

自殺リスク要因を検討するマイクロデータ分析環境の構築と 自殺総合対策に資する統計的活動

研究分担者	椿 広計	統計数理研究所
研究協力者	岡本 基	情報・システム研究機構
研究協力者	岡 檀	統計数理研究所
研究協力者	久保田貴文	多摩大学
研究協力者	竹林 由武	福島県立医大
研究協力者	谷道正太郎	総務省統計局統計データ利活用センター

研究要旨：自殺総合対策を地域プロフィールに基づき提示するため、総務省統計局統計データ利活用センターの協力を受け、平成 30 年度構築した情報・システム研究機構データサイエンス共同利用基盤施設オンサイト拠点で、総務省社会生活基本調査生活編個票データを用いた自殺総合対策に資するデータ分析を行い国内外のシンポジウムで発表した。

方法：自殺総合対策に資するデータとして、総務省社会生活基本調査生活時間編を申請し、オンサイト拠点で分析すると共に、自治体別自殺率データとのリンケージを行うために公表されていない全国 1741 自治体別集計結果のオンサイト拠点外持ち出し申請を行った。また、平成 27 年度に実施した厚生労働省国民生活基礎調査匿名化データ分析において、自殺リスク増大要因として摘出されていた高介護負担について、社会生活基本調査介護時間を従属変数とし、個人属性、自治体のマクロ経済変数を独立変数とした探索的分析を統計的機械学習手法である回帰樹で行い、高介護負担の要因候補の発見も試みた。

結果と考察：1741 市区町村別の生活時間情報の基本集計を行い、自治体当たりの抽出標本 10 以上の 1501 自治体についてオンサイト拠点外への持ち出しが可能となり、自治体の自殺リスク要因データの一部とすることができた。一方、現状のオンサイト拠点分析結果の持ち出し基準では、調査標本の割り当てが小さい自治体の集計結果は持ち出しができず、自治体にフィードバック可能な情報とはできない。従って、現状の社会生活基本調査における標本配分に自治体別集計ができる制約（例えば各自治体、最低 10 名分の世帯抽出）を付けることが必要と考える。ただし、これは同調査の集計精度維持のために全標本の大きさが若干増大する可能性もある。

統計的機械学習手法で、導かれた要因は多くは極めて常識に合致したものであるが、自治体のマクロ統計変数だけでなく、県による介護負担の差があるという仮説が浮き彫りになった。県による介護負担の差の原因については、現時点では不明である。

なお、K6 と地域情報とが含まれる国民生活基礎調査マイクロデータのオンサイト拠点利用については、政府方針として 3 年以内に基幹統計調査をオンサイト拠点で利用可能とする方針は示された。

A. 研究目的

本研究は、自殺総合対策に資する公的統計データの利用環境を各府省と連携して構築整備すること、さらに具体的にその種のデータを利活用して、自殺対策に資する実証研究を加速することを目的としている。また、本研究において協力者竹林を中心に平成 27 年度に実施した国民生活基礎調査匿名データにおける K6 に対するリスク分析を基に、地域政策に資する、地域情報とのリンケージを可能にするデータ分析環境構築を目指すことを研究目的とした。

特に令和元年度は、構築したマイクロデータ環境における分析の有効性検証を主要な目的とした。

B. 研究方法

(1) オンサイト環境の有効性検証

平成 29 年 4 月 1 日に和歌山市に総務省統計局統計データ利活用センターが設置された。当該センターは、公的統計マイクロデータの公益性の高い政策研究を支援することをミッションとしており、本研究の目的と整合的であった。このため、平成 29 年度以降、協力者谷道統計データ利活用センター長、(独)統計センターの指導の下、自殺総合対策に資する総務省社会生活基本調査生活編マイクロデータ並びに自治体が活用可能な全自治体のマクロデータの利活用環境整備を実施してきた。平成 31 年度には、情報・システム研究機構データサイエンス共同研究基盤施設に協力者岡本を管理責任者とし、マイクロデータを直接分析可能なオンサイト拠点が整備・設置された。同拠点は、(独)統計センターから正式にオンサイト拠点としての認可を受けた。これを通じて、広く全国の研究者が活用可能な同拠点での公的統計マイクロデータの分

析が可能となり、自殺総合対策に資する公的統計マイクロデータの利活用基盤の構築に関する初期の目的は果たされた。

令和元年度は、実際にオンサイト拠点でのマイクロデータ分析が十分有効であることを検証することを目的として、平成 27 年度協力者竹林・久保田によって厚生労働省国民生活基礎調査匿名化データで分析された在宅介護者のうつ状態からの K6 悪化検討を基に、平成 28 年度総務省社会生活基本調査生活時間編個票データ（全国 8 万世帯、10 歳以上の世帯員 20 万名が調査対象）の分析を 2 つ行った。

最初の分析は、自治体にオンサイト拠点で集計した市区町村単位の集計結果を提供できるか否かの検討である。総務省は社会生活基本調査の集計は、目標精度上の問題から公表していないからである。市区町村別集計にあたっては、単純集計ではなく、集計用乗率（ウェイト）を用いた集計を椿が行うこととした。

第 2 の分析は、協力者の岡と椿による介護時間に影響をあたえる要因を統計的機械学習手法である樹形モデルを用いて行った探索的解析である。個票データに含まれる個人属性、地域属性以外に、統計センターが提供する SSDSE

（教育用標準データセット：Standardized Statistical Data Set for Education）の全国市区町村集計データを社会生活基本調査マイクロデータに結合させ、家族介護の実態と地域差についても分析した。

これらのデータ分析を通じて、整備したオンサイト拠点における分析が自殺総合対策に資するものか検証を行った。

(2) オンサイト拠点の普及

椿と協力者の岡本はこの種のマイクロデータ

分析環境を多くの自治体や行政が利活用できるための展開するために、オンサイト拠点を設置した大学並びにその形成に関心のある大学とのコンソーシアム「公的統計マイクロデータ研究コンソーシアム」を運営してきた。令和元年度は、協力者谷道と共に、その運営を、総務省統計データ利活用センターと共同で行う体制に切り替え、2020年1月28日に「公的統計マイクロデータ研究コンソーシアム」において、研究協力者である谷道の「公的統計における2次利用制度とオンサイト利用の概要」を中心としたチュートリアルも実施した。このコンソーシアムでは、「EBPM（証拠に基づく政策立案）におけるマイクロデータ利用」のセッションも企画した。当該セッションで樺は、「EBPMに向けたオンサイト施設におけるマイクロデータ分析」をオンサイト拠点形成に関心のある研究者に紹介した。

（3）オンサイト拠点で分析可能なデータ

自殺総合対策に資するデータで、オンサイト拠点で分析可能な公的統計マイクロデータは、国民の生活状況を明らかにする、総務省「社会生活基本調査」である。これを本研究が注目してきた厚生労働省国民生活基礎調査に拡大することについては、これまで協力者の谷道センター長を中心に交渉は続けてきたが実現しておらず、引き続き交渉を続けた。

（4）自治体が利活用可能なデータの構築

平成30年度に引き続き、市区町村など自治体が利活用可能な公的統計データの利用環境の整備、あるいは整備活動の（独）統計センターにおける恒常業務措置を検討した。

また、住環境が住民の精神的健康に好影響を

あたえる空間構造特性があるという仮説に立ち、「路地」に着目した分析を行った。平成31年度までに三重県のデータを整備し、分析を行い、市区町村ごとの路地存在率と自殺率との間に有意な負の相関が示されたことをふまえ、令和元年度は東日本大震災被災3県（岩手、福島、宮城）の海岸部に位置する市区町村ごとに路地存在率を算出するとともに、空間構造特性のひとつとして「土地利用の多様性（医療福祉、商工業や農業、教育などの土地利用の種類、もしくはその種類の多さ）」の指標作成とデータセット構築を行った。

C. 研究結果

（1）オンサイト拠点の有効性検証

平成30年5月に施行された、新統計法、関連する総務省令の整備で、オンサイト拠点における公的統計マイクロデータ分析の具体像が明確になり、本研究のオンサイト拠点利用の有効性検証も、情報・システム研究機構データサイエンス共同利用基盤施設に「社会データ構造化プロジェクト」と連携し一挙に進捗した。

令和元年8月から令和元年12月まで、共同利用基盤施設オンサイト拠点において社会生活基本調査マイクロデータ分析が行われた。オンサイト環境には、必要な統計ソフトウェアを利用可能とすると共に、1741自治体の統計数値

（教育用標準データセット125変数）、厚生労働省が公開している自治体別自殺死亡者数、死亡率データをアップロードした。

令和元年9月から令和2年3月まで、分析結果・分析プログラムの持ち出し申請を行った。

以下では、オンサイト拠点におけるマイクロデータ解析が自殺総合対策に資すると考えられる分析結果の概要を示す。

(1-a) 市区町村別集計の有用性検証

平成 28 年度社会生活基本調査生活時間編の市区町村別集計を次の 13 変数について行った。市区町村別集計結果の中で、その結果をオンサイト拠点外に持ち出し申請し、公開可能な集計自治体は 1506 市区町村であった。集計結果の中で、10 人以下の集計結果は調査客体の保護のため、オンサイト拠点外には持ち出すことができない。人口が小さな自治体は、抽出標本の大きさが 10 未満なので公開はできないのである。参考までに、今回作成した市区町村別集計(政令市区を含む)の平均値と標準偏差を表 1 に示す。なお、単独行動時間には睡眠時間も含まれる。

表 1 市区町村別集計変数平均と標準偏差
生活時間データ：単位は分(1 日当たり)

変数名	平均値	標準偏差
平均年齢	51.4	6.2
単独行動時間	773	70
介護看護時間	3.86	4.85
単独介護看護時間	0.416	0.840
家事時間	85.3	19.2
単独家事時間	38.4	15.4
くつろぎ時間	101.2	25.3
マスメディア時間	146.7	38.1
趣味時間	46.3	20.2
社会活動時間	5.06	7.38
スポーツ時間	14.56	9.30
交際時間	17.89	11.45
睡眠時間	303.7	66.5

この市区町村データを厚生労働省が発表している市区町村別自殺死亡率（住居地，平成 31 年 6 月暫定値）データとリンクし、次の 2 つの解

析を行った。

第一の分析は、自殺死亡率が 0 でない 67 自治体のリスク要因を表 1 のデータから見出した。この際、市区町村の平均年齢と生活時間の交互作用効果を独立変数とし、その効果を線形ロジスティック回帰モデルで推定した上で、AIC を用いてモデル選択をおこなった。自殺死亡率が正になる確率の予測モデルを導出した。平均スポーツ時間は年齢に依らずリスク低減要因（10%有意）、テレビ・新聞などを見るマスメディア時間・家事時間は、高齢化自治体ほどリスクを低減する交互作用効果（家事時間との交互作用は 10%有意）、くつろぎ時間は高齢化自治体ほどリスクを増大する交互作用効果（5%有意）が検出された。

第 2 の分析は、自殺死亡率が 0 でない 67 自治体について、自殺死亡率を従属変数とし、第一の分析同様、市区町村の平均年齢と生活時間の交互作用効果を独立変数とし、その効果を線形ロジスティック回帰モデルで推定した上で、AIC を用いてモデル選択をおこなった。

高齢化と共に自殺率増大という交互作用が認められるのは、睡眠時間（1%有意）、単独介護看護時間(p value =0.21)、非単独介護看護時間（10%有意）、マスメディア時間(p value = 0.11)、交際時間(1%有意)である。逆に高齢化と共に自殺率減少という交互作用があるのは、スポーツ時間(0.1%有意)、家事時間(p value =0.12)、単独家事時間(5%有意)である。この種の交互作用が認められないものとして、くつろぎ時間は自殺率増大傾向(p value = 0.23)が認められた。

限られた分析結果ではあるが、総務省社会生活基本調査の市区町村別集計は自殺総合対策に資する情報を提供できる可能性が示唆された。

(1-b) ミクロデータ分析の有用性

オンサイト拠点でのマイクロデータ分析は、社会生活基本調査個票の中で、介護を必要とする家族が存在する 37060 名を分析対象とした。分析対象の介護時間平均は単純集計で 24.5 分であるが、介護対象の年齢が 65 歳以上だと 48.1 分となることも分かった。

ここでは、介護時間を従属変数とし、就業状態、家計収入、性、年齢、世帯主との続柄、結婚状態（未婚・既婚・離別・死別）、労働時間、学業時間、通勤・通学時間、外部介護支援状況、介護休暇日か否かの識別情報、介護が必要な者の年齢、自宅介護か否かといった社会生活基本調査個票データに含まれるマイクロ情報を独立変数とし、更に、調査世帯が居住する都道府県情報、SSDSE に含まれる各自治体の人口統計学的属性、財政状況、医療施設情報なども独立変数に追加し、基本的な統計的機械学習モデル（第 2 世代人工知能）である樹形モデルあてはめを行い、完全に自動的に要因を抽出した。社会生活基本調査匿名データの分析と異なるのは、各世帯の居住地情報が存在することである。

オンサイト拠点内での樹形モデルあてはめ結果について、(独)統計センターに結果の持ち出し申請を行った。樹形モデルの終端節に含まれる標本の大きさが 10 以上になる処理を行った結果について持ち出しが許可された。

統計的機械学習で得られた介護時間推定モデルで、オンサイト拠点外に持ち出し、公表が許可された結果の概略を図 1 に示す。

人工知能は、本人の年齢が 24 歳以下か 25 歳以上かで、まず層別を行った。次に 25 歳以上の者についてはフルタイムの就業かパートタイムないしは非労働かを自動的に層別した。フ

ルタイムの就業者については、調査日が介護休暇であるか否かで層別し、パートタイムないしは非労働者については、介護を受ける家族が外部サービスを週 3 日以下受けるか 4 日以上受けるかといった家族の要介護度の代替変数で層別を行っている。これらは、一切人為的な操作を加えず、アルゴリズムが自動層別した結果である。

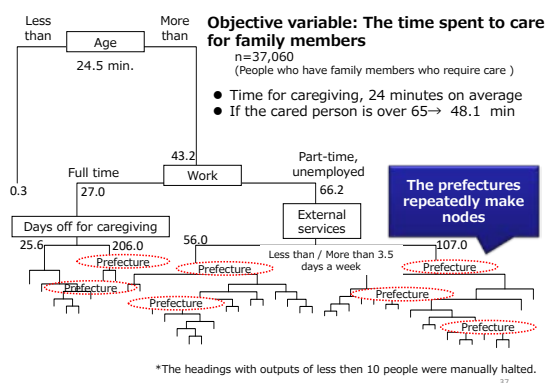


図 1 機械学習による自動層別結果の概要

次に細かい層別としてアルゴリズムが探索したのは、都道府県情報で、県によって介護時間に差異があるというものであった。

更に、細かい層別として各自治体のマクロ変数が抽出された。例えば、人口当たりの保育園児数が 0.03 人以上であること、女性の就業率が 0.16 以下であること、第二次産業従業者率が 0.087 以上であること、人口当たりの病院数が 0.0002 以下であることなどが、介護時間を増大させる場合があることが示唆された。図 1 の自動層別ルールの終端節には、1 日 6 時間以上の介護時間となるグループも生じている。

(2) オンサイト拠点の普及

統計データ利活用センター並びに、「公的統計マイクロデータ研究コンソーシアム」の活動により、オンサイト施設は、全国に 11 か所設立

された。自殺総合対策に係るマイクロデータ分析を行っているのは、2施設である。

これらの公的統計マイクロデータに基づく分析環境の整備に関する貢献で、総務省統計局統計データ利活用センター並びに公的統計マイクロデータ研究コンソーシアムは、令和2年度一般社団法人日本統計学会活動賞受賞が決まった。

(3) オンサイト拠点で分析可能なデータ

平成30年末に、厚生労働省人口動態統計のマイクロデータがオンサイト環境で利用可能となったが、その後、自殺総合対策に資するK6情報と地域情報を含む厚生労働省国民生活基礎調査については、地域情報を含まない匿名化データだけが提供可能な状況が続いている。統計データ利活用センターによる交渉が続いている状況であった。

しかし、令和元年末に、国民生活基礎調査も含まれる、各府省の基幹統計調査個票については、3年以内にオンサイト拠点で活用する方針が明確化された。

(4) 自治体が利活用可能なデータの構築

椿が、平成29年10月地方自治体職員に対する「データに基づく問題解決」研修(EBPM研修)に、全国市区町村の基本状態を表す70変数のデータセットを試作して以降、平成30年6月に(独)統計センターは、1741市区町村の111変数からなる「教育用標準データセット(SSDSE)」を公表した。

(独)統計センターは、SSDSE作成を法人業務計画に公式に組み込んだ。椿は、同センター顧問としてこのデータ作成のプロジェクトに参画し、令和元年6月には市区町村情報を125変数に増大した2019年版を公表した。現在

2020年版の公表が準備中である。これらは、教育用とは称しているが、自殺総合対策に係る自治体プロフィール情報として活用可能である。

また、研究協力者の岡が、構築した土地利用の多様性データについても、その自殺総合対策への有用性検証を行い、東北被災3県(岩手、福島、宮城)の海岸部に位置する市区町村ごとに路地存在率を算出し、自殺率との関係を分析した。被災3県市区町村の路地存在率は、標準化自殺死亡比30年間平均値との間にやや強い負の相関を示したが、最近の人口10万対自殺率5年間平均値との間には有意な相関は示されなかった。今後さらなる検討を要する。

D. 考察

オンサイト拠点における、公的統計マイクロデータ分析について、オンサイト拠点でしか集計できない、社会生活基本調査の自治体別集計は、自殺総合対策に資する情報であることが示された。しかし、現状の社会生活基本調査標本設計では、全ての自治体が世帯員合計10名以上の世帯抽出が行われているわけではない。従って、全自治体にプロフィール情報を提供できるわけではない。今後、社会生活基本調査の標本設計に際し、各自治体最低5世帯は抽出するといった制約をつけることで、全自治体の集計結果の利活用が可能になる。ただし、公表される公的統計自体の精度が劣化しない配慮も必要である。また、自治体別集計値は、今回のようなリスク分析に基づく政策推奨の参考情報として使われるべきであり、個別自治体の生活時間の推定精度が十分とは言えないことには注意が必要である。

オンサイト拠点における総務省社会生活基本

調査マイクロデータの機械学習による探索的分析も、膨大な情報からの分析なので、人間の直観にあった自動層別が行われることが分かった。

また、地域情報も人工知能は自動的に層別情報として提示することが分かった。一方で、それらの都道府県が何故介護時間が長くなるかについては、現時点では解釈できておらず、今後の課題である。

また、介護時間が自殺率に与える影響については、別途行ったマクロ分析で解釈しているのが現状である。やはり、K6が含まれる厚生労働省国民生活基礎調査のマイクロデータがオンサイト拠点で分析できた段階で、自殺総合対策に直接資するマイクロデータ分析が可能になると考える。

オンサイト拠点の整備とその有効性に関する啓発は進み、霞が関にも府省専用のオンサイト拠点が立ち上がった。しかし、膨大なマイクロデータ分析を実施する要員を府省・自治体レベルでどう育成するかが、今後の課題である。政府・自治体職員に統計データアナリスト資格を付与し、EBPMを加速する構想が示されたが、それがまずはSSDSEやe-STATを用いたマクロデータ分析活用を、更にオンサイト拠点におけるマイクロデータ分析結果活用に繋がることが期待される。

F. 健康危険情報 なし

G. 研究発表

1) 論文発表

椿広計；データ利活用とEBPM（特集行政におけるデータ利活用の動向），統計 Vol. 70 (4), pp.2-5, 2019.

2) 学会発表

岡本基，山下智志：情報・システム研究機構におけるオンサイト施設の整備と構築—公的統計マイクロデータリモートアクセス型利用と「国際マイクロ統計データベース」，2019年度統計関連学会連合大会,滋賀大学, 滋賀, 2019年9月.

谷道正太郎：統計データ利活用センター活動を通じたEBPMへの貢献，企画セッション「マイクロデータの利用技術とEBPM」2019年度統計関連学会連合大会, 滋賀大学, 滋賀, 2019年9月.

岡檀：社会生活基本調査マイクロデータを利用した介護高負担要因の探索的分析，企画セッション「マイクロデータの利用技術とEBPM」2019年度統計関連学会連合大会, 滋賀大学, 滋賀, 2019年9月.

椿広計：オンサイト拠点の活用について—提供者視点から利用者視点へ—，共同研究集会「官民オープンデータ利活用の動向及び人材育成の取り組み(2019年度)」，統計数理研究所, 東京, 2019年11月

岡檀：社会生活基本調査マイクロデータを利用した介護高負担要因の探索的解析，共同研究集会「官民オープンデータ利活用の動向及び人材育成の取り組み(2019年度)」，統計数理研究所, 東京, 2019年11月

Tsubaki, H.: New Data Sources of Japanese Official Statistics in Big Data Era, the 11th International Workshop on Micro Data of Official Statistics, The Institute of Statistical Mathematics, Tokyo, 2019/11.

Tsubaki, H: Using Statistics Bureau Japan Data at an onsite facility, Introduction to Data and Resources Available at Statistics Bureau Japan organized by JSPS, 2020/01, Oxford University, UK.

谷道正太郎：公的統計における2次利用制度
とオンサイト利用の概要，公的統計マイクロデー
タ研究コンソーシアムシンポジウム，2020年
1月，エッサム神田ホール，東京。

椿広計：EBPMにおけるマイクロデータ利用，
公的統計マイクロデータ研究コンソーシアムシン
ポジウム，2020年1月，エッサム神田ホー
ル，東京。

H. 知的財産権の出願 なし