様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月 １３ 日現在

機関番号：１４３０１
研究種目：基盤研究(C)（一般）
研究期間：2020～2022
課題番号：２０Ｋ１０３７６

研究課題名（和文）機械学習による経時的なQOL変化、及び質調整生存年(QALY)の予測に関する研究

研究課題名（英文）Estimating quality of life changes based on machine learning methods

研究代表者
山本　洋介（Yosuke, Yamamoto）

京都大学・医学研究科・教授

研究者番号：３０５８３１９０

交付決定額（研究期間全体）：（直接経費）　　3,300,000円

研究成果の概要（和文）：本研究の目的は、１）一般住民集団に基づくコホートに基づき、機械学習等を用いた一時点の効用値（ならびに経時的変化）推定の可能性を探ること、２）新型コロナウイルス感染症流行下におけるわが国のQOLの記述や様々な臨床疫学研究を行うことである。
１）では、既存のコホート等に含まれる変数を説明変数、従属変数を１）一時点での効用値２）1年後の効用値の変化とした様々な機械学習モデルを比較検討した。過学習が一部認められたものの、概ね使用に耐えうる推定アルゴリズムが得られた。
２）では、高齢者における孤立とワクチン接種忌避との関連性、新型コロナウイルスへの感染状況とそう痒との関連性について、新たな知見を得た。

研究成果の学術的意義や社会的意義
本研究の結果、既存のコホートに含まれる変数から、機械学習を活用した一時点での効用値の推定が一定の精度でもって可能であることが明らかとなった。また効用値の経時的変化の推定にも拡張して、やや精度は劣るものの推定の可能性を示した点で意義があると考える。さらには、前の課題から継続してコホートデータ構築に取り組むという連続性のある課題設定の結果、新型コロナウイルス前後におけるQOLや諸問題を精緻に測定、追跡することにも成功した。その結果として、英文原著論文２報が受理（うち1報はすでに掲載済み）されており、本課題終了後もこの一大データベースに基づく知見の継続的な発信が期待できる。

研究成果の概要（英文）：The objectives of this study are 1) to estimate health utilities using machine learning and other methods based on a cohort of the general population, and 2) to describe QOL in Japan during a COVID-19 epidemic and conduct various clinical epidemiological studies.
1) From the variables included in existing cohorts, we estimated 1) health utilities at baseline and 2) changes in health utilities after one year follow-up. Although problems regarding over-learning was observed in several models, we obtained usable estimation algorithms.
2) New findings were obtained on the association between isolation and avoidance of vaccine uptake in the elderly and the association between infection status with COVID-19 and development of pruritus.

研究分野：臨床疫学

キーワード：Quality of Life　患者報告型アウトカム　効用値　機械学習

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

健康関連 QOL は、患者の主観を反映したアウトカム指標として、臨床疫学研究だけではなく臨床試験など医薬品承認申請の際にも重視されるようになった。昨今、医療費の増大により適正な医療資源の分配がますます重要視される、医療経済評価研究が国レベルで推進されている。実際、「中央社会保険医療協議会における費用対効果評価の分析ガイドライン」にも QOL のひとつである効用値に基づく質調整生存年（QALY）での評価が明記されており、正確なアウトカム評価のためにも、患者の効用値を様々な時点で把握することが求められている。とりわけ、経時的な QOL の変化を把握することは今後の患者の予後を多面的に予測する上で重要であると考えた。一方、日本におけるコホート研究およびデータベース研究における効用値評価の現状としては、QOL 評価を含んだコホートやレジストリは多数存在するが、いずれの研究集団もばらばらの尺度により、限られたタイミングのみでの評価がなされていた。そのため、各研究集団間における比較可能性が乏しいだけでなく、その測定頻度も不十分である。すなわち、日本の一般住民集団において、様々な疾患や状態を有する対象における効用値とその経時的な変化について十分な知見があるとは言い難いのが実情であった。

２．研究の目的

本研究の目的は、一般住民集団から得られた大規模なコホートを構築し、従来の診療情報やコホートに含まれるような変数に基づき、機械学習等を用いた効用値の推定が可能かについて検討を行うこと、そしてさらにその経時的変化の推定にも拡張した上で、得られた値について検証を行うこと、以上 2 点に加えて、新型コロナウイルス感染症流行下という特殊事情も鑑み、日本における新型コロナウイルス感染症流行下におけるさまざまな QOL の指標の推移、ならびに縦断的評価も行うことをも目的に加えた。結果として、新型コロナウイルス感染症流行下の状況において、従来の評価法では実現が困難であった多時点での測定に基づき、さまざまな手法を用いて主観に基づく健康状態の推定を試みる点において学術的独自性がある。また、既存のデータベースやコホートに含まれうる変数からどのように効用値が経時的に遷移していくかを知ることができ、その測定方略に新たな知見を与えうるものであると考えられる。

３．研究の方法

１）大規模コホートデータから効用値を推定するアルゴリズムの検討：

コホートやデータベースに一般に含まれる変数から機械学習を始めとする各種モデルにより効用値を推定するための方法を検討した。具体的には、機械学習のモデル構築（後藤）、コホートデータ分析（大前）の専門家とともに、効用値予測を機械学習で行うにあたって適切なモデルの選択、ならびに各パラメータのチューニング設定についても検討を加えた。

２）既存のコホートデータ・データベースから一時点での効用値の推定の試行

実際の既存のコホートデータならびにデータベースに含まれる変数を用いて、上記 1 で検討された手法に基づき、一時点での効用値の推定を試行した。推定する効用値は EQ-5D に基づくものとし、推定値と、観測値の比較を行った。RMSE など統計学的に妥当な指標により推定値を検証した。

３）経時的な効用値の変化の予測

上記１）～２）で、あるベースラインの一時点での効用値の推定がなされた対象者が、その後追加されたコホートに含まれる変数などを含めた上で、各人の効用値変化を予測するアルゴリズムを別に作成し、観測値と推定値の比較を行った。

４）日本における新型コロナウイルス感染症流行下における QOL の指標の推移、ならびに様々な要因との縦断的な関連性の評価

本研究課題はちょうど新型コロナウイルス感染症流行下の時期であったため、その時期の日本の地域・年齢・性別の人口構成比に従ってサンプリングした一般住民集団のプロファイル型尺度での OQL の推移を記述した。また特に行動制限やワクチンなど感染症がもたらす不安の高まる背景において、心の健康や孤立といった要因とワクチン接種忌避との関連性、さらには新型コロナウイルス感染状況とそう痒との関連性についても検討を加えた。

４．研究成果

１）大規模コホートデータから効用値を推定するアルゴリズムの検討：

まず、説明変数として、性・年齢・独居・婚姻・世帯収入・教育歴・就労状況・併存疾患数（高血圧・糖尿病・高脂血症・脳血管疾患・心疾患・呼吸器疾患・消化器疾患・腎疾患・悪性腫瘍）飲酒の有無・Lubben Social Network Scale・新型コロナウイルスに対する不安・新型コロナウイルス感染症への周囲ならびに自らの罹患・SF ツールで測定可能な３つのコンポーネントサマリースコアとした。また、検討するモデルとしては、①least absolute shrinkage and selection operator（Lasso）回帰②ニューラルネットワーク③ランダムフォレスト④勾配ブースティング回帰木⑤重回帰分析の 5 つとした。具体的には、Lasso 回帰にて削減できそうな変数の候補を絞り込み、その上でそれらの変数について各モデルに投入し、学習データ並びにテストデータにて、Root Mean Square Error（RMSE）ならびに R2（決定係数）によりモデルの評価を実施することとした。

２）既存のコホートデータ・データベースから一時点での効用値の推定の試行

回答を得た 3500 人の背景を【表 1】に示す。平均年齢 48.7 歳（標準偏差 16.9）、男性 49.5%、であった。コホートに含まれる変数としては、併存疾患数中央値は 0（四分位範囲 0-1）と健康な一般住民集団であり、また、SF ツールから測定された各サマリースコア、①サマリースコア（身体）である PCS、②サマリースコア（精神）である MCS③サマリースコア（役割）である RCS については、当該サンプルはおおむね国民標準に合致していることが示唆された。なお、効用値の平均は 0.89 であった。

この効用値を推定するために以下のモデルを実行した。

まず、least absolute shrinkage and selection operator（Lasso）回帰にて、コホートに含まれる変数の項目を絞り込むことを期待して、Lasso 回帰を実施した。最終的には年齢カテゴリ、独居の有無、収入カテゴリ、教育年数、併存疾患数、新型コロナウイルス感染症への不安の有無、各サマリースコア、飲酒の有無、家族も含めた新型コロナウイルス感染症の有無以外の変数の係数が 0 になったため、これらの変数をニューラルネットワーク、ランダムフォレスト、勾配ブースティング回帰木、ならびに重回帰分析に投入し、あらかじめ訓練データとテストデータを３：１に分けた上で解析を行った。なおパラメータチューニングにおいては 10 分割の交差検証法を用いたグリッドサーチを行い、過学習の予防に努めた。

【表１】ベースラインの背景因子

背景因子		N=3,493
年齢，歳	mean（SD）	48.7 (16.9)
性別，男性	N（%）	1,729 (49.5%)
独居	N（%）	632 (18.1%)
既婚	N（%）	1,915 (54.8%)
世帯収入	N（%）	
	～３００万円	1,312 (37.6%)
	３００～５００万円	976 (27.9%)
	５００～７００万円	630 (18.0%)
	７００～１０００万円	314 (9.0%)
	１０００～１２００万円	163 (4.7%)
	１２００万円以上	98 (2.8%)
教育年数，年	median（IQR）	14.0 (12.0-16.0)
併存疾患数，個	median（IQR）	0.0 (0.0-1.0)
新型コロナウイルスへの不安	N（%）	2,822 (80.8%)
新型コロナウイルス感染の既往（家族含む）	N（%）	55 (1.6%)
過去1年間の仕事の状況	N（%）	
	変わりなし	1,840 (52.7%)
	失業/休職の経験あり	202 (5.8%)
	もともと仕事をしていない	1,311 (37.5%)
	新たに就職した	140 (4.0%)
飲酒の有無	N（%）	1,441 (41.3%)
効用値	mean（SD）	0.89 (0.14)
LSNS	mean（SD）	9.6 (6.3)
サマリースコア（身体）	mean（SD）	53.1 (10.5)
サマリースコア（精神）	mean（SD）	46.3 (11.3)
サマリースコア（役割）	mean（SD）	49.3 (11.4)

最終的に訓練データならびにテストデータでの推定結果を【表 2】に示す。過学習の予防にはつとめたものの、ランダムフォレストならびに勾配ブースティング回帰木では過学習を示唆する結果となった。

３）経時的な効用値の変化の予測

1 年後の効用値の変化を推定することを目的として、1 年間追跡できた対象者 1,884 名を対象に、上記の変数に加えて 1 年後に収集した同様の変数を追加してモデルを構築した。最終的に訓練データならびにテストデータでの推定結果を【表 3】に示す。一時点での推定と同様、経時的な変化の推定においても、全体的に過学習を示唆する結果となった。

【表２】一時点での効用値推定におけるモデルの比較

	訓練データ			テストデータ		
	RMSE	R2	MAE	RMSE	R2	MAE
重回帰分析	0.092	0.577	0.063	0.099	0.561	0.065
Lasso回帰	0.092	0.577	0.062	0.099	0.561	0.064
ニューラルネットワーク	0.091	0.584	0.062	0.098	0.569	0.064
ランダムフォレスト	0.044	0.924	0.029	0.101	0.544	0.067
勾配ブースティング回帰木	0.041	0.922	0.031	0.101	0.489	0.071

【表３】1年間での効用値の変化の推定におけるモデルの比較

	訓練データ			テストデータ		
	RMSE	R2	MAE	RMSE	R2	MAE
重回帰分析	0.081	0.548	0.056	0.086	0.377	0.061
Lasso回帰	0.081	0.545	0.056	0.085	0.382	0.06
ニューラルネットワーク	0.071	0.65	0.052	0.088	0.354	0.062
ランダムフォレスト	0.041	0.939	0.026	0.085	0.385	0.059
勾配ブースティング回帰木	0.023	0.968	0.017	0.093	0.291	0.067

４）日本における新型コロナウイルス感染症流行下における QOL の指標の推移、ならびに様々な要因との縦断的な関連性の評価

本研究課題はちょうど新型コロナウイルス感染症流行下の時期であったため、その時期の日本の地域・年齢・性別の人口構成比に従ってサンプリングした一般住民集団のプロファイル型尺度での OQL **の推移を記述した。また特に行動制限やワクチンなど感染症がもたらす不安の高まる背景において、心の健康や孤立といった要因とワクチン接種忌避との関連性、さらには新型コロナウイルス感染状況とそう痒との関連性についても検討を加えた。**

４－１）周囲の新型コロナウイルス感染症罹患の状況と皮膚のかゆみとの関連性

20 歳以上の 3330 **名を対象に、家族ならびに自身の新型コロナ感染症と皮膚のかゆみの有病との関連性（調整リスク比** = 1.45, 95%**信頼区間** 1.14–1.86）**、ならびにベースラインでかゆみのなかった** 2549 **名を対象に、**1 **年後のかゆみの発症との縦断的な関連性（調整リスク比** = 1.97, 95%**信頼区間** 1.48–2.64）**を示した。**

４－２）高齢者における社会的孤立と、ワクチン忌避との関連性

65 歳以上の高齢者 910 **名を対象に、**Lubben **ソーシャルネットワークスケールで測定した社会的孤立の有無と、ワクチン忌避との関連性について縦断的に解析した。結果、社会的に孤立している高齢者は、そうでない高齢者よりも新型コロナウイルス感染症ワクチン接種を忌避する傾向にあることが明らかとなった（調整リスク比** 1.98, 95% **信頼区間** 1.18 - 3.32）。

5．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　0件／うち国際共著　0件／うちオープンアクセス　0件）

1．著者名 Kogame T.、Ogawa Y.、Kabashima K.、Yamamoto Y.	4．巻 36
2．論文標題 At‐risk circumstances for COVID‐19 increase the risk of pruritus: cross‐sectional and longitudinal analyses	5．発行年 2021年
3．雑誌名 Journal of the European Academy of Dermatology and Venereology	6．最初と最後の頁 e174-e175
掲載論文のDOI（デジタルオブジェクト識別子） 10.1111/jdv.17809	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計0件

〔図書〕　計0件

〔産業財産権〕

〔その他〕

－

6．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	大前　憲史 (Omae Kenji) (60645430)	福島県立医科大学・公私立大学の部局等・准教授 (21601)	
研究分担者	後藤　匡啓 (Goto Tadahiro) (80622894)	東京大学・大学院医学系研究科（医学部）・客員研究員 (12601)	

7．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

8．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関