

大和証券ヘルス財団の全研究業績集の計量テキスト分析 — 38年間の成果報告書の可視化 —

鹿児島大学大学院 医歯学総合研究科
助教 下高原 理恵

(共同研究者)

鹿児島大学医学部 保健学科 教授 緒方 重光
鹿児島大学医学部 保健学科 助教 西本 大策
鹿児島大学医学部 保健学科 助教 李 慧瑛

はじめに

昭和47年の財団設立とともに研究助成が開始され1,041編の研究論文¹⁾が生み出された。これまでの38年間の4期(「第1～4次国民健康づくり運動」政策を基準に分類)にヘルスケアに関わる専門家がどのようなテーマに関心を持ち、研究を行ってきたかを調査した。データの分析はテキストマイニングの手法²⁻⁷⁾を用い、質的帰納的に行なうために報告書に書かれた生データを繰り返し読み、全体を理解した。そして未加工データをExcelに入力して、報告書の電子テキスト化を行った。次に、解析ソフトウェアとして、KH Coder (ver. 2.00)とText Mining Studio (数理システム)を用いて内容分析した。

結 果

38年間の全成果報告書のテキストデータは、文章数161,480、総抽出語数3,172,487語、分析対象語数1,275,652語であり、異なり語数118,583語、分析対象異なり語数91,185語であった。

1) 全報告書の上位20頻出語

頻出語上位5語の使われ方について、KWICコンコーダンスで分析した。「細胞」は7,292回出現しているが、「肝細胞」「細胞障害」「細胞治療」「細胞株」「癌細胞」「骨芽細胞」「内皮細胞」「腸管細胞」「神経細胞」等の文脈で使用されていた。「血」は4,868回出現しているが、「虚血性」「冠血流」「採血」「血清」「高脂血症」「観血的」「脳血管」「便潜血」等の文脈で使用されていた。「癌」は4,589回出現しているが、「癌患者」「癌免疫」「大腸癌」「癌抗原」「癌治療」「腺癌」「癌遺伝子」「癌組織」等の文脈で使用されていた。「患者」は4,067回出現しているが、「生活習慣病患者」「糖尿病患者」「透析患者」「患者背景」「対象患者」「患者血清」等の文脈で使用されていた。「調査」は3,461回出現しているが、「疫学調査」「満

足度調査」「質問紙調査」「面接調査」等の文脈で使用されていた(表1)。

2) 各年度の上位10の頻出語

年度毎に表れる頻出語には、国民健康づくり運動の政策が反映された特徴語が見られる(表2)。

表1 全報告書の上位20頻出語

	抽出語	出現回数
1	細胞	7292
2	血	4868
3	癌	4589
4	患者	4067
5	調査	3461
6	健康	3371
7	疾患	3266
8	治療	3226
9	運動	2525
10	症例	2405
11	死亡	2392
12	発現	2390
13	検診	2348
14	血圧	2304
15	高齢	2239
16	機能	2225
17	検査	2205
18	地域	2169
19	因子	2079
20	肝	2046

表2 各年度の上位10の頻出語

抄録	年号	西暦	1位	2位	3位	4位	5位	6位	7位	8位	9位	10位
第1期：第1次国民健康づくり対策(1978年～1987年)												
1	52	1977	血	運動	肥満	心臓	負荷	血糖	脳卒中	脂肪	症例	血圧
2	53	1978	血	死亡	健康	血圧	脳卒中	心臓	運動	地域	糖尿	疾患
3	54	1979	血	糖尿	健康	死亡	血圧	心臓	測定	運動	体力	蛋白
4	55	1980	血	尿	血圧	検診	肥満	健康	地域	喫煙	蛋白	年齢
5	56	1981	血	糖尿	患者	死亡	運動	血圧	癌	脂肪	細胞	心臓
6	57	1982	血	糖尿	疾患	運動	健康	肝	摂取	症例	死亡	地域
7	58	1983	血	健康	疾患	血圧	肝	検診	運動	対象	喫煙	死亡
8	59	1984	血	健康	死亡	癌	調査	疾患	変化	保健	比較	患者
9	60	1985	健康	患者	調査	癌	地域	保健	医療	大腸	アルコール	発症
10	61	1986	健康	保健	血	蛋白	運動	地域	肝	細胞	脳	対象
11	62	1987	癌	細胞	患者	疾患	血圧	検診	危険	肝	因子	死亡
第2期：第2次国民健康づくり対策：アクティブ80ヘルスプラン(1988年～1999年)												
12	63	1988	細胞	癌	調査	療法	胃	疾患	摂取	健康	血	死亡
13	64	1989	癌	疾患	死亡	心	調査	患者	地域	発症	健康	健康
14	2	1990	癌	患者	血圧	疾患	血	調査	治療	検診	地域	肝
15	3	1991	健康	細胞	患者	調査	検診	老人	地域	発症	治療	保健
16	4	1992	癌	細胞	患者	症例	組織	血	大腸	高齢	老人	検診
17	5	1993	癌	大腸	細胞	血	効果	生活	血圧	対象	摂取	因子
18	6	1994	細胞	検診	癌	保健	健康	医療	蛋白	受診	活性	血圧
19	7	1995	老人	骨	検診	動脈	血	血管	血圧	医療	細胞	受診
20	8	1996	研究	因子	疾患	健康	検討	患者	原因	高齢	手術	予防
21	9	1997	高齢	健康	研究	血管	因子	変化	ケア	運動	影響	危険
22	10	1998	研究	高齢	要因	健康	開発	癌	細胞	在宅	腫瘍	免疫
23	11	1999	研究	高齢	中年	効果	治療	リスク	因子	患者	骨	骨折
第3期：第3次国民健康づくり対策：健康日本21(2000年～2012年)												
24	12	2000	高齢	研究	開発	運動	肝炎	予防	ウイルス	システム	ストレス	モデル
25	13	2001	高齢	研究	因子	影響	遺伝子	介護	細胞	基礎	調査	分析
26	14	2002	研究	開発	機能	解析	高齡	機構	硬化	動脈	因子	感染
27	15	2003	高齢	血管	インスリ	療法	運動	影響	解析	改善	患者	関与
28	16	2004	治療	細胞	開発	研究	高齡	解明	解析	脳	肝	機構
29	17	2005	研究	予防	開発	解明	発症	基礎	機能	血管	高齢	細胞
30	18	2006	研究	障害	検討	治療	解明	開発	高齢	生活	確立	患者
31	19	2007	研究	開発	癌	療法	治療	遺伝子	応用	化学	関連	機能
32	20	2008	解明	癌	治療	開発	研究	疾患	関連	中年	解析	改善
33	21	2009	細胞	研究	治療	開発	癌	診断	患者	機能	検討	高齢
34	22	2010	開発	研究	治療	解明	標的	解析	血管	新規	早期	制御
35	23	2011	解明	癌	治療	マーカー	開発	関連	研究	高齢	新規	因子
36	24	2012	開発	研究	治療	癌	血糖	解析	高齡	新規	システム	マーカー
第4期：第4次国民健康づくり対策：第2次健康日本21(2013年～)												
37	25	2013	開発	治療	細胞	応用	解析	解明	研究	新規	癌	脂肪
26	26	2014	検討	調査	癌	診断	患者	因子	予防	開発	危険	研究

3) 共起ネットワーク

我が国の国民健康づくり政策を基準に、報告書を次の4期に分類して共起ネットワーク(サブグラフ媒介)を作成した。

①第1期：第1次国民健康づくり対策開始(1978年～1987年)

当期10年間の成果報告書のテキストデータは、文章数51,521、総抽出語数936,906語、分析対象語数379,390語であり、異なり語数50,766語、分析対象異なり語数37,390語であった。

研究内容は、健康の増進、健康診断、成人病に関する業績が多くみられ、疾患では胃がん、脳卒中、糖尿病、肝炎、循環器疾患であった。

第1期の報告書の全文章を統合した共起ネットワークの構造は、「血」と「健康」が中心

a) 健康年齢を引き上げるための健康診断の諸問題、b) 以前は成人病と呼ばれた生活習慣病の原因究明・予防・治療に関する問題、c) 中高年者や高齢者の保健・福祉の向上を目指したヘルスプロモーションに関する問題、d) 遺伝子解析を中心としたマクロ的な課題、e) がん（癌）の予防・治療を主とした基礎研究の課題、であった。

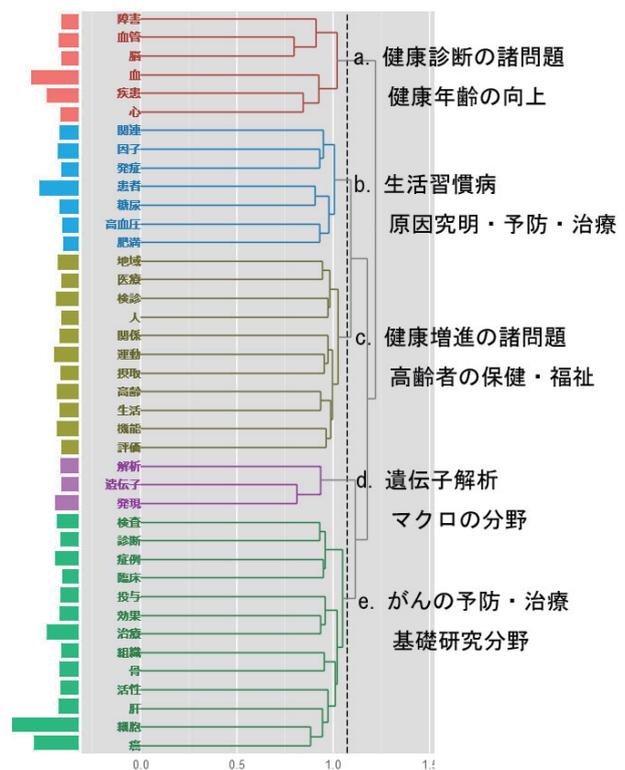


図5 全抄録集の階層的クラスター分析

考察

1) 全報告書の頻出語分析

頻出語の第1位の語は、「細胞」であり、「肝細胞」や「神経細胞」のように器官名を冠されて使用されることが多い。頻出語の第2位の語は、「血」で、血液を表現する語であるが、「虚血性心疾患」や「高脂血症」のように様々な文脈で使われている。頻出語の第3位の語は、「癌（がん）」であり、生活習慣病対策と併行してがん対策は永遠のテーマであると考えられる。

2) 各年度の上位10の頻出語

年度毎に表れる頻出語と「第1～4次国民健康づくり運動」とを見比べると、ヘルスケアに関わる専門家がどのようなテーマに関心を持ち、研究を行なうかは政策と密接に関連していることがわかる。

3) 共起ネットワークの構造分析と解釈

各期の研究を共起ネットワークにして俯瞰すると、ヘルスプロモーションの推進に腐心する研究者の姿勢やその時代に関心が集中した疾患が把握できる。

4) クラスタ分析

38年間の各年度に特徴的な言葉を探るためにクラスタ分析を行った。これにより各期の抽出語の全体的布置において、年代ごとに語の使われ方がいかに異なっていたのか、どの年代とどの年代の内容が似通っていたのかといったことが俯瞰できる。

要 約

全研究業績集の計量テキスト分析を実施することにより、各期の研究傾向や特長や重要語、キーワードが抽出され、出現頻度や同時出現関係等を分析することにより、各期の助成研究内容の変遷が整理できた。内容は、生活習慣病をはじめとした中高年・高齢者の病気の予防と治療に関する研究、また、対象が生涯を通じて健やかで充実した生活を過ごせる豊かな長寿社会を築くことに寄与する研究であった。

文 献

1. 大和証券ヘルス財団抄録集(第1～38集)
2. 樋口耕一、社会調査のためのテキスト分析、第4版、2015
3. 上田太一郎 監修、村田真樹他：事例で学ぶテキストマイニング、共立出版、2008.
4. 服部兼敏：テキストマイニングで広がる看護の世界、ナカニシヤ出版、2010.
5. 那須川哲哉：テキストマイニングを使う技術／作る技術、東京電機大学出版局、2006.
6. 湖上美喜：事例で学ぶテキストマイニング、共立出版、2008.
7. 岸江信介：テキストマイニングによる言語研究、ひつじ書房、2014.