol Int*. 2013 Nov 12. doi: 10.1111/ggi.12168. [Epub ahead of print]

1 9 . Hayashi T, Araki A, Kawashima S, Sone H, Watanabe H, Ohru T, Yokote K, Takemoto M, Kubota K, Noda M, Noto H, Ina K, Nomura H. Metabolic predictors of ischemic heart disease and cerebrovascular attack in elderly, diabetic individuals: difference in risk by age. *Cardiovascular Diabetology* 2013, 12:10 doi:10.1186/1475-2840-12-10

2 0 . Kishimoto M, Inoue K, Noda M: The effect of alogliptin and metformin combination therapy in type 2 diabetes: A pilot study. *J Endocrinol Metab* 3(4-5):111-118, 2013

2 1 . Noto H, Goto A, Tsujimoto T, Noda M: Cancer risk in diabetic patients treated with metformin: A systematic review and meta-analysis. *PLoS One* 7(3): e33411. doi:10.1371/journal.pone.0033411 (2012)

2 2 . Noto H, Tsujimoto T, Noda M: Significantly increased risk of cancer in diabetes mellitus patients: A

meta-analysis of epidemiological evidence in Asians and non-Asians. *J Diabetes Invest* 3:24-33, 2012

2 3 . Noto H: Impaired functionality of HDL in diabetes. *Diabetol Int* 3:5-7,2012

H. 知的所有権の取得状況
なし

厚生労働科学研究費補助金（循環器疾患・糖尿病等生活習慣病対策総合研究事業）

分担研究報告書

保健指導等を活用した総合的な糖尿病治療の年代別要因を踏まえた研究

スタチンの脳梗塞予防に対する直接効果と間接効果の推定

分担研究者 久保田 潔（東京大学大学院医学系研究科薬剤疫学講座 教授）

研究要旨

「保健指導等を活用した総合的な糖尿病治療の年代別要因を踏まえた研究」（課題 ID：11103634）は、平成 21-2 年度厚労省研究「糖尿病患者心血管イベントの後ろ向きコホート研究」（課題番号：21180301）におけるケース・コホート研究を踏まえ、保険指導等を活用する新たな研究であるが、平成 21-22 年度厚労省研究「糖尿病患者心血管イベントの後ろ向きコホート研究」自体を継続し、新たなデータを追加収集し、ケース・コホートデータの最終解析を行うことをもその課題としている。

平成 25 年度の本分担研究では、スタチン使用による脳梗塞発生予防の効果を Lange らの方法により、血清脂質の変化を介する間接効果と、それとは異なる直接効果に分離することを試みた。Lange らの方法では、まず、スタチン使用後の血清脂質値を各個人の性、年齢その他の因子、曝露の有無により予測するモデルを作成する。本研究では、LDL コレステロール、HDL コレステロール、総コレステロール、トリグリセリドを mediator として検討したが、スタチンによって変化するのは平均で 5mg/dL 前後と予測された。間接効果、直接効果の推定では、4 種類の mediator のいずれについても、また本研究で試みられた 2 種類の異なる予測モデルのいずれを使った場合でも、間接効果は無視できる程度であり、「直接効果」がほとんどであり、「直接効果」により脳梗塞発生のリスクが約 0.6 倍になることが示された。この 0.6 倍という値は臨床試験における値よりも低く、薬の真の直接効果以外に healthy user bias（スタチンを使用開始する患者はそうでない患者よりも健康志向が強く、実際により健康である）にもよるのかもしれない。しかし、「間接効果がほとんどない」との結論については healthy user bias の有無にかかわらず成立すると考えられる。また、この結果は、スタチン使用者における脳梗塞の発生低下が、LDL や総コレステロールの広い範囲で見られること、脳梗塞の発生低下が、LDL や総コレステロールが上昇した群でも見られたこととも一致する。ただし、スタチン使用者においては、非使用者においてと同様、観察終了時の LDL や総コレステロールが低いほど脳梗塞発生が低いことも認められている。本研究から得られるのは「血清脂質は正常化させることが望ましいが、正常化しないからと言って、スタチンの脳梗塞の予防効果が期待できないわけではない」というメッセージと考えられる。

A. 研究目的

ケース・コホートデータを用いたスタチンの脳梗塞発生に対する直接・間接効果の推定

「保健指導等を活用した総合的な糖尿病治療の年代別要因を踏まえた研究」（課題ID: 11103634、以下「本研究」）は平成 21、22 年度の厚生労働科学研究「糖尿病患者における心血管イベント発症に関する後ろ向きコホートに関する研究」（課題番号：21180301）を踏まえて実施されている。「糖尿病患者における心血管イベント発症に関する後ろ向きコホートに関する研究」では、ケース・コホートデザインが採用されたが、さらに平成 23 年度以後のデータも追加収集して実施するケース・コホートデータの最終解析も本研究における課題の一つである。本研究に関連して本分担研究者と研究協力者はケース・コホート研究のサンプルサイズについての論文を公表している¹⁾。

本報告書では、「保健指導効果の疫学的解析」についてはすでに昨年度報告済であることから、糖尿病患者におけるスタチンによる血清脂質の改善がいかに脳卒中の発生予防に寄与しているかについて、本研究から得られたデータを用いて検討した。

スタチンの心血管イベント予防に対する効果として、血清脂質を改善することを介する間接的効果とともに、血清脂質レベルの変化を介さない直接的効果を考えることができる。スタチンの心血管イベント予防効果において間接効果と直接効果のいずれがより重要であるかを明らかにすることは、薬物の作用メカニズムを考える上での参考となるのみならず、目標の血清脂質の達成の臨床的重要性に関する示唆を与える。

B. 研究方法

1. 研究対象患者

本研究で対象としたのは、平成 21、22 年度の厚生労働科学研究「糖尿病患者における心血管イベント発症に関する後ろ向きコホートに関する研究」において 20 の医療機関で登録された 3080 人の 2 型糖尿病の患者である。平成 24 年度報告書では、対象者は 3088 人とされているが、そのうち 2 例が登録日の 2004 年 9 月 1 日以前に死亡していることが判明した。また、2004 年 9 月 1 日以前に脳梗塞が発生している 6 例については、本研究では対象者から除外した。

研究期間は 2004 年 9 月 1 から 2010 年 9 月 30 日までである。各患者の追跡期間は 2004 年 9 月 1 日から 2010 年 9 月 30 日の研究終了日、イベント発生、脱落のいずれかのうち最も早い時点までとした。

研究の効率性を上げるために、観察開始時のサブコホート 404 例をランダムに抽出したが、ケースについてはサブコホート内かサブコホート外かによらず、3080 人のいずれかに発生したケースをすべて特定した。ケースとしては心筋梗塞、脳梗塞、総死亡のいずれかの発生としたが、本報告では脳梗塞に絞った再解析結果を示す。

2. 薬の効果推定における中間因子の扱い

一般に薬物の効果は何らかの mediator を介して発現する場合、その mediator で「調整」する解析を行うと、薬の効果を正しく示しえなくなる事が知られている。高脂血症治療薬の効果を例にとると、mediator と考えられる治療薬開始後の血清脂質の値による「調整」は、薬の効果の正しい推定を妨げる。たとえば血清脂質が正常化しない (M=0) 患者における心血管イベント発生割合 P0 と血清脂質の改善が

見られた ($M=1$) 患者におけるイベント発生割合を $P1$ ($P0>P1$) が薬の使用・非使用には全く無関係と考える。すなわち、高脂血症治療薬の効果はもっぱら、血清脂質の正常化を介して間接的にもたらされると仮定する。この際に、血清脂質が改善するかどうか (M) で「調整」を行う (M の値で層別化するか多変量モデルに M を説明因子として含める) と、血清脂質レベルの正常化はイベント発生のリスクを下げるが (多変量解析では血清脂質改善のイベント発生のリスク比またはハザード比 <1)、薬はリスクに影響しない (薬のイベント発生のリスク比またはハザード比 $=1$) と推定され、薬の効果は皆無であるとの誤った結論が導かれる。コホート研究や臨床試験の解析などにおいて、投与前の *baseline* の変数で「調整」することはあっても、投与後の *mediator* による調整を行ってはならないとされる理由である。

1986年に Baron²⁾らは曝露の効果がすべて間接効果である時、*mediator* で「調整」することにより、曝露の効果が見かけ上消えてしまうことを利用する間接効果と直接効果の推定法を提唱した。Baron らが提唱したのは、*mediator* をモデルに加えた時の薬の効果を「直接効果」として扱い、*mediator* をモデルに含めない時の「全効果」から「直接効果」を差し引いて「間接効果」を求める方法である。Baron らの方法は、心理学の分野で広く使われたが、*mediator* とイベント発生の関係に、または曝露 (薬の使用) と中間因子との間に「交絡」として働く因子が存在すると、正しい「直接効果」「間接効果」が求められないことが知られている^{3,4)}。たとえば、ある因子が

mediator とイベント発生の間に「交絡」として働くのなら、その因子の異なるレベルにおける *mediator* の値の分布は異なり、*mediator* を除いたモデルでその因子は、*mediator* のいわば「代替」として働き、*mediator* を除外したにも関わらず、*mediator* を介した「間接効果」を完全に取り除くことができない。

「直接効果」「間接効果」の推定は固有の困難さを伴っている。たとえば薬の投与後、*mediator* の値が変化するとともに、投与しなければ発生しないと考えられるイベントが発生したことが観察されても、イベント発生が *mediator* の変化を介して起こったのか、*mediator* の変化とは無関係に起こったのかの判別は困難である。薬非投与時のイベント発生 Y (イベント発生 Yes/No の時 $Y=1/0$) を薬非投与時 ($D=0$) に観察される非投与時の *mediator* の値 ($M(D^*=0)$) を用いて $Y[(D=0), M(D^*=0)]$ と表し、薬投与時のイベント発生 Y (イベント発生 Yes/No の時 $Y=1/0$) を薬投与時 ($D=1$) に観察される薬投与時の *mediator* の値 ($M(D^*=1)$) を用いて $Y[(D=1), M(D^*=1)]$ と表す。ある患者に薬を投与しない時、イベントが発生しないのなら $Y[(D=0), M(D^*=0)]=0$ と表され、薬を投与した時、イベントが発生したのなら $Y[(D=1), M(D^*=1)]=1$ と表される。ここで曝露時と非曝露時で *mediator* の値が変化したのなら $M(D^*=0) \neq M(D^*=1)$ である。

仮に同一の患者における薬は非投与 ($D=0$) であるにも関わらず、*mediator* は投与後の値 $M(D^*=1)$ である時のイベント発生 ($Y[(D=0), M(D^*=1)]$)、あるいは薬投与後 ($D=1$) であるにも関わらず、*mediator* の値が薬非投与時の値 ($M(D^*=0)$) である時のイベント発生 ($Y[(D=1), M(D^*=0)]$) が推定可能なら、イベント発生が *mediator* を介したものであるか否

かを推定できる。たとえば $Y(D=0, M(D^*=0))=0$ 、 $Y(D=0, M(D^*=1))=0$ 、 $Y(D=1, M(D^*=1))=1$ であるのなら、薬非投与では起こらなかったイベントが、薬投与後起こるのは、mediator の変化 ($M(D^*=0) \rightarrow M(D^*=1)$) によるのではなく、薬の直接効果によると考えることができる。これに対し、 $Y(D=0, M(D^*=0))=0$ 、 $Y(D=0, M(D^*=1))=1$ 、 $Y(D=1, M(D^*=1))=1$ であるのなら、イベント発生は、薬の直接効果ではなく、mediator の変化 ($M(D^*=0) \rightarrow M(D^*=1)$) を介した間接効果であると考えることができる。より一般的には $Y(D=0, M(D^*=0))=Y(D=0, M(D^*=1)) \neq Y(D=1, M(D^*=1))$ であるのなら直接効果、 $Y(D=0, M(D^*=0)) \neq Y(D=0, M(D^*=1))=Y(D=1, M(D^*=1))$ なら間接効果と考えることができる。同様に $Y(D=1, M(D^*=1))=Y(D=1, M(D^*=0)) \neq Y(D=0, M(D^*=0))$ が推測されるのなら直接効果、 $Y(D=1, M(D^*=1)) \neq Y(D=1, M(D^*=0))=Y(D=0, M(D^*=0))$ なら間接効果と考えることができる。しかし、一人の対象者は（同時刻には）薬は使用するかしないかのいずれかであり、実際に観察可能なのは $Y(D=0, M(D^*=0))$ と $Y(D=1, M(D^*=1))$ のうち、いずれか一つであり、当然のことながら薬非使用の患者では使用時の、薬を投与された患者では非使用時のイベントは（同時には）観察できない。「直接効果」、「間接効果」の推定では、さらに、そもそも実際には観察することが全く不可能な $Y(D=0, M(D^*=1))$ と $Y(D=1, M(D^*=0))$ の2つを加えた4つの状態（曝露の有無と曝露/非曝露時の mediator の値の組み合わせ）におけるイベント発生の有無

を想定する。すなわち、「直接効果」「間接効果」は現実には観察されえない状態を仮想する「反事実仮想 (counter-factual)」のアプローチによる因果モデル (causal model) を用いることが必要と考えられている。

2011年にはLangeら⁵⁾は、周辺構造モデル (marginal structural model) を用いて比較的簡単に「直接効果」と「間接効果」を推定する方法を発表した。本報告ではLangeの方法を用いてスタチンの脳梗塞予防に対する「直接効果」と「間接効果」を推定する。Langeらの方法を含め、推定しようとするのは”natural”な「直接効果」と「間接効果」である。”natural”な「間接効果」は $Y(D=0, M(D^*=0))$ と $Y(D=0, M(D^*=1))$ の差から求められ、”natural”な「直接効果」は $Y(D=0, M(D^*=1))$ と $Y(D=1, M(D^*=1))$ の差から求められる。ここで、 $M(D^*=0)$ と $M(D^*=1)$ は薬非使用時と使用時に個別の対象者に（”natural”に）見られる mediator の値であるが、「mediator の値を薬使用/非使用にかかわらず、操作可能」とする想定も可能である。「対象者全員に薬使用/非使用の有無によらず、mediator の値を操作し、特定の値 m に固定することが可能」との想定に立つ”controlled”の効果を求めるアプローチも可能である。たとえば”controlled”の「直接効果」（”controlled direct effect”）が存在するのは $Y(D=0, m)$ と $Y(D=1, m)$ が異なる時である。しかし、ここでは、”controlled”ではなく”natural”な「直接効果」と「間接効果」を求める。

3. 周辺構造モデルの利用

臨床試験でも観察研究でも同一の対象者における薬の非使用時におけるイベント発生 $Y(D=0, M(D^*=0))$ と使用時におけるイベント発生 $Y(D=1, M(D^*=1))$ を観察することは

きないので、薬を使用しなかった集団におけるイベント発生の平均 $E(Y[(D=0), M(D^*=0)])$ と薬を使用した集団におけるイベント発生の平均 $E(Y[(D=1), M(D^*=1)])$ を求めて比較する。観察研究では通常、両集団における曝露以外の因子の分布は異なるので、層別化や多変量解析などにより「調整」を行う。

これに対し、観察することが全く不可能な $Y[(D=0), M(D^*=1)]$ あるいは $Y[(D=1), M(D^*=0)]$ の合理的な推定は可能なのだろうか？ Lange らの方法⁵⁾では、個々の対象者における $Y[(D=0), M(D^*=1)]$ 、 $Y[(D=1), M(D^*=0)]$ を推測することは不可能であっても、 $M(D^*)$ の分布が推定可能なら、集団におけるイベント発生の平均 $E(Y[(D=0), M(D^*=1)])$ 、 $E(Y[(D=1), M(D^*=0)])$ を、周辺構造モデルを用いて推定可能であることを利用して間接効果と直接効果を推定する。

初めに行うのは、 D (薬の使用/非使用) と $baseline$ の共変量の値 (C) と (観察終了直前またはイベント発生直前の) と $mediator$ の値 (M) の実測のデータを用いて D と C から M を推定するモデルの作成である。推定モデルとしては、 M の値が正常/高値など 2 値をとるのならロジスティックモデルなどを用いることができる。高/中/低などの 3 値以上の有限の数の値をとるのであれば、“multinomial” のロジスティックモデルが利用できる。血清脂質などを連続量として扱う場合は、正規分布などの分布を仮定し、一般化線形モデルなどを用いてモデルのパラメータ値を推定することができる。

M を推定するモデルを用いることにより、それぞれの患者における、「観察された交絡

因子 C をもつ患者が観察された曝露の状態 D にある時に、観察された $mediator$ の値が得られる確率」とともに、「観察された交絡因子 C をもつ患者が観察された曝露とは逆の状態 D^* にある時に、観察された $mediator$ の値が得られる確率」を計算することができる。

$Mediator$ が 2 値変数 (例: 正常/高値) の場合には、第 i 番目の患者の年齢や性や $baseline$ などの交絡因子 (観察開始時の血清脂質の値を含む) C_i と観察された曝露の状態 D_i から、 $mediator$ が「正常」($M=1$) の期待値が 0.3 と予測され、実測の $mediator$ の値 (M_i) が 1 (正常) なら $P^*(M=M_i | D=D_i, C=C_i) = P^*(M=1 | D=D_i, C=C_i) = 0.3$ であり、実測の $mediator$ の値が 0 である (高値) なら $P^*(M=M_i | D=D_i, C=C_i) = P^*(M=0 | D=D_i, C=C_i) = 0.7$ である。年齢や性などの C と観察された曝露と逆の状態 D^* から、 $mediator$ が正常化すること ($M=1$) の期待値が 0.6 と予測されるのなら、実測の $mediator$ の値が 1 (正常化) なら $P(M=1 | D=D^*_i, C=C_i) = 0.6$ であり、実測の $mediator$ の値が 0 である (正常化しなかった) のなら $P(M=0 | D=D^*_i, C=C_i) = 0.4$ である。

$Mediator$ が 3 つの値 (例: 高/中/低) をとるのなら、multinomial のロジスティック回帰により、標準となる状態 (例: 低) に対する他の二つの状態の (例: 高と中) のオッズ比 (OR) を求めることができる。たとえば第 i 番目の患者の年齢や性などの交絡因子 C_i と観察された曝露の状態 D_i から、「中」の「低」に対する $OR_{iD_{中}}$ と「高」の「低」に対する $OR_{iD_{高}}$ を求めることができる。これらの値から、 C_i と D_i の患者の「高」「中」「低」の確率はそれぞれ、

$$P(M = \text{"高"}|D = D_i, C = C_i) \\ = \frac{OR_{iD\text{高}}}{1 + OR_{iD\text{中}} + OR_{iD\text{高}}}$$

$$P(M = \text{"中"}|D = D_i, C = C_i) \\ = \frac{OR_{iD\text{中}}}{1 + OR_{iD\text{中}} + OR_{iD\text{高}}}$$

$$P(M = \text{"低"}|D = D_i, C = C_i) \\ = \frac{1}{1 + OR_{iD\text{中}} + OR_{iD\text{高}}}$$

と与えられ、交絡因子が C_i だが曝露については観察値とは逆の D_i^* の場合に予測される「高」「中」「低」の確率はそれぞれ

$$P^*(M = \text{"高"}|D = D_i^*, C = C_i) \\ = \frac{OR_{iD^*\text{高}}}{1 + OR_{iD^*\text{中}} + OR_{iD^*\text{高}}}$$

$$P^*(M = \text{"中"}|D = D_i^*, C = C_i) \\ = \frac{OR_{iD^*\text{中}}}{1 + OR_{iD^*\text{中}} + OR_{iD^*\text{高}}}$$

$$P^*(M = \text{"低"}|D = D_i^*, C = C_i) \\ = \frac{1}{1 + OR_{iD^*\text{中}} + OR_{iD^*\text{高}}}$$

と与えられる。

次に Mediator が正規分布に従う連続量であると仮定するモデルを使う場合を考える。第 i 番目の患者における年齢や性などの交絡因子 C_i と実際の曝露の状態 D_i から計算される mediator の期待値を $E(m|D_i, C_i)$ 、交絡因子 C_i と観測された曝露と逆の状態 D_i^* から計算される期待値を $E(m|D_i^*, C_i)$ とする。実測値が M であるのなら、標準偏差 σ (D の値によらず一定であることを仮定) を使って期待値が $E(m|D_i, C_i)$ と $E(m|D_i^*, C_i)$ である時に M の値が観測される正規密度関数の値を P 、 P^* として求める。すなわち

$$P(M = M_i|D = D_i, C = C_i) \\ = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(M_i - E(m|D_i, C_i))^2}{2\sigma^2}\right)$$

$$P^*(M = M_i|D = D_i^*, C = C_i) \\ = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(M_i - E(m|D_i^*, C_i))^2}{2\sigma^2}\right)$$

を計算する。

周辺構造モデルは、自らが受ける曝露の確率を用いて重みを計算し、「擬似集団」(pseudo-population) を作り、擬似集団におけるイベント発生を推定し、因果関係の推定に利用する因果モデルの一つである。

間接効果の推定では、曝露は観察された D_i であるが、mediator は観察されたのと逆の状態の時の確率または確率分布をもつと仮定した時の重みを、以下のように設定する。

$$W_i = \frac{P^*(M = M_i|D = D_i^*, C = C_i)}{P(M = M_i|D = D_i, C = C_i)}$$

たとえば、Mediator が 2 値変数 (例: 正常/高値) で、第 i 番目の患者で実際に測定された mediator の値が $M=1$ (正常) なら $W_i = \frac{0.6}{0.3} = 2$ (第 i 番目の患者は pseudo-population では、「1 人 2 役」)、Mediator が「高」「中」「低」の 3 つの値をとりうるが、第 i 番目の患者の実測の mediator の値が「高」であるのなら、

$$W_i = \frac{P^*(M = \text{"高"}|D = D_i^*, C = C_i)}{P(M = \text{"高"}|D = D_i, C = C_i)}$$

である。また、LDL コレステロールを連続量の mediator と考え、第 i 番目の患者における観察終了直前またはイベント発生直前の LDL の値が 150mg/dL であるのなら、

$$W_i = \frac{P^*(M = 150|D = D_i^*, C = C_i)}{P(M = 150|D = D_i, C = C_i)}$$

と計算される。

周辺構造モデルを用いる Lange の方法では、曝露の状態が実際とは逆の D_i^* において期

待される個々の対象者における mediator の特定の値を推定するのではない。推定されるのは、曝露が実際と同一の D_i である時及びそれとは逆の値 D^*_i をとる時の「確率分布」である。その確率分布にもとづいて、実際に測定された mediator の値（「正常」、「高」、150mg/dL）が見られる確率（または確率密度）を計算する。同様に曝露が実際と逆の D^*_i である時の各患者のイベント発生 $Y[(D=D_i), M(D=D^*)]$ を推定するわけではなく、各患者に重み (W_i) を与えることによって、mediator が $D=D^*_i$ で見られる集団 (pseudo-population) を構成し、その集団におけるイベント発生の平均 $E(Y[(D=0), M(D^*=1)])$ 、 $E(Y[(D=1), M(D^*=0)])$ を推定する。

4. データの解析

1. サブコホートの抽出率の補正

ケース・コホートデータの解析では、Barlow らの方法⁶⁾にもとづく stratified sampling に対応する Cox 回帰モデルを用いた。Stratified sampling は random sample に対比する概念であり、全コホートを何等かの特徴（たとえば性・年齢）でサブグループに区分し、それぞれのサブグループから観察開始時において異なる抽出割合でサンプルする方法である。本研究ではサブコホートはランダムサンプルによって得られているが、サンプリングエラーにより、性・年齢によって無視できない程度の抽出割合の差がみられたことから、性（男/女）と年齢（65歳未満、65-74歳、75歳以上）で6群に分け、それぞれの抽出割合 F_i ($i=1, 2, \dots, 6$) (range : 0.08~0.14) を求め、サブコホートメンバー404人がこの6群のいずれに属するかによって抽出率の逆

数 ($1/F_i$) で重みづけを行った。

以下に全コホート 3830 人における性・年齢分布、サブコホート 404 人における性・年齢分布と、抽出割合を示す。

表1 性・年齢による6サブグループ： 全コホート

| | Female | Male | Total |
|-------|--------|------|-------|
| <65 | 429 | 673 | 1102 |
| 65-74 | 913 | 864 | 1777 |
| 75- | 528 | 423 | 951 |
| Total | 1870 | 1960 | 3830 |

表2 性・年齢による6サブグループ： サブコホート

| | Female | Male | Total |
|-------|--------|------|-------|
| <65 | 55 | 95 | 150 |
| 65-74 | 71 | 81 | 152 |
| 75- | 58 | 44 | 102 |
| Total | 184 | 220 | 404 |

表3 性・年齢による6サブグループ： 抽出割合

| | Female | Male | Total |
|-------|--------|------|-------|
| <65 | 0.13 | 0.14 | 0.14 |
| 65-74 | 0.08 | 0.09 | 0.09 |
| 75- | 0.11 | 0.10 | 0.11 |
| | 0.10 | 0.11 | 0.11 |

2. 欠測値と multiple imputation

サブコホートとケースについては、観察開始前と観察終了直前（ケースについてはイベント発生直前）の検査値と体重、身長、スタチンの開始日と終了日に関する情報を取得した。検査値については収縮期血圧 (SBP)、拡張期血圧 (DBP)、空腹時血糖 (SB)、ヘモグロビン A1C (HbA1c)、総コレステロール (TC)、トリグリセリド (TG)、HDL コレステロール (HDL)、LDL コレステロール (LDL) の8種類である。観察開始前のこれらの検査値を SBP0、DBP0、BS0、HbA1c0、TC0、TG0、HDL0、LDL0 と定義し、観察終了直前（ケー

スについてはイベント発生直前)のこれらの検査値を SBP99、DBP99、BS99、HbA1c99、TC99、TG99、HDL99、LDL99と定義した。

これらの値と観察開始日、観察終了日に関するサブコホートとケースにおける欠測の状況を以下に示す。サブコホート、ケースともに欠測はどの項目についても10%未満であった。

表4 サブコホート (N=404) における欠測

| | | | | |
|------------|-------|-------|---------|-------|
| | 性 | 年齢 | 体重 | 身長 |
| Missing(N) | 0 | 27 | 29 | FALSE |
| Missing(%) | 0.0 | 6.7 | 7.2 | 0.0 |
| | 観察開始日 | 観察終了日 | | |
| Missing(N) | 0 | 0 | | |
| Missing(%) | 0.0 | 0.0 | | |
| | SBP0 | DBP0 | HbA1c0 | BS0 |
| Missing(N) | 0 | 0 | 0 | 0 |
| Missing(%) | 0.0 | 0.0 | 0.0 | 0.0 |
| | TC0 | TG0 | HDL0 | LDL0 |
| Missing(N) | 0 | 0 | 0 | 0 |
| Missing(%) | 0.0 | 0.0 | 0.0 | 0.0 |
| | SBP99 | DBP99 | HbA1c99 | BS99 |
| Missing(N) | 29 | 29 | 29 | 26 |
| Missing(%) | 7.2 | 7.2 | 7.2 | 6.4 |
| | TC99 | TG99 | HDL99 | LDL99 |
| Missing(N) | 34 | 24 | 26 | 24 |
| Missing(%) | 8.4 | 5.9 | 6.4 | 5.9 |

表5 ケース (N=97) における欠測

| | | | | |
|------------|-------|-------|---------|-------|
| | 性 | 年齢 | 体重 | 身長 |
| Missing(N) | 0 | 0 | 9 | 9 |
| Missing(%) | 0.0 | 0.0 | 9.3 | 9.3 |
| | 観察開始日 | 観察終了日 | | |
| Missing(N) | 0 | 0 | | |
| Missing(%) | 0.0 | 0.0 | | |
| | SBP0 | DBP0 | HbA1c0 | BS0 |
| Missing(N) | 0 | 0 | 0 | 0 |
| Missing(%) | 0.0 | 0.0 | 0.0 | 0.0 |
| | TC0 | TG0 | HDL0 | LDL0 |
| Missing(N) | 0 | 0 | 0 | 0 |
| Missing(%) | 0.0 | 0.0 | 0.0 | 0.0 |
| | SBP99 | DBP99 | HbA1c99 | BS99 |
| Missing(N) | 0 | 0 | 1 | 0 |
| Missing(%) | 0.0 | 0.0 | 1.0 | 0.0 |
| | TC99 | TG99 | HDL99 | LDL99 |
| Missing(N) | 0 | 0 | 0 | 0 |
| Missing(%) | 0.0 | 0.0 | 0.0 | 0.0 |

欠測については、多重補完法 (multiple imputation) により補完した。解析では欠測を補完したデータを用い、SAS の proc MIANALYSE を用いて最終結果を得た。

3. ケース・コホートデータの解析：粗解析

イベント発生率を人-年法を用いて new use、prevalent use、non-use に区分した。スタチンについては、現在日本ではアトルバスタチン、ロスバスタチン、ピタバスタチン、プラバスタチン、シンバスタチン、フルバスタチンの6種類が承認されている。このうち最後に承認されたのはロスバスタチンであり、2005年4月に販売が開始された。本研究では2004年9月から2005年4月まではロスバスタチンが使用されていないが、この期間を除き、対象患者は全てのスタチンを本研究の全研究期間において使用することが可能な状態にあり、本研究でも、6種類のスタチンのいずれかが使用されていた。これら6種類のスタチンの区別は可能であるが、イベント発生数が比較的小さいことから、本研究では、6種類のスタチンを区別せず、全て同じ「スタチン」として扱った。Non-use を reference として率比 (Rate Ratio) を求めた。また、性で層別化して、発生率を求め、対応する性の non-use を reference として率比を求めた。一般に薬の非使用期間と使用期間におけるイベント発生の比較を行う研究で、使用者の使用開始までの期間を除外すると、immortal time bias というバイアスが起る。これは薬を観察開始後新規に使う者 (new user) の使用開始までの期間は本来非使用の分母情報に含めるべきだが、この期間を解析から除外する (または非曝露の期間とすべきである) にもかかわらず、誤って曝露の期間に誤分類することにより、非曝露の期間の発生率が過大評価され (誤分類の場合には、非曝露の期間の

発生率が過大評価されることに加えて曝露期間の発生率が過小評価され、結果として発生率比が過小評価されるために起こるバイアスである。このバイアスを避けるために、薬の new user については、観察開始から曝露（スタチン使用）開始までを non-use の期間と扱い、それ以後を曝露期間と扱った。人年法による粗解析では non-user の観察期間と new user の観察開始から薬開始までの期間の総計を分母情報に含めた。

また、new use、prevalent use の期間のうち、投与中止例については、中止前までの期間と中止後の期間を区別する解析も行った。これは、スタチン投与中止が CVA イベント発生に与える影響を検討するためである。

4. ケース・コホートデータの解析：Cox 回帰による通常の解析

Barlow の方法⁶⁾により、Cox 回帰モデルを用いてスタチンを曝露変数とし、脳血管疾患（cerebrovascular accident、CVA、実質的には脳梗塞）の発生をアウトカムイベントとする通常のケース・コホートデータの解析法により評価した。スタチン使用期間におけるイベント発生をスタチン非使用期間におけるイベント発生と比較した。粗解析においてと同様、immortal time bias を避けるために、薬の new user については、観察開始から曝露開始までを non-use の期間と扱い、それ以後を曝露期間と扱った。性・年齢に関する 6 つのサブグループの互いに異なる抽出率の逆数で重みづけを行った。一般に薬開始後のイベント発生率が薬開始後の時間によって変化する場合には、観察開始時にすでに薬を使用していた薬の継続使用患者（prevalent

users）を除外しないとバイアスを生ずることが知られており、prevalent users を除外する”new user design”が薬剤疫学研究の標準になりつつある⁷⁾が、本研究では、new user の割合が少なく、prevalent use と new use を区別しなかった。ただし、prevalent users では観測開始時にすでにスタチンを使用しており、観察開始時の「ベースライン」の TG、TC、HDL、LDL は、「スタチン使用開始後」の値である点は配慮した。すなわち、本研究では prevalent user を解析に含めたのでベースラインの共変量には、観察開始時の TG、TC、HDL、LDL を含めなかった。スタチンによって直接影響を受けることが考えにくい、血圧（SBP、DBP）、血糖（BS）、HbA1c については、観察開始時の値（SBP0、DBP0、BS0、HbA1c0）をベースラインの共変量として解析に含めた。共変量にはそのほか、性（“sexv”、0：女、1：男）、年齢（“age”）、身長（“bh”）、体重（“bw”）を含めた。欠測値を multiple imputation により補完したデータを用いた。

以下に statin（変数 drug）以外に 8 因子を含めた解析の SAS プログラムを示す。使用される SAS のデータセット “mioutallwithweight”には性・年齢で区分した 6 つの subgroup の抽出率の逆数が重み（“w”）として含まれているほか、多重補完法によって欠測値を補完した観察された値の 5 倍のレコード数（_imputation_=1, 2, ..., 5）が含まれている。使われる変数のうち”studyid”は患者 ID、“eventcva”は CVA イベントの発生の有/無(1/0)を表し、“nonu”は薬非使用なら「1」、users（new または prevalent users）なら「0」の値をとる変数、“PrevalentU”は prevalent user なら「1」それ以外なら「0」の値をとる変数、“scv”はサブコホートメンバーな

ら「1」サブコホート外なら「0」をとる変数、「tend」は観察終了日（ケースではケース発生日）,「w」は性・年齢で区分した6つの subgroup の抽出率の逆数を表す変数である。

SAS プログラム 1 通常の解析

```

data d1; set
sasf.mioutallwithweight;
data d2; set d1;
* サブコホート非ケース、薬非使用;
if eventcva=0 and nonu=1 then do;
drug=0; case=0; starttime=0;
survtime=tend; logw=log(w); output;
end;
* サブコホート非ケース、new user;
if eventcva=0 and newU=1 then do;
drug=0; case=0; starttime=0;
survtime=tdrug;
logw=log(w); output;
    drug=1;
case=0; starttime=tdrug;
survtime=tend; logw=log(w); output;
end;
* サブコホート非ケース、prevalent
user;
if eventcva=0 and PrevalentU=1 then
do; drug=1; case=0; starttime=0;
survtime=tend; logw=log(w); output;
end;
* サブコホート外ケース、薬非使用;
if eventcva=1 and scv=0 and nonu=1
then do; drug=0; case=1;
starttime=tend-0.00005; survtime=t
end; logw=0; output; end;
* サブコホート外ケース、new user;

```

```

if eventcva=1 and scv=0 and newu=1 then
do; drug=1; case=1;
starttime=tend-0.00005; survtime=tend
; logw=0; output; end;
* サブコホート外ケース、prevalent user;
if eventcva=1 and scv=0 and
PrevalentU=1 then do; drug=1; case=1;
starttime=tend-0.00005; survtime=tend
; logw=0; output; end;
* サブコホート内ケース、薬非使用;
if eventcva=1 and scv=1 and nonu=1 then
do; drug=0; case=0; starttime=0;
survtime=tend-0.00005; logw=log(w); ou
tput;
    drug=0; case=1;
starttime=tend-0.00005; survtime=tend
; logw=0; output; end;
* サブコホート内ケース、new user;
if eventcva=1 and scv=1 and newu=1 then
do; drug=0; case=0; starttime=0;
survtime=tdrug; logw=log(w); output;
    drug=1; case=0;
starttime=tdrug; survtime=tend-0.0000
5; logw=log(w); output;
    drug=1; case=1;
starttime=tend-0.00005;
survtime=tend; logw=0; output; end;
* サブコホート内ケース、prevalent user;
if eventcva=1 and scv=1 and
Prevalentu=1 then do; drug=1; case=0;
starttime=0;
survtime=tend-0.00005; logw=log(w); ou
tput;
    drug=1; case=1;
starttime=tend-0.00005; survtime=tend
; logw=0; output; end;

```

```

proc phreg data=d2 covs(aggregate)
outest=outcasecohortUall
covout;title 'case-cohort usual All
Users 4+1 factors ';
model (starttime
survtime)*case(0)=drug age sexv bw
bh SBP0 DBP0 BS0 HbA1c0 /
ties=breslow offset=logw; id
studyid;
by _imputation_;
run;
proc mianalyze
data=outcasecohortUall;
title 'case-cohort usual
mianalyze';
modeleffects drug age sexv bw bh SBP0
DBP0 BS0 HbA1c0 ; run;

```

上記プログラムは Barlow ら⁶⁾が示したプログラムを改変したものであるが、サブコホート外のケースについては、ケース発生直前 (tend-0.00005) からケース発生までの時間間隔のみが解析対象に含まれ、ケースであるサブコホートメンバーについては、観察開始時からケース発生の直前 (tend-0.00005) までの期間とそれ以後ケース発生までの2つの期間に区分する点に特徴がある。このうち抽出率の逆数で重みづけするのはケース発生の直前までであり、ケース発生の直前からケース発生までの期間の重みは1である。ケースについては全例特定されており、抽出率は1であるからである。またサブコホート内の new user のケースについては、抽出率の逆数で重みづけする薬使用開始 (tdrug) までの薬非使用 (drug=0) の期間、同じく抽出率の逆数

で重みづけするその後イベント発生直前までの薬使用 (drug=1) の期間 (tdrug から tend-0.00005)、さらに、その後のイベント発生までの重み1の期間の3つの期間に区分される。またCox回帰 (proc phreg の「covs(aggregate)」はロバスト分散を用いた解析を指定するものである)。使用されたデータは多重補完法によって欠測値を補完した観察された値の5倍のレコード数をもつデータであるが、最後のproc mianalyzeによって補完されたデータからの最終結果が求められる。

5. 間接効果と直接効果の推定

ここでは2と3で叙述した間接効果と直接効果の推定方法にもとづく、SASプログラムを示し、解析方法について説明する。以下に示すのは、プログラム1と同様に prevalent users までをも含め、プログラム1と同じ8因子で解析するためのプログラムである。用いるSASデータセットはプログラム1におけるものと同じであるが、LDL99をmediatorと考える解析である。

SASプログラム2 間接・直接効果: mediator
が正規分布に従うことを仮定

```

-----
data d1; set sasf.mioutallwithweight;
if nonU=0 then drug=1; else drug=0;
* mediator (LDL99)を予測するモデル
もともとの1レコードに対し、D*=0、D*=1
の各2レコードをoutput;
data d2; set d1;
drugstar=0; Mtemp=.;
if drugstar=drug then Mtemp=LDL99;
output;
drugstar=1; Mtemp=.;
if drugstar=drug then Mtemp=LDL99;

```

```

output;
*sc外のケースは推定から外す;
data d3; set d2;
if scv=0 then Mtemp=.;
*重みの分母推定;
data d4; set d3;
Atemp=drug; run;
proc genmod data=d4;
  model Mtemp=Atemp age sexv bw bh
  SBP0 DBP0 hbalc0 bs0 /error=n;
  output out=d5 P=PredM; run;
data d6; set d5;
  weightdir=pdf('normal', LDL99,
  predM, 29.3604);
*重みの分子推定;
data d7; set d6;
atemp=drugstar; run;
proc genmod data=d7;
  model Mtemp=Atemp age sexv bw bh
  SBP0 DBP0 hbalc0 bs0 /error=n;
  output out=d8 P=PredMStar; run;
data d9; set d8;
weightindir=pdf('normal', LDL99,
  predmstar, 29.3604); run;
*pseudo-populationの重み (w2) 推定;
data d10; set d9;
w2=weightindir/weightdir;
data d12; set d10;
* サブコホート非ケース、薬非使用;
if eventcva=0 and nonu=1 then do;
drug=0; case=0; starttime=0;
survtime=tend; logw=log(w)+log(w2)
;output;end;
* サブコホート非ケース、prevalent
user;
if eventcva=0 and Prevalentu=1 then
do; drug=1; case=0; starttime=0;
survtime=tend; logw=log(w)+log(w2);ou
tput;end;
* サブコホート非ケース、new user;
if eventcva=0 and newU=1 then do;
drug=0; case=0; starttime=0;
survtime=tdrug;
logw=log(w)+log(w2);output;
  drug=1; case=0;starttime=tdrug;
survtime=tend; logw=log(w)+log(w2);ou
tput; end;
* サブコホート外ケース、薬非使用;
if eventcva=1 and scv=0 and nonu=1 then
do; drug=0; case=1;
starttime=tend-0.00005;survtime=tend
; logw=log(w2); output;end;
* サブコホート外ケース、prevalent user;
if eventcva=1 and scv=0 and
Prevalentu=1 then do; drug=1; case=1;
starttime=tend-0.00005;survtime=tend
; logw=log(w2); output;end;
* サブコホート外ケース、new user;
if eventcva=1 and scv=0 and newu=1 then
do; drug=1; case=1;
starttime=tend-0.00005;survtime=tend
; logw=log(w2); output;end;
* サブコホート内ケース、薬非使用;
if eventcva=1 and scv=1 and nonu=1 then
do; drug=0; case=0; starttime=0;
survtime=tend-0.00005; logw=log(w)+lo
g(w2);output;
  drug=0; case=1;
starttime=tend-0.00005;survtime=tend
; logw=log(w2);output; end;
* サブコホート内ケース、prevalent user;
if eventcva=1 and scv=1 and

```

```

Prevalentu=1 then do; drug=1;
case=0; starttime=0;
survtime=tend-0.00005; logw=log(w)
+log(w2); output;
  drug=1; case=1;
starttime=tend-0.00005; survtime=t
end; logw=log(w2); output; end;
* サブコホート内ケース、new user;
if eventcva=1 and scv=1 and newu=1
then do; drug=0; case=0;
starttime=0;
survtime=tdrug; logw=log(w)+log(w2
); output;
  drug=1; case=0;
starttime=tdrug; survtime=tend-0.0
0005; logw=log(w)+log(w2); output;
  drug=1; case=1;
starttime=tend-0.00005;
survtime=tend; logw=log(w2);
output; end;
proc phreg data=d12 covs(aggregate)
outest=outcasecohortU covout; title
'case-cohort direct indirect';
model (starttime
survtime)*case(0)=drug drugstar
age sexv bw bh SBP0 DBP0 hbalc0 bs0
/ ties=breslow offset=logw; id
studyid;
by _imputation_; run;
proc mianalyze
data=outcasecohortU;
title 'case-cohort usual
mianalyze';
modeleffects drug drugstar age sexv
bw bh SBP0 DBP0 hbalc0 bs0 ; run;
-----

```

上記のプログラムではデータセット d2 は元のデータの 1 レコードについて D* (変数名 "drugstar") の値が D*=0 と D*=1 の 2 レコードが output され、元のデータの 2 倍のレコードが作られる。2 レコードのうち D と D* が一致する 1 レコードのみについて Mtemp という変数に mediator の値を代入する。Mediator の予測モデルは全コホートを代表するサブコホートから作成することが必要であり、サブコホート外の患者 (scv=0、すべてケース) については Mtemp を欠測 (.) とするデータセットを d3 として作成する。SAS は一般に欠測データをもつレコードを無視して解析を実施するので、この指定により、mediator の推定はサブコホートメンバーのみについて実測の D (スタチンの有無) と共変量と測定された LDL99 を用いて推定が行われる。実際の曝露の状態 (drug) を用いて正規分布を前提に (error=n)、mediator (LDL99) の分布の予測モデルのパラメータが proc gemod data=d4 で特定され、期待値が PredM に記録される。期待値については、Mtemp が欠測であるか否かにかかわらず推定されるので、サブコホート外のケースを含めて、全レコードに対して計算される。期待値が PredM で分散が 29.3 (genmod の結果に示される "Scale" から読み取り) の正規分布において観察された LDL99 の値が得られる確率密度の大きさが weightdir という変数に記録される。元の 1 レコードから作られる 2 レコードで drug の値は同一なので weightdir も同一の値である。次いで D* を用いて mediator (LDL99) の分布の予測モデルのパラメータが proc gemod data=d7 で特定される。パラメータは Mtemp が欠測でない (D と D* の値が同一の) レコードから特定されるので、推定されるパラメータは、proc gemod data=d4 で特

定されるものと全く同一である。期待値が PredMStar という変数に記録され、その期待値と分散が 29.3 の正規分布において観察された LDL99 の値が得られる確率密度の大きさが weightindir という変数に代入される。期待値は、Mtemp が欠測であるか否かにかかわらず推定されるが、proc gemod data=d4 と異なり、曝露の状態は D* で与えられる。元の 1 レコードから作られる 2 レコードのうち 1 つのレコードで D* は D (変数名 "drug") と等しいが、他の 1 レコードで D* は D と異なる値である。PredMStar と weightindir は D と等しい値の D* についてはそれぞれ PredM、weightdir と同一の値をとるが、D と異なる D* については、PredMStar と weightindir は PredM と weightdir とは異なる。したがって data d10 で計算される重み (w2) は、D と一致する D* については "1" となるが、D と一致しない D* については pseudo-population を作るための重みが計算される。

Pseudo-population のための重み計算の後のケース・コホートデータの解析の手順はプログラム 1 と同じであるが、ケース発生直前かまでの期間についての重みは w ではなく $w \cdot w2$ (log スケールでは $\log(w) + \log(w2)$) である。ケース発生直前からケース発生までの重みは 1 ではなく $\log(w2)$ であるが、w2 は観察された元の集団では "1" であるので、観察された元の集団におけるケースについては、ケース発生直前からケース発生までの重みは 1 となる。ロバスト分散を用いた解析を行い (「 covs(aggregate) 」)、proc mianalyze によって補完されたデータか

らの最終結果が求める手順はプログラム 1 と同一である。

上記、プログラム 2 では、mediator は正規分布に従うことが仮定されていたが、以下のプログラム 3 では、LDL99 の値を (<100mg/dL)、(≥ 100 , <140mg/dL)、(≥ 140 mg/dL) の 3 つのカテゴリに分け、mediator の値を multinomial のロジスティック回帰を用いて予測する。

SAS プログラム 3 間接・直接効果: mediator が multinomial logistic 分布に従うことを仮定

```
-----
data d1; set sasf.mioutallwithweight;
if nonu=0 then drug=1; else drug=0;
*Mediator (LDL99) を3Categoryに区分;
data d1; set d1;
if LDL99<100 then LDL99cat=1;
else if LDL99<140 then LDL99cat=2;
else LDL99cat=3;
data d2; set d1;
drugstar=0; Mtemp=. ;
if drugstar=drug then Mtemp=LDL99cat;
output;
drugstar=1; Mtemp=. ;
if drugstar=drug then Mtemp=LDL99cat;
output;
*sc外のケースは推定から外す;
data d3; set d2;
if scv=0 then Mtemp=. ;
*multinomial logistic 回帰による推定;
*/重み分母推定;
data d4; set d3;
Atemp=drug;
run;
proc logistic data=d4;
model Mtemp=Atemp age sexv bw bh SBPO DBPO
```



```

hbalc0 bs0 / link=glogit;
  output out=d5 P=PredM;
run;
  *実測のカテゴリの推定値を採用;
data d6; set d5;
  weightdir=PredM;
  if _LEVEL_=LDL99Cat then output;
*/重み分子推定;
data d7; set d6;
atemp=drugstar;
run;
proc logistic data=d7;
  model Mtemp=Atemp age sexv bw bh SBPO DBPO
hbalc0 bs0 / link=glogit;
output out=d8 P=PredMStar;
run;
  *実測のカテゴリの推定値を採用;
data d9; set d8;
weightindir=PredMStar;
  if _LEVEL_2=LDL99Cat then output;
run;
*/重み計算;
data d10; set d9;
w2=weightindir/weightdir;
*/ケース・コホート観察期間の処理;
data d12; set d10;
  * サブコホート非ケース、薬非使用;
if eventcva=0 and nonu=1 then do;
drug=0; case=0; starttime=0;
survtime=tend;logw=log(w)+log(w2)
;output;end;
  * サブコホート非ケース、prevalent
user;
if eventcva=0 and Prevalentu=1 then
do; drug=1; case=0; starttime=0;
survtime=tend;logw=log(w)+log(w2)

```

```

;output;end;
  * サブコホート非ケース、new user;
if eventcva=0 and newU=1 then do;
drug=0; case=0; starttime=0;
survtime=tdrug;
logw=log(w)+log(w2);output;
  drug=1; case=0;starttime=tdrug;
survtime=tend;logw=log(w)+log(w2);ou
tput; end;
  * サブコホート外ケース、薬非使用;
if eventcva=1 and scv=0 and nonu=1 then
do; drug=0; case=1;
starttime=tend-0.00005;survtime=tend
; logw=log(w2); output;end;
  * サブコホート外ケース、prevalent user;
if eventcva=1 and scv=0 and
Prevalentu=1 then do; drug=1; case=1;
starttime=tend-0.00005;survtime=tend
; logw=log(w2); output;end;
  * サブコホート外ケース、new user;
if eventcva=1 and scv=0 and newu=1 then
do; drug=1; case=1;
starttime=tend-0.00005;survtime=tend
; logw=log(w2); output;end;
  * サブコホート内ケース、薬非使用;
if eventcva=1 and scv=1 and nonu=1 then
do; drug=0; case=0; starttime=0;
survtime=tend-0.00005;logw=log(w)+lo
g(w2);output;
  drug=0; case=1;
starttime=tend-0.00005;survtime=tend
;logw=log(w2);output; end;
  * サブコホート内ケース、prevalent user;
if eventcva=1 and scv=1 and
Prevalentu=1 then do; drug=1; case=0;
starttime=0;

```

```

survtime=tend-0.00005; logw=log(w)
+log(w2); output;
  drug=1; case=1;
starttime=tend-0.00005; survtime=t
end; logw=log(w2); output; end;
* サブコホート内ケース、new user;
if eventcva=1 and scv=1 and newu=1
then do; drug=0; case=0;
starttime=0;
survtime=tdrug; logw=log(w)+log(w2
); output;
  drug=1; case=0;
starttime=tdrug; survtime=tend-0.0
0005; logw=log(w)+log(w2); output;
  drug=1; case=1;
starttime=tend-0.00005;
survtime=tend; logw=log(w2);
output; end;
proc phreg data=d12 covs(aggregate)
outest=outcasecohortU covout; title
'case-cohort direct indirect';
model (starttime
survtime)*case(0)=drug drugstar
age sexv bw bh SBP0 DBP0 hba1c0 bs0
/ ties=breslow offset=logw; id
studyid;
by _imputation_; run;

proc mianalyze data=outcasecohortU;
title 'case-cohort usual mianalyze';
modeleffects drug drugstar age sexv bw bh
SBP0 DBP0 hba1c0 bs0; run;
-----

```

上記のプログラム 3 でも元になるデータセットはプログラム 1、プログラム 2 で用いているものと同一であるが、mediator (ここ

では LDL99) の値を 3 つのカテゴリに区分する (変数: LDL99cat)。プログラム 2 と同様、元のデータの 1 レコードについて $D^*=0$ と $D^*=1$ の元のデータの 2 倍のレコードが作られ、2 レコードのうち D と D^* が一致する 1 レコードのみについて Mtemp という変数に mediator の値 (3 つのカテゴリを示す 1, 2, 3) が示される。プログラム 2 と同様サブコホート外の患者 (scv=0、すべてケース) については Mtemp を欠測 (.) とし、mediator の推定モデルにはサブコホートメンバーのデータのみを利用する。推定には multinomial のロジスティック回帰を利用 (proc logistic / link=glogit) 実際の曝露の状態 (D 、変数名 "drug") を用いて期待値が PredM に記録される。この場合には、期待値として、各レコードについて 3 つの値の確率 (3 つのカテゴリのそれぞれをとる確率、合計=1) が記録されるが、このうち、実測の LDL99 のカテゴリのものをとりだし (if _LEVEL_=LDL99Cat then output) weightdir という重みの分母とする。次いで D^* を用いて mediator (LDL99cat) の分布の予測モデルのパラメータが multinomial のロジスティック回帰で推定されるが、推定されるパラメータは、プログラム 2 と同様、実際の曝露の状態 D を用いて特定するものと全く同一である。期待値 (3 つのカテゴリに対応する確率) が PredMStar という変数に記録され、実測の LDL99 と同一カテゴリのものを取り出し (if _LEVEL_2=LDL99Cat then output)、weightindir という重みの分子とする。計算される重み ($w2$) は、 D と一致する D^* については "1" となるが、 D と一致しない D^* については pseudo-population を作るための重みが計算される点はプログラム 2 と同じである。

Pseudo-population のための重み計算の後のケース・コホートデータの解析の手順はプログラム 3 と全く同一である。

6. 観察終了日、観察開始日における mediator の値とイベント発生の関連

また、どの値を目標にして治療すべきか？という臨床的なプラグマティックな問いに対する回答を得る上で参考となる情報を示すことを目的とし、観察終了日直前の mediator の値 (TG99、TC99、HDL99、LDL99) と曝露の有無で区分したグループにおけるイベント発生率を人-年法で求めた。発生率については、prevalent user を含めた発生率のほか、new user に限った発生率、prevalent user に限った発生率も求めた。

さらに、曝露の有無と mediator の変化 (TG99-TG0、TC99-TC0、HDL99-HDL0、LDL99- LDL0) を 3 つのカテゴリに分け、それぞれのグループにおけるイベント発生率を人一年法で求めた。

発生率の算出にあたっては、それぞれのグループに含まれるサブコホートメンバーをさらに 6 つの性・年齢グループにわけ、それぞれのグループに関する抽出率の逆数をかけて、全コホートにおける人-年を推定し、イベント数をその推定人-年でわって算出し、この算出方法で求めたイベント発生率を粗の発生率とした。

またグループ間で性・年齢分布が異なりうるので、3080 人の全コホートを標準集団とし、各グループにおける 6 つの性・年齢分布が全コホートと等しいと仮定したときの発生率を性・年齢調整の発生率として算出した。

C. 研究結果

1. 粗解析の結果

3830 人の 2 型糖尿病の患者 (男性 : 1960 人、登録時年齢 66.9 ± 10.0 歳、女性 : 1870 人、登録時年齢 69.7 ± 8.7 歳) が登録された。3830 人の全コホートから 404 人のサブコホートが抽出された。スタチンの使用期間については、全コホートからではなく、サブコホートとケースから情報を得た。以下に new users, prevalent users, non-users の観察期間の分布を示す。New users か prevalent users かによらず、スタチン使用者は男性が半分弱であるが、non-user では男性が 6 割以上を占めていた。年齢に関しては、スタチン使用者のうち、new users は prevalent users より 4.5 歳若く、prevalent users は非使用者より 1.7 歳若かった。また、スタチン開始からの期間は new users で 5.63 年、prevalent users では 5.12 年であるが、その約 3 割で観察終了前に使用を中止していた。

表 6 サブコホート (N=404) のスタチンの使用による区分

| | New Users | Prevalent Users | Non-Users | Total |
|---------------------------------|-------------|-----------------|-------------|-------------|
| N | 89 | 121 | 194 | 404 |
| 男性 (%) | 47.2 | 47.9 | 61.9 | 54.5 |
| 年齢 (平均 (SD)) | 62.4 ± 11.2 | 66.9 ± 10.3 | 68.6 ± 11.6 | 66.7 ± 11.3 |
| 観察開始から薬開始までの平均 (年) | 1.48 | 0.00 | - | - |
| 観察終了前の中止 (N) | 27 | 36 | - | - |
| 観察終了前の中止 (%) | 30.3 | 29.8 | - | - |
| 薬中止から観察終了までの平均 (年) ^a | 2.16 | 3.04 | - | - |
| 薬開始から観察終了までの平均 (年) | 4.15 | 5.12 | - | - |
| 観察開始から観察終了までの平均 (年) | 5.63 | 5.12 | 5.00 | 5.18 |

^a 薬を観察終了前に中止した人における平均

観察期間中に 97 件の CVA イベントが発生した。このうち 66 件はスタチン非使用者に、31 件はスタチン使用者 (New user 7 件、prevalent user 24 件) であった。

サブコホートのスタチン使用区分 (new

users、prevalent users、non-use を構成するスタチン非使用者では CVA イベントの粗発生率は 1000 人年あたり 6.35 件であるのに対しスタチン使用者では 1000 人年あたり 3.34 件であり、粗の率比(Rate Ratio)は 0.53 (95%信頼区間: 0.43-0.64)、new users に限ると 0.33 (0.26-0.42)、prevalent user で 0.64 (0.51-0.80) であった。

表 7 CVA イベント: 粗発生率と Rate Ratio

| | N (subcohort) | Pt-yr (subcohort) | Pt-yr (whole cohort) ^a | CVA cases (N) | Rate (/1000 Pt-yr) | Rate Ratio |
|----------------------|---------------|-------------------|-----------------------------------|---------------|--------------------|------------------|
| New Use | 89 | 369.1 | 3343.7 | 7 | 2.09 | 0.33 (0.15~0.72) |
| Prevalent Use | 121 | 619.6 | 5927.1 | 24 | 4.05 | 0.64 (0.40~1.02) |
| All Use | 210 | 988.7 | 9270.8 | 31 | 3.34 | 0.53 (0.34~0.81) |
| Non-Use ^b | 194 | 1102.1 | 10392.9 | 66 | 6.35 | ref |

^a サブコホートの Pt-yr から表 3 の抽出率の逆数をかけて算出した全コホートの人年
^b Non-Use の期間には New user の観察開始から薬開始までの期間を含む

性と投与中止前と投与中止後の期間を区別した詳細な粗発生率を以下に示す。女性の prevalent use におけるイベント発生は non-use とそれほど変わらないが、男性の prevalent use、女性・男性の new use において発生率は低い。New use では中止後の発生率が高いのに対し、prevalent use では中止後の発生率はむしろ低いが、特に new use では信頼区間が広く、結果は不安定である。

表 8 CVA イベント: 粗発生率と Rate Ratio (性・投与中止前後の期間に区分)

| | N (sub-cohort) | Pt-yr (subcohort) | Pt-yr (whole cohort) ^a | CVA cases (N) | Rate (/1000 Pt-yr) | Rate Ratio (vs Non-Use) |
|----------------------------|----------------|-------------------|-----------------------------------|---------------|--------------------|-------------------------|
| New Use | | | | | | |
| Female | | | | | | |
| On Tx | 29 | 145.2 | 382.8 | 2 | 1.43 | 0.30 (0.07~1.29) |
| Post-dose | 18 | 39.3 | 1779.5 | 3 | 7.84 | 1.65 (0.49~5.53) |
| sum | 47 | 184.6 | 1409.6 | 5 | 2.81 | 0.59 (0.22~1.57) |
| Male | | | | | | |
| On Tx | 33 | 165.4 | 154.7 | 2 | 1.42 | 0.19 (0.05~0.78) |
| Post-dose | 9 | 19.1 | 1564.2 | 0 | 0.00 | 0 |
| sum | 42 | 184.5 | 2806.3 | 2 | 1.28 | 0.17 (0.04~0.70) |
| Sum | | | | | | |
| On Tx | 62 | 310.7 | 537.4 | 4 | 1.43 | 0.22 (0.08~0.62) |
| Post-dose | 27 | 58.4 | 2519.4 | 3 | 5.58 | 0.88 (0.28~2.80) |
| Prevalent Use | | | | | | |
| Female | | | | | | |
| On Tx | 37 | 236.2 | 790.1 | 2 | 5.95 | 1.25 (0.65~2.43) |
| Post-dose | 26 | 75.9 | 3309.5 | 17 | 2.53 | 0.53 (0.12~2.27) |
| sum | 63 | 312.2 | 2380.2 | 7 | 5.14 | 1.08 (0.57~2.05) |
| Male | | | | | | |
| On Tx | 48 | 273.8 | 257.4 | 0 | 2.97 | 0.39 (0.18~0.87) |
| Post-dose | 10 | 33.7 | 2617.6 | 7 | 0.00 | 0 |
| sum | 58 | 307.4 | 4879.6 | 22 | 2.67 | 0.35 (0.16~0.79) |
| Sum | | | | | | |
| On Tx | 85 | 510.0 | 1047.5 | 2 | 4.51 | 0.71 (0.44~1.15) |
| Post-dose | 36 | 109.6 | 4422.6 | 21 | 1.91 | 0.30 (0.07~1.23) |
| Non-Use^b | | | | | | |
| Female | 74 | 437.8 | 4422.6 | 21 | 4.75 | ref |
| Male | 120 | 664.3 | 5970.3 | 45 | 7.54 | ref |
| Total | 194 | 1102.1 | 10392.9 | 66 | 6.35 | ref |

^a サブコホートの Pt-yr から表 3 の抽出率の逆数をかけて算出した全コホートの人年
^b Non-Use の期間には New user の観察開始から薬開始までの期間を含む

2. 通常のケース・コホートデータの解析

以下にスタチン使用の CVA イベント発生に対するハザード比を通常のケース・コホート解析 (プログラム 1) で求めた結果を示す。

表 9 CVA イベント: Cox 回帰による解析結果 (プログラム 1)

| スタチンのCVAイベントに対する効果 (全ユーザー) | | |
|----------------------------|-------------------|-------|
| | 曝露+8変数 | |
| | ハザード比 (95%信頼区間) | P値 |
| スタチン | 0.57 (0.35, 0.94) | 0.027 |
| 年齢(1歳あたり) | 1.03 (1.00, 1.05) | 0.065 |
| 男(女性に対して) | 1.23 (0.74, 2.04) | 0.560 |
| 体重(1Kgあたり) | 1.01 (0.98, 1.04) | 0.545 |
| 身長(1cmあたり) | 0.99 (0.95, 1.04) | 0.890 |
| SBP0(1mmHgあたり) | 1.00 (0.98, 1.02) | 0.717 |
| DBP0(1mmHgあたり) | 1.00 (0.97, 1.02) | 0.878 |
| BS0(1mg/dLあたり) | 1.00 (1.00, 1.00) | 0.919 |
| HbA1c0(1%あたり) | 1.16 (0.96, 1.38) | 0.124 |

粗解析と同様にスタチン使用者では CVA 発生が 0.57 倍になることが示された。また、高齢は CVA 発生のリスク因子となりうることも示唆された。

3. 直接・間接効果の推定: mediator の分布が正規分布であることを仮定するモデル

表 10 から表 13 に mediator が正規分布をすることを仮定した (LDL についてはプログラム 2 による) 直接・間接効果の推定結果を示す。表 10 LDL を介する間接効果と直接効果 (LDL が正規分布をすることを仮定)

| スタチンのCVAイベントに対する LDLを介した直接・間接効果 | | |
|---------------------------------|-------------------|--------|
| | 曝露+8変数 | |
| | ハザード比 (95%信頼区間) | P値 |
| スタチン (直接効果) | 0.57 (0.35, 0.93) | 0.0256 |
| スタチン (間接効果) | 1.00 (0.98, 1.02) | 0.7649 |
| 年齢(1歳あたり) | 1.03 (1.00, 1.06) | 0.0573 |
| 男(女性に対して) | 1.16 (0.64, 2.09) | 0.6186 |
| 体重(1Kgあたり) | 1.01 (0.98, 1.04) | 0.5298 |
| 身長(1cmあたり) | 1.00 (0.95, 1.05) | 0.9135 |
| SBP0(1mmHgあたり) | 1.00 (0.98, 1.02) | 0.715 |
| DBP0(1mmHgあたり) | 1.00 (0.97, 1.03) | 0.9324 |
| HbA1c0(1%あたり) | 1.16 (0.96, 1.39) | 0.1218 |
| BS0(1mg/dLあたり) | 1.00 (1.00, 1.00) | 0.9215 |