

3. NIPPON DATA80/90 追跡・分析ワーキンググループ

研究分担者 早川 岳人 立命館大学衣笠総合研究機構地域健康社会学研究センター
教授・センター長
研究分担者 原田亜紀子 滋賀医科大学 NCD 疫学研究センター医療統計学部門 准教授
研究分担者 門田 文 滋賀医科大学 NCD 疫学研究センター予防医学部門 准教授

【目的】

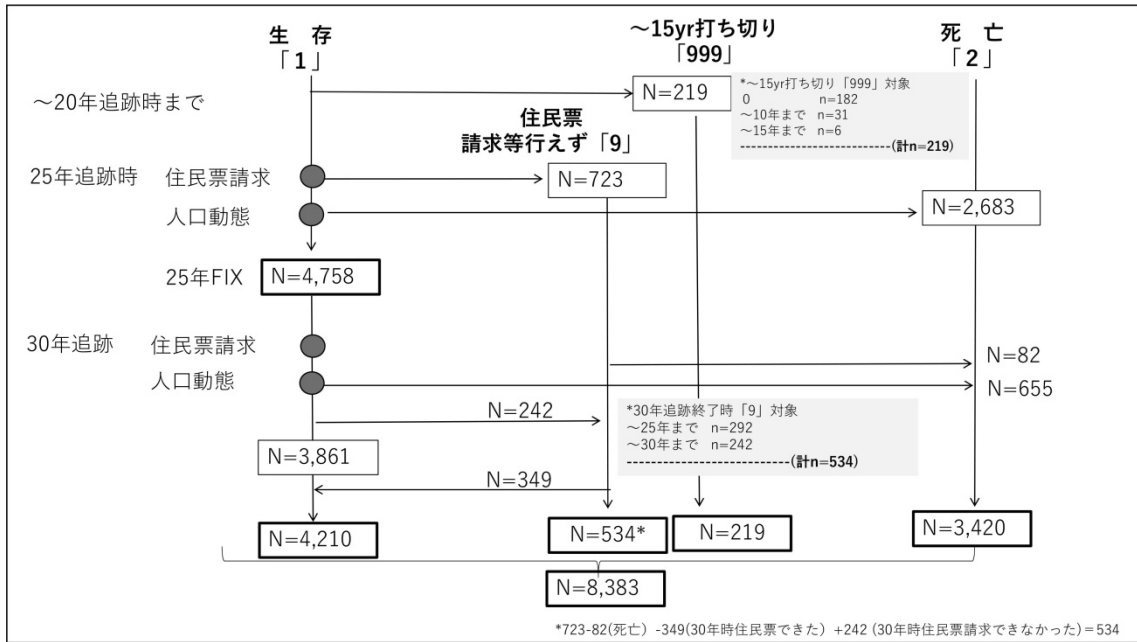
研究期間中、NIPPON DATA90 コホート約 8000 人の 30 年目生死追跡調査(住民票請求)、死因調査(人口動態統計使用請求)を行った結果を受けて、30 年追跡データセットを作成する。

また、NIPPON DATA 研究では新旧三つのコホートがあり、これらのコホートから各種要因と主要疾患に及ぼす影響について時代変遷の分析を行っていく。

【NIPPON DATA90 35 年追跡】

NIPPON DATA90 コホートは、1990 年の循環器疾患基礎調査受診者を追跡するコホート研究であり、5 年後の 1995 年、10 年後の 2000 年、15 年後の 2005 年、20 年後の 2010 年、25 年後の 2015 年、30 年後の 2020 年に、生命予後を対象者の住民票、死亡個票によって追跡を行ってきた。また、2025 年時点の 35 年追跡も今後実施していく予定である。なお、ベースラインの循環器疾患基礎調査は 1990 年 11 月に実施されているので、その後の各追跡は、11 月が経過した段階で実施している。

本年度は、30 年追跡データセットを完成させる目的で、これまでに、死亡者に対して、国から使用承認を得た人口動態統計と照合し、死亡原因を同定している。



上図に示したように、最終的に、対象者 8,383 名において、30 年追跡で生存 4,210 名、死亡 3,420 名、住民票請求が行えなかった者 534 名、15 年追跡で追跡打ち切りになった者 219 名であった。この情報をもとに、NIPPON DATA90 30 年追跡データを作成した。

【各種要因と主要疾患に及ぼす影響について時代変遷の分析】

ND80 集団の追跡開始から ND2010 の追跡 10 年まで 40 年近くが経過しており、その間のリスク因子の変化や集団の平均寿命、暦年の寄与（同じ 60 歳であっても、各集団で意味合いが異なる）も大きく変化している可能性がある。本 WG では、新旧 3 コホートでのリスク因子の寄与の変化を検討することを研究計画としており、リスク因子重要度の変化、リスク因子間の交互作用に着目した検討テーマを設定している。

1. 各集団でのリスク因子の重要度（寄与）の変化（機械学習による検討）
2. 各集団でのリスク因子の重要度（寄与）の変化（Generalized additive model による定量評価）
3. 新旧 3 コホートでのリスク因子の集団寄与危険割合の検討

ND 研究では、リスク因子の変化について経時的な測定は行っていないことから、3 集団のベースラインデータとアウトカムの関係を比較することになる。このため、リスク寄与を考えるうえで、国民健康栄養調査等を用いた経時的な検討が別途必要と考えている。また、死亡率の地域差があること、都市部と地方の生活環境の違いも大きくなっていることを考慮し、リスク因子の寄与を考えるうえで地域別の時系列推移の確認も必要と考えている。

このような時系列データとアウトカムを解析する手法については、経時的に値が測定され、時間依存性共変量を伴う生存モデルの対応として、動的予測モデルが用いられている。ジョイントモデルやランドマークモデルなどが提案されているが、前者はパラメトリック仮定が必要で、実データでは当てはまりが悪いことが多い。推移別に集団をグループ化する Trajectory 解析、LTA などの潜在変数を想定した解析も行われている。こうした状況で、近年、非線形関係のモデル化という点で、各観測値に関数を対応させ関数化データ集合を対象とする分析手法（理論）である関数主成分回帰（Functional Principal Component Analysis）を組み入れた応用が行われており、疫学領域での活用も期待される。そこで今年度は、ND 3 集団のベースライン値とアウトカムの関連の比較に入る前段階として、上記の点に着目し、関数主成分回帰を用いた予備的な検討を実施した。

1. 検討課題

平均寿命の延びの要因については、「特定死因を除去した寿命ののび」などの指標や APC 解析などが行われているが、経時的な推移において、死因構造や年齢構成などが複雑にかかわるため、有効な分析方法が提案されていない。時間など連続的な変数に対して記録されたデータを関数データと呼ぶが、本検討では、関数に対する主成分（関数主成分）を抽出することにより、関数の形状を特徴量とできる関数主成分解析を提案する。

社会保障人口問題研究所が提供する「日本版死亡データベース」の都道府県別死因（1975-2022 年）、国民健康栄養調査の都道府県別食事摂取量、都道府県生命表による平均寿命を用い、以下の検討を行った。

課題 1：関数主成分解析を用いた死因別死亡率・平均寿命の関連

課題 2：関数主成分解析を用いた経時的な死因別死亡率・平均寿命と食事摂取の関係（今回は、死因は脳血管疾患の検討）

2. 方法

解析は、1) 各死因の主成分抽出し、主成分要素を軸にプロットし、平均寿命高位県と下位県の布置状況を視覚的に確認する。2) 平均寿命の主成分の抽出。3) 直近の平均寿命(2020年)、平均寿命の主成分1~3をそれぞれ従属変数、各死因の主成分を独立変数とした一般化回帰を実施した。モデルの選択は、Ridge回帰、Lasso回帰、Elastic-net、ステップワイズ法の変数減少法を採用し、AICで最適モデルを選択した。

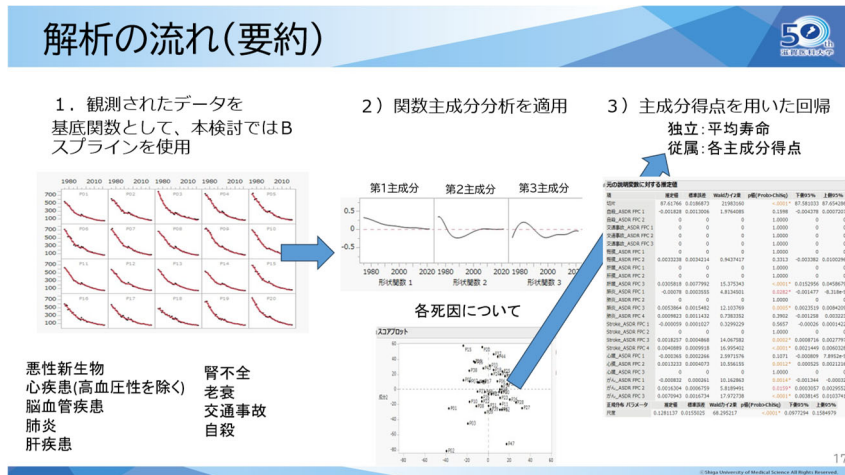


図1: 本検討における死因やリスク因子の経時主成分の抽出(関数主成分解析)と経時主成分を用いた回帰の実施手順

3. 結果

1) 平均寿命と死因別死亡率の経時推移(主成分)

一般化回帰において、従属変数を2020年の平均寿命とした場合には、女性では、がん、心疾患、脳卒中中の要素に加え、肝疾患、肺炎、男性では、これらに加え腎臓の抽出成分が選択された。平均寿命増加を示す第3成分を従属変数とした検討において、女性では、がん、脳卒中、肝疾患、肺炎、男性では、がん、心疾患、肝疾患、老衰の成分が抽出された(図2)。

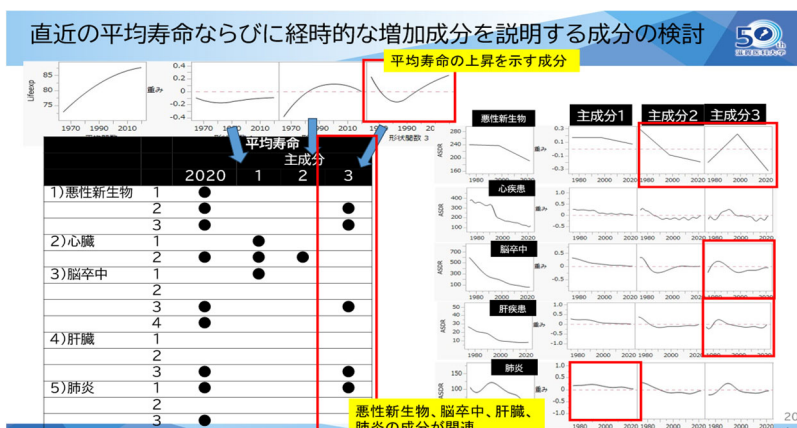


図2

直近の平均寿命、平均寿命の経時主成分を従属変数、各死因の経時主成分を独立変数とした検討(女性の結果)

左表●に該当する死因主成分が関連のあった成分。右図に各死因の経時主成分(1-3)の形状を示した

2) 平均寿命・死亡率の経時推移(主成分)と食事摂取量の経時推移(主成分)

国民健康栄養調査の地域(ブロック別)食事摂取量と平均寿命、経時的な死亡率(脳血管疾患)の関連においては、直近の平均寿命に対しては、エネルギー、脂質の第2主成分(増加傾向を示す成分)が、脳血管疾患死亡率の経時的成分(低下成分)に対しては、脂質の第1主成分、第2主成分、炭水化物

の第1主成分が脳血管疾患死亡率の低下に関連していた。

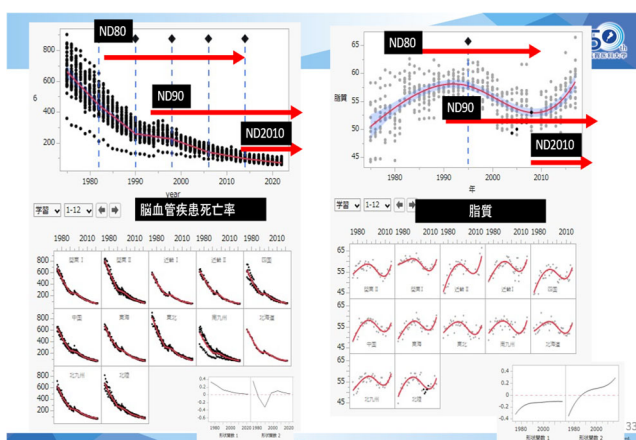
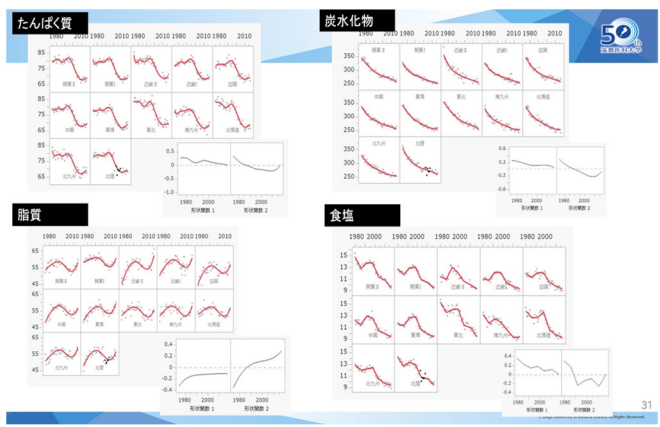
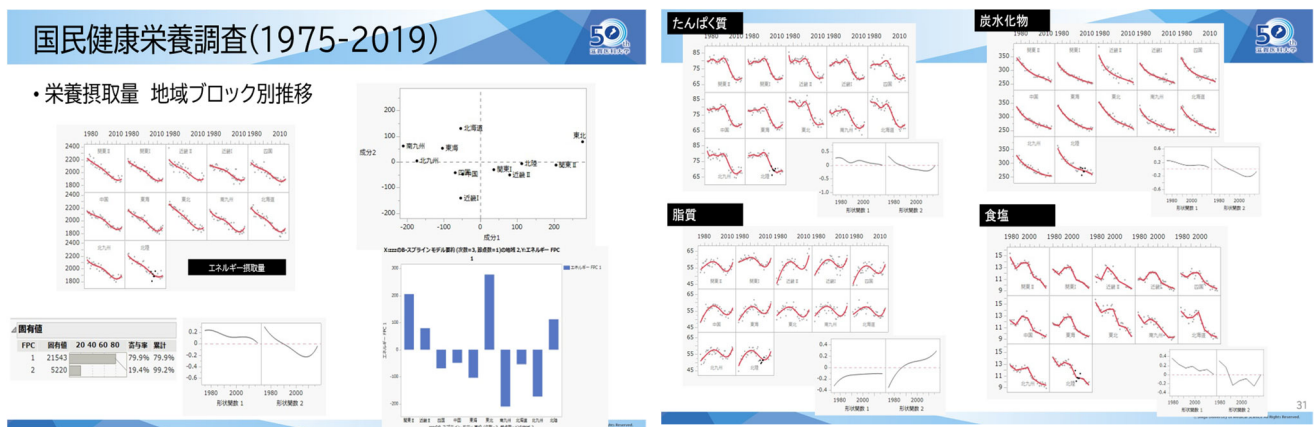


図3-1(左上)
国民健康栄養調査 栄養摂取量の地域ブロック別経時推移と経時主成分:エネルギー

図3-2(右上)
同上たんぱく質、炭水化物、脂質、食塩

図3-3(左下)
脳血管死亡率経時推移と経時主成分

4. 考察

平均寿命に対して、死亡率の高い三大死因の寄与が大きいのはもちろんであるが、近年の治療や予防対策により、死亡率低下が顕著な疾患の寄与が大きいことが考えられた。肝疾患などで、過去に死亡率が高率で、その後の死亡率低下が大きい県において、近年の平均寿命のびや都道府県順位の上昇がみられる傾向が確認された。リスク因子としては、今回は食事因子を取り上げたが、脳血管疾患死亡に対しては、脂質や炭水化物摂取との関連がみられた。今回の検討では、1970年代からの長期的変動を解析対象としたが、関数主成分解析は、区間の取り方で抽出成分が変わるため、さらに年代を区切った検討も行うとともに、食事以外のリスク因子についても同様に検討を行う予定である。

5. 新旧3コホート集団の解析を行う予備解析としての本検討の意義

関数主成分解析手法上のメリット

- ・アウトカムの時系列推移において、増加成分、減少成分が分離できれば、このような動的特性に対し寄与が大きいリスク因子成分を抽出できる可能性がある
- ・死亡率、平均寿命の地域差の要因（時系列を伴った地域差の発生）特定に寄与できる可能性がある
- ・経時データの解析手法の1つの選択肢となりうる（相関の強い変数の取り扱い、経時成分の解釈や分析のしやすさ）

今後、新旧3コホート集団の解析を進めるうえで

- ・アウトカムとする死亡率が新旧コホートで大きく異なるため、相対リスクと絶対リスクの解釈を行っていくうえでの予備的な情報となる
- ・アウトカムとリスク因子の検討において、それぞれの時系列推移を考慮したうえでの結果の解釈も可能となるので、これまでの新旧コホートでの既報と長期追跡結果を比較検討するうえでの確認資料となりうる。

参考文献

- 1) 松井秀俊. 関数データに基づく統計的モデリング. 統計数理.2019;67:73-96
- 2) Shi B, Wei P, Huang X. Functional principal component based landmark analysis for the effects of longitudinal cholesterol profiles on the risk of coronary heart disease. Stat Med. 2021 Feb 10;40(3):650-667
- 3) Dey D, Ghosal R, Merikangas K, Zipunnikov V. Functional Principal Component Analysis for Continuous Non-Gaussian, Truncated, and Discrete Functional Data. Stat Med. 2024 Dec 10;43(28):5431-5445.
- 4) JMPPro: Functional Data Explorer Functional DOE
<https://www.jmp.com/support/help/ja/17.2/index.shtml#page/jmp/functional-data-explorer.shtml>
- 5) R : fda <https://cran.r-project.org/web/packages/fda/fda.pdf>
- 6) Python : [scikit-FDA](https://fda.readthedocs.io/en/latest/auto_examples/plot_fpca.html) https://fda.readthedocs.io/en/latest/auto_examples/plot_fpca.html