

# 歩数とレセプトの匿名加工情報を用いた歩行不足による生活習慣病の罹患リスク

進藤 翔太†

池上 和輝‡

伊藤 聡志‡

菊池 浩明†

明治大学総合数理学部†

明治大学大学院‡

表 1 匿名加工データの統計量

データ名	レセプトデータ	歩数データ	健康診断データ
属性数	15	4	49
期間	2012–2018	2014–2018	2008–2018
レコード数	3,936,877	24,936,120	964,636
対象者	288,568	58,655	198,740

表 2 各匿名加工データの属性

データ名	属性	例
レセプト	仮個人 id	user_A
	診療開始日	2010/11/10
	icd10 コード 1_基本疾患_細分類	I10
歩数	仮個人 id	user_A
	年月日	2015/6/11
	歩数	5602
	歩数データソース	スマートフォン自動連携
健康診断	仮個人 id	user_A
	健診受診月	2010/11/10
	BMI	21.2
	喫煙	はい

## 1 はじめに

歩数と死亡リスクには強い相関がある。Pedro らは 40 歳以上の人々を対象に、最長 7 日間歩数を計測した後に、約 10 年間の追跡した結果、1 日 4000 歩の人たちに比べ、8000 歩の人たちは死亡リスクが半減したことを報告している [1]。しかし、計測した歩数が最大 7 日間であることが信憑性に欠ける。

そこで、本研究ではウェアラブルデバイスに着目する。ウェアラブルデバイスは身に着けるだけで、体の状態や行動を簡単に計測でき、私たちの健康維持に役立つとされている。本研究ではウェアラブルデバイスから得られた日々の歩数の記録とレセプトのデータを組み合わせた匿名加工情報を用いて、歩数と 3 年以内の生活習慣病の罹患リスクの相関を明らかにする。本稿では、歩数の記録が 1 年間で 330 日以上ある 40~59 歳の男性 2159 人を対象に調査を行った結果を報告する。

## 2 データ分析

### 2.1 データ概要

本研究では、あるヘルスケアサービス事業者から法律に従って適切に措置され、提供されたヘルスケアデータの匿名加工情報 [2] を使用する。表 1 に使用した匿名加工データの統計量を示す。レセプトデータの属性、歩数データの属性、健康診断データの属性を表 2 に示す。

### 2.2 データの前処理

ウェアラブル端末の歩数の記録が 2016 年、330 日以上ある 40~59 歳のユーザー 2159 人を対象とする。2016 年の健康診断データから、BMI、喫煙の有無の問診票の項目、2016 年~2018 年のレセプトデータから、表 3 に示す 5 つの生活習慣病の ICD-10 コードを取り出す。ICD-10 コードは、世界保健機関

表 3 ICD-10 コードによる生活習慣病の分類

生活習慣病	ICD-10 コード
高血圧症	I10
糖尿病	E11, E12, E13, E14
脂質異常症	E780, E781, E785
虚血性心疾患	I209, I219, I259
高尿酸血症	E790

(WHO) が作成している疾病の分類である。ICD-10 コードによる生活習慣病の分類は [3] を参考にした。

### 2.3 分析手法

5 つの生活習慣病について、歩数が 4000 未満、4000~5999、6000~7999、8000~9999、10000 以上の 5 つのグループに分け、次の 2 つの分析を行う。

(1) 6000~7999 歩を基準にして、各疾病に 3 年以内に罹患するリスク比 Relative Risk(RR) を求め、独立性の  $\chi^2$  検定を行う。

(2) ロジスティック回帰分析を行う。目的変数  $y$  を 3 年以内の罹患の有無、説明変数を歩数を 2000 歩ずつに分け、6000~7999 歩を基準にして、相対的に比較する歩数群  $x_1$ 、年齢  $x_2$ 、BMI  $x_3$ 、喫煙の有無  $x_4$  としたロジスティックモデル

$$y = \frac{1}{1 + e^{-(a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4)}}$$

を用いて、交絡因子の影響を考慮した各疾病のオッズ比 Odds Ratio(OR) を求める。

### 2.4 分析結果

ユーザの特徴、罹患率、リスク比、 $\chi^2$  検定の結果を表 4 に示す。歩数の項目のロジスティック回帰分析の結果を表 5 に、高血圧症における全項目のロジスティック回帰分析の結果を

A risk of lifestyle diseases incurred by lack of daily walking steps analyzed from anonymously processed data of daily steps count and health insurance claims

†Shota Shindo and Hiroaki Kikuchi, School of Interdisciplinary Mathematical Science, Meiji University.

‡Kazuki Ikegami and Satoshi Ito, Meiji University Graduate School.

表4 平均歩数による各疾患の罹患率

平均歩数	<4000	4000-5999	6000-7999	8000-9999	≥10000
全体	266	679	674	363	177
平均年齢	48.9	49.0	50.0	48.8	49.9
BMIの平均	24.2	24.0	24.0	23.9	24.0
喫煙の有無 (%)	70(26.3)	169(24.9)	146(21.7)	82(22.6)	36(20.3)
罹患率 (%)	109(40.1)	240(35.3)	212(31.5)	121(33.3)	59(33.3)
生活習慣病全体	RR 1.30	1.12	1	1.06	1.06
	p-value 0.000	0.144	1	0.583	0.699
高血圧症	罹患率 (%) 41(15.4)	92(13.5)	56(8.3)	37(10.2)	13(7.34)
	RR 1.86	1.63	1	1.22	0.88
	p-value 0.002	0.003	1	0.369	0.792
糖尿病	罹患率 (%) 63(23.7)	146(21.5)	117(17.4)	75(20.7)	47(26.6)
	RR 1.36	1.24	1	1.19	1.52
	p-value 0.033	0.063	1	0.222	0.008
脂質異常症	罹患率 (%) 53(19.9)	112(16.5)	82(12.2)	45(12.4)	27(15.3)
	RR 1.63	1.35	1	1.02	1.25
	p-value 0.003	0.028	1	0.993	0.333
高尿酸血症	罹患率 (%) 25(9.4)	51(7.5)	35(5.2)	20(5.5)	7(3.4)
	RR 1.80	1.45	1	1.06	0.76
	p-value 0.026	0.10	1	0.943	0.630
虚血性心疾患	罹患率 (%) 25(9.4)	41(6.0)	39(5.8)	20(5.5)	17(9.6)
	RR 1.62	1.04	1	0.95	1.66
	p-value 0.066	0.94	1	0.966	0.098

表5 ロジスティック回帰分析の結果(歩数)

平均歩数	$x_1$	<4000	4000-5999	6000-7999	8000-9999	≥10000
生活習慣病全体	OR	1.50	1.19	1	1.10	1.12
	p-value	0.008	0.141	1	0.464	0.52
高血圧症	OR	2.01	1.74	1	1.30	0.89
	p-value	0.002	0.002	1	0.248	0.727
糖尿病	OR	1.47	1.30	1	1.28	1.82
	p-value	0.033	0.064	1	0.147	0.003
脂質異常症	OR	1.78	1.42	1	1.04	1.35
	p-value	0.003	0.025	1	0.854	0.220
高尿酸血症	OR	1.82	1.46	1	1.08	0.86
	p-value	0.029	0.096	1	0.792	0.734
虚血性心疾患	OR	1.68	1.04	1	0.96	1.74
	p-value	0.054	0.854	1	0.891	0.071

表6 高血圧症におけるロジスティック回帰分析

説明変数 $x_i$	$a_i$	OR	p-value
歩数 (1:<4000 0:6000-7999)	0.698	2.01	0.002
年齢	0.087	1.08	0.000
BMI	0.123	1.13	0.000
喫煙 (1: はい 0: いいえ)	-0.032	0.97	0.903

表6に、先行研究の4000歩を基準とした約10年以内の死亡リスクと本研究の6000~8000歩を基準とした3年以内の生活習慣病の罹患リスクの比較を表7に示す。

表4から4000未満のグループは全疾病について罹患率が高く、6000~8000歩の基準群と比べるとリスクが4つの生活習慣病で1.36~1.86倍、生活習慣病全般で1.3倍に上がっていた。表5でロジスティック回帰分析の交絡因子を考慮した分析を行っても、統計的に有意な水準で(p値が0.05未満)、歩行不足により、3年以内の短期間で、高血圧で2倍、生活習慣病全般でもリスクを1.5倍に上げることが判明した。表6で歩数、年齢、BMIが有意な水準で、生活習慣病のリスクを上げる因子であるとわかったが、喫煙の項目に有意差は見られなかった。表7で、先行研究は歩数に比例して死亡リスクが減ったのに対して、

表7 先行研究とのリスク比較

平均歩数(先行研究)	本研究(生活習慣病)	先行研究(死亡)
10000以上(12000)	1.06	0.35
6000-8000(8000)	1	0.49
4000未満(4000)	1.30	1

本研究では、10000歩以上では1.06倍と罹患リスクが再び増加する非単調な振る舞いをする結果となった。

## 2.5 考察

歩数はBMI、年齢と同様に生活習慣病をリスクを上げる要因となることが判明した。喫煙の項目に有意差が見られなかったのは、問診票の虚偽の記載を指摘する研究[4]があることから、喫煙者が非喫煙者であると虚偽の記載をした可能性があると考えられる。

3年以内の短期間で、4000歩未満の群は生活習慣病全体のリスクを1.3倍に上げてしまうが、4000歩以上の4つの群では有意差が見られなかったことから、歩数が極端に少ない場合は短期間で、健康に悪影響を及ぼしてしまうと考えられる。本研究では3年以内の罹患の影響を見たが、先行研究の様に長期的に見たら、影響を及ぼすことが示唆される。

## 3 おわりに

本研究では歩数と生活習慣病の罹患リスクについて、匿名加工情報を用いて調査を行った。歩数が4000歩未満であると、3年以内の生活習慣病の罹患リスクを上げてしまうことがわかった。ただし、それ以上の歩数では罹患リスクがほとんど変わらず、歩行による健康の向上の明確な証拠は得られなかった。

## 参考文献

- [1] Pedro F. Saint-Maurice, et al. "Association of Daily Step Count and Step Intensity With Mortality Among US Adults", JAMA, 323(12) pp.1151-1160, 2020.
- [2] 伊藤 聡志, 池上 和輝 "匿名加工情報の応用 (1): 健康診断データとレセプトデータの分析とプライバシーリスク評価", 情報処理学会 CSS 2020, pp. 1230-1237, 2020.
- [3] 全国健康保険協会, "生活習慣関連 10 疾患の動向に関する調査" ([https://www.kenporen.com/toukei\\_data/pdf\\_chosa\\_r01\\_06\\_01.pdf](https://www.kenporen.com/toukei_data/pdf_chosa_r01_06_01.pdf), 2020年9月参照)
- [4] 笹原 妃佐子, 西村 瑠美, 深田 恵里, 二川 浩樹, "大学生の健康診断時の喫煙に対する虚偽回答に関する研究", 総合保健科学: 広島大学保健管理センター研究論文集, Vol. 34, 1-12, 2018.