慶應義塾大学学術情報リポジトリ
Keio Associated Repository of Academic resouces

Title	ベイズ統計を用いた医療費指数の改良 : エビデンスに基づく医療政策に向けて
Sub Title	Improvement of the health care expenditure index using Bayesian statistics : towards the implementation of the evidence-based health policy
Author	黒島, テレサ(Kuroshima, TeresaInnami, Ichiro) 印南, 一路
Publisher	慶應義塾大学湘南藤沢学会
Publication year	2010-03
Jtitle	優秀修士論文
JaLC DOI	
Abstract	平成18年度医療制度改革関連法案は,国及び都道府県に,医療費適正化計画を策定することを義務づけた.都道府県は,地域の実情に応じて計画を作成するものと定められており,その前提として詳細な医療費分析を行うことが求められている. 地域の医療費水準を相対的に把握する方法の一つに,医療費の地域差指数の算出がある.しかし,現在の算出方法では,地域別あるいは性,年齢,疾患別等の細かな区分に分けて計算する際に,分析単位の小ささから,誤差変動が大きくなるという,いわゆるSmall Area Estimation(SAE)の問題が生じてしまう. SAEに関する国内の既存研究によれば, SAE問題を回避するため,経験ベイズ法による補正は市区町村別の死亡率等に使われているが,医療費の分析では未だ実施されていない.また,国外の研究を含めると,補正方法として経験ベイズ法以外にもいくつか提案されているが,日本では適切な手法選択について十分な議論がなされておらず,また,精度を高めることを念頭に複数の補正方法が開発された結果,補正方法が高度に発達する一方で,精度と簡便性の間にトレードオフの関係が生じていることが知られている. 本研究の目的は,2つある.第1にベイズ統計を用いて医療費指数の改良を行うこと,第2に小地域推定の方法選択に関する知見を獲得することである.具体的には,はじめにSAEの補正手法に関する既存研究のレビューを行い,医療費指数の改良に有用と考えられる4つの方法(Locally Weighted Average(LWA),経験ベ イズ法(EB),階層ベイズ法(HB)のポアソンガンマモデルとポアソン対数正規モデル)があることを明らかにした. 次に,この4つの方法を実際の高血圧疾患の医療費データに適用し,補正の精度を実証的に分析した.分析結果を基に,データ区分(性・ 年齢区分)ごとに適切な補正手法と重視すべき点(精度あるいは簡便性のどちらか)を明らかにし,さらに補正前と補正後で選択される適正化重点対象地域に違いがないかを比較分析した. 本研究の分析結果から得られた結論をまとめると,以下のようになる. 高血圧疾患の医療費データでは,データ区分によっては,学術的に支持されている精度の高い手法の代わりに簡便な補正手法を代用したとしても,一定の推定精度を保つことが可能であることを示した.具体的には,高血圧疾患の発生頻度の高い年齢階層のデータ(高血圧,男性高血圧,女性高血圧,60-60歳高血圧,70 歳以上高血圧)では,簡便な手法であるLWA, EBで問題がないことがわかった,一方,① 人口規模の小さな地域のデータ(期待件数がおおむね500件以下),② 高血圧疾患の発生頻度の低い年齢階層のデータ(0-39歳, 40-49歳, 50-59歳)を補正する場合には,より精度の高い方法であるHB, EBを選択する必要があることがわかった.最後に,補正を行うことにより,適正化重点対象地域が大きく変化することを確かめ,指数の補正が必要であることを実証的に示した. 今後の課題として,誤差変動が大きくなる期待受診件数(対人口比)の規模を疾患別に明らかにすることが必要である.
Notes	医療福祉政策・ 経営プロジェクト2009年度
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=0302-0000-0636

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

ISBN 978-4-87762-235-0
SFC-MT 2009-004

優秀修士論文

べ

イズ統計を用いた医療費指数の改良

-エビデンスに基づく医療政策に向けて-

2009年度

黒島 テレサ 政策・メディア研究科 修士課程

医療福祉政策・経営プロジェクト

慶應義塾大学湘南藤沢学会

Keio SFC Academic Society

優秀修士論文推薦のことば

本論文は、実際の医療費適正化政策を進めるために不可欠な医療費指数を改良した点、ベイズ統計を医療政策分野で応用した先駆的な研究である点、さらに小地域推定問題の解決における手法間のトレードオフを部分的に明らかにしたという方法論上の貢献もある点で優れた論文である。学会専門誌にも投稿すれば採択は確実な内容とレベルを備えているため、優秀修士論文として推薦した。

慶應義塾大学
総合政策学部教授
印南一路

修士論文　2009 年度(平成 21 年度)

ベイズ統計を用いた医療費指数の改良

―エビデンスに基づく医療政策に向けて―

慶應義塾大学　大学院　政策・メディア研究科

黒島テレサ

修士論文要旨 2009年度(平成21年度)
ベイズ統計を用いた医療費指数の改良—エビデンスに基づく医療政策に向けて—

平成18年度医療制度改革関連法案は，国及び都道府県に，医療費適正化計画を策定することを義務づけた．都道府県は，地域の実情に応じて計画を作成するものと定められており，その前提として詳細な医療費分析を行うことが求められている．

地域の医療費水準を相対的に把握する方法の一つに，医療費の地域差指数の算出がある．しかし，現在の算出方法では，地域別あるいは性，年齢，疾患別等の細かな区分に分けて計算する際に，分析単位の小ささから，誤差変動が大きくなるという，いわゆるSmall Area Estimation (SAE)の問題が生じてしまう．

SAEに関する国内の既存研究によれば，SAE問題を回避するため，経験ベイズ法による補正は市区町村別の死亡率等に使われているが，医療費の分析では未だ実施されていない．また，国外の研究を含めると，補正方法として経験ベイズ法以外にもいくつか提案されているが，日本では適切な手法選択について十分な議論がなされておらず，また，精度を高めることを念頭に複数の補正方法が開発された結果，補正方法が高度に発達する一方で，精度と簡便性の間にトレードオフの関係が生じていることが知られている．

本研究の目的は，2つある．第1にベイズ統計を用いて医療費指数の改良を行うこと，第2に小地域推定の方法選択に関する知見を獲得することである．具体的には，はじめにSAEの補正手法に関する既存研究のレビューを行い，医療費指数の改良に有用と考えられる4つの方法(Locally Weighted Average(LWA)，経験ベイズ法(EB)，階層ベイズ法(HB)のポアソンガンマモデルとポアソン対数正規モデル) があることを明らかにした．

次に，この4つの方法を実際の高血圧疾患の医療費データに適用し，補正の精度を実証的に分析した．分析結果を基に，データ区分(性・年齢区分)ごとに適切な補正手法と重視すべき点(精度あるいは簡便性のどちらか)を明らかにし，さらに補正前と補正後で選択される適正化重点対象地域に違いがないかを比較分析した．

本研究の分析結果から得られた結論をまとめると，以下のようになる．

高血圧疾患の医療費データでは，データ区分によっては，学術的に支持されている精度の高い手法の代わりに簡便な補正手法を代用したとしても，一定の推定精度を保つことが可能であることを示した．具体的には，高血圧疾患の発生頻度の高い年齢階層のデータ(高血圧，男性高血圧，女性高血圧，60—60歳高血圧，70歳以上高血圧)では，簡便な手法であるLWA，EBで問題がないことがわかった．一方，①人口規模の小さな地域のデータ(期待件数がおおむね500件以下)，②高血圧疾患の発生頻度の低い年齢階層のデータ(0－39歳，40－49歳，50－59歳)を補正する場合には，より精度の高い方法であるHB，EBを選択する必要があることがわかった．最後に，補正を行うことにより，適正化重点対象地域が大きく変化することを確かめ，指数の補正が必要であることを実証的に示した．

今後の課題として，誤差変動が大きくなる期待受診件数（対人口比）の規模を疾患別に明らかにすることが必要である．

キーワード

1 小地域推定　2 医療費指数　3 ベイズ統計　4 スムージング　5 医療費適正化計画

慶應義塾大学 大学院 政策・メディア研究科

黒島テレサ

Abstract of Master's Thesis Academic Year 2009

Improvement of the Health Care Expenditure Index using Bayesian Statistics
—Towards the Implementation of the Evidence-Based Health Policy—

The Health Care Reform Act of 2006 mandated Japanese central and local governments to devise Medical Expenditure Controlling Plans, in which they need to analyze the health care expenditures of their areas to meet the local circumstances. The Health Care Expenditure Index (HCEI) is developed to clarify the relative magnitude of the expenditure of one area to other areas, yet the current HCEI is unstable when the covered population is small and the unit of analyses is too large, which is often called the Small Area Estimation (SAE) problem.

Although some previous studies used Empirical Bayes method to calibrate the SMR (Standardized Mortality Ratio) to avoid the SAE problem, but no study did to the HCEI so far. In addition, despite a variety of methods for circumventing the SAE problem have been proposed worldwide, those methods have not been fully discussed in Japan. Some methods are more accurate but too complex to used by local government staffs, thus there being a trade-off between methodological accuracy and practical simplicity.

The purpose of this study is twofold: to improve the accuracy of the HCEI using Bayesian Statistics and to acquire knowledge about how we pick up the right method to avoid the SAE problem. Specifically, literature review revealed that there are four methods to resolve the SAE problem --Locally Weighted Average (LWA) method, Empirical Bayes (EB) method, and two Hierarchical Bayes (HB) methods (Poisson-gamma distribution model and Log-normal distribution model). These four methods were applied to the HCEI for Hypertensive disease to observe the actual level of calibration.

The following are findings. First, for Hypertensive disease, more simple methods such as the LWA and EB can substitute more complex methods such as HB without losing the accuracy in general. For example, you can safely use LWA and EB when you modify the HCEI for segments on age groups which has large frequency of high blood pressure (male, female, 60-69 years, over 70 years). Second, there are, however, exceptions. You need to employ HB methods to calculate the HCEI for segments in small population sizes (less than about 500) and on age groups which has low frequency of high blood pressure (0-39 years, 40-49 years, 50-59 years). Third, the study showed empirically that different calibration methods lead to different policy targets to optimize the total health care expenditure, thus calibration itself is very much needed. The limitation of the study and future research direction are discussed.

Key Word
1 Small Area Estimation 2 Health Care Expenditure Index
3 Bayesian statistics 4 Smoothing 5 Medical Expenditure Controlling Plans

Keio University Graduate School of Media and Governance
Teresa Kuroshima

目次

1. 研究背景と目的 1
1.1. 医療費の分析単位の細分化の必要性 1
1.1.1. 医療費適正化計画 1
1.1.2. 求められる医療費分析と現在の医療費分析の実態 3
1.1.3. 医療費指数の改良と Small Area Estiomation(SAE) 7
2. Small Area Estiomation(SAE)に関する既存研究 9
2.1. Small Area Estiomation(SAE)の定義 9
2.2. 既存研究 9
2.3. SAE の手法分類 10
2.4. 手法の精度と簡便性 11
3. 本研究の目的と意義 15
4. 分析方法 16
4.1. 分析データ 16
4.2. 分析の流れ 21
4.3. 補正手法 22
4.3.1. Locally Weighted Average (LWA) Method 22
4.3.2. Empirical Bayes(EB) Method 24
4.3.3. Hierarchical Bayes(HB) Method 26
5. 分析結果と考察 28
5.1. 補正手法の選択に関する分析 28
5.1.1. 分析結果 28
5.1.2. 考察 50
5.2. 適正化重点対象地域の選定 54
5.2.1. 結果 54
5.2.2. 考察 57
6. 研究のまとめと課題 58
参考文献 61
謝辞 65

図 1 国民医療費と対国民所得比の年次推移 2
図 2 社会保障費, 国民所得の動向 2
図 3 医療費の分析単位の実態 5
図 4 人口規模による医療費指数の変動 8
図 5 Model-based Method と Locally Weighted Average Method の効果の比較 13
図 6 推定精度と手法の複雑さによる手法の位置づけ 14
図 7 生活習慣病の医療費の割合 17
図 8 疾患別受診率(合計・性別) 18
図 9 疾患別受診率(年齢階層別) 19

図 10 高血圧と脳卒中の性・年齢階層別受診率 19
図 11 実績件数，期待件数と被保険者数 20
図 12 各データ区分の期待件数の要約統計量 20
図 13 分析の流れ 22
図 14 Locally Weighted Averages Method による重み付け 23
図 15 データ区分の違いによる各手法の指数の補正結果 30
図 16 データ区分の違いによる各手法の指数の補正結果 30
図 17 すべてのデータ区分における期待件数と補正差 31
図 18 データ区分別の期待件数と補正差 32
図 19 データ区分別の期待件数と補正差 33
図 20 データ区分別の補正差の標準偏差 34
図 21 手法別の単純指数との平均二乗誤差 36
図 22 手法間の平均二乗誤差 36
図 23 手法間の補正結果の比較 37
図 24 階層ベイズと経験ベイズの補正差と期待件数 38
図 25 0－39 歳高血圧，医療費指数の補正結果 50
図 26 高血圧，医療費指数の補正結果 51
図 27 男性高血圧，医療費指数の補正結果 51
図 28 手法とデータ区分の位置づけ 52
図 29 各手法による重要地域の選定結果(0－39 歳，高血圧) 57
図 30 手法とデータ区分の位置づけ（再掲） 59
図 31 各手法による重要地域の選定結果(0－39 歳，高血圧)（再掲） 60

表 1 医療費分析の指標，単位，比較対象の種類 4
表 2 各手法の概要と利点・欠点 14
表 3 各データ区分の要約統計量 20
表 4 手法別の単純指数との平均二乗誤差 36
表 5 手法間の平均二乗誤差 37
表 6 階層ベイズ法によるモデルの DIC 値 39
表 7 山形県 高血圧データの補正結果 42
表 8 山形県 男性 高血圧データ補正結果 43
表 9 山形県 女性 高血圧データ補正結果 44
表 10 山形県 0-39 歳 高血圧データ補正結果 45
表 11 山形県 40-49 歳 高血圧データ補正結果 46
表 12 山形県 50-59 歳 高血圧データ補正結果 47
表 13 山形県 60-69 歳 高血圧データ補正結果 48
表 14 山形県 70 歳以上 高血圧データ補正結果 49
表 15 標準偏差分類による重要地域の選定(0－39 歳高血圧) 56

1. 研究背景と目的

1.1. 医療費の分析単位の細分化の必要性

1.1.1. 医療費適正化計画

平成18年度の国民医療費は33兆1276億円であり，国民所得に対する割合は約8.88%に達している．また，国民所得に対する割合は年次推移でみても上昇の一途を辿っており，平成 2 年度から平成18年度にかけて約6%から約9%へと増加し，国民所得の伸びを上回る勢いで増加し続けていることが問題視されている(図 1参照)．さらに，このような傾向は今後も高齢化等の影響により継続することが予想され，超高齢化社会の到来を前に国家財政の見地から国民医療費の伸びと国民所得との均衡を図ることが求められている．

さらに，医療費や年金を含む社会保障給付費の総額は国民医療費と同様，増加の一途を辿っており，平成20年度には95.7兆円，対国民所得比でみれば24.9%を示している(図 2参照)．社会保障給付費の中でも，高齢化の影響から年金給付費や福祉関連費の伸びを抑えることは難しいと考えられるため，政策的に対応する余地が大きいと考えられる医療費について，その動向を捉え，またその伸びを適切な範囲に抑えることが国家財政上重要な課題となっている．

このような国民医療費の増加を引き起こした背景には，医療提供体制の整備と充実，老人医療費の無料化など保険給付費等の充実に加えて，近年では高齢化や医療技術の高度化，さらに生活習慣病患者の増加等が指摘されている(平成19年度版 厚生労働白書)．このように，医療費の増加を引き起こした原因としては複数の要因が考えられるが，医療提供体制の充実や医療技術の進歩については，国民の健康状態の向上に寄与するものと考えられるため，それによって引き起こされる医療費の増加は必然の結果であり，また適切な範囲に抑えるべき対象とは考えにくい．一方，生活習慣病患者の増加については，国民の健康状態の向上にも寄与すべく適切な対応を行い，患者数の減少を目指すことは，政策的視点からも国民の健康水準の維持向上という視点からも妥当だと考えられる．

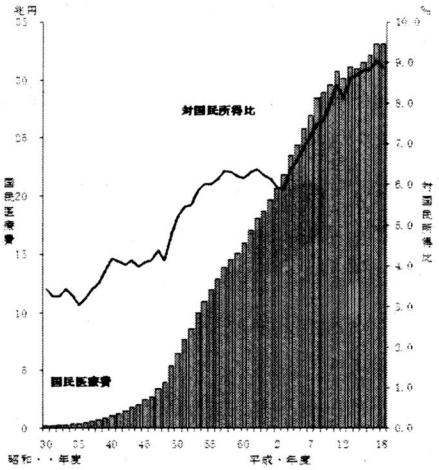
このような現状を受け，わが国では平成 18 年度医療制度改革関連法案により，計画的な医療費の適正化対策を推進することとして，医療費適正化計画が策定された．医療費適正化計画は，医療給付費の伸びと国民負担の間の均衡の確保を目的に，患者自己負担の見直し等を中心とした短期的な方策だけではなく，中長期的な医療費の適正化を目指すものとされている．そして，医療費適正化を総合的かつ計画的に進めるために，国は全国医療費適正化計画と医療費適正化に関する施策についての基本的な方針を，また都道府県は国の方針に基づき都道府県医療費適正化計画を策定することと定められている．

具体的に，全国医療費適正化計画と都道府県医療費適正化計画は，平成 20 年度を初年度とする5か年計画として策定されており，その政策目標として，①国民(住民)の健康の保持の推進，②医療の効率的な提供の推進を掲げている．また，平成24年時点には，国民(住民)の健康の保持の推進に関する目標として，①特定健診の実施率を 70%以上とすること，②特定保健指導の実施率を 45%以上とすること，③メタボリックシンドローム(内臓脂肪症候群)の該当者及び予備群の減少率を 10%以上とすることなどの目標が定められている(平成19年度版 厚生労働白書)．

このように具体的な目標の中に，医療費の削減という文字はなく，健康状態の改善に寄与する旨の目標が多く並んでおり，医療費適正化の推進の主な目的は，短期的な医療費削減ではなく，住民の健康を改善することにより，結果として中長期的な意味で医療費の適正化を行うことであると解釈する方が適切であろう．特に，医療費適正化に関する施策についての基本的な方針(案)にも記述されてい

るように，国民医療費の約 3 分の 1 を占める老人医療費は，生活習慣病の外来受診率の上昇と共に引き起こされた一面があることから，生活習慣病患者への予防あるいは重症化予防に注目が集まっている．

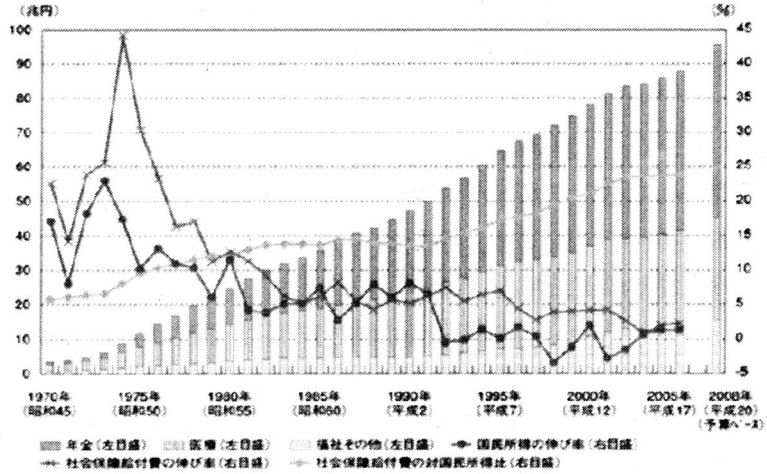
さらに，医療費適正化計画の策定にあたっては，各都道府県自らが医療費を含めた医療の地域特性を調査，分析，評価し，実態に即した医療費適正化の取り組みを行うことと定められている．したがって，これまで以上に，疾病の発症予防や重症化予防といった予防対策が重要となってきているのと同時に，各都道府県に求められる政策上の役割が大きくなり，これまでに比してより正確かつ精緻な分析を都道府県自らが行うことが必要だといえる．



出典：平成 18 年度国民医療費の概況

図 1 国民医療費と対国民所得比の年次推移

	1970年度	1980年度	1990年度	2005年度	2008年度（予算ベース）
国民所得額（兆円）（A）	61.0	203.2	348.3	367.6	384.4
社会保障給付費総額（兆円）（B）	3.5（100.0%）	24.8（100.0%）	47.2（100.0%）	87.9（100.0%）	95.7（100.0%）
（内訳）　年金	0.9（24.3%）	10.5（42.2%）	24.0（50.9%）	46.3（52.7%）	50.5（52.8%）
医療	2.1（58.9%）	10.7（43.3%）	18.4（38.3%）	28.1（32.0%）	29.8（31.1%）
福祉その他	0.5（16.8%）	3.6（14.5%）	4.8（10.2%）	13.5（15.4%）	15.4（16.0%）
社会保障給付費の対国民所得比（B／A）	5.77%	12.19%	13.55%	23.91%	24.90%



出典：平成 20 年度版 厚生労働白書

図 2 社会保障費，国民所得の動向

1.1.2. 求められる医療費分析と現在の医療費分析の実態

1.1.2.1. 求められる医療費分析

各都道府県が医療の地域特性を細かく調査・分析・評価すること，またそれに合わせて医療費適正化計画を策定することが，医療費適正化計画の実施にあたって重要かつ必須となっていることを述べた．このような医療の実態を指し示すものとして，医療供給体制や住民の健康状態等があげられ，それらについての分析を行うことが重要であることは当然だといえる．また，実態を指し示す指標の中でも，とりわけ医療費が重要な位置を占め，その分析を実行することが重要であることは言うまでもない．医療費適正化計画を策定する主体である各都道府県は，医療費適正化計画の指針において自ら医療費の実態を把握するための分析を行うものと定められており，以前にも増して**より実態を把握するための**医療費分析が求められている．

一方で，医療費は地域，性別や年齢階層，あるいは疾患によって，その傾向が大きく異なることが知られている．これは，患者特性(性別，年齢など)によって疾患への罹患率が異なることや，疾患の種類によって平均的に生じる医療費が異なることから当然の結果だといえる．したがって，実態に即した施策を検討するためには，地域別あるいは性別，年齢階層別，疾患別など細かな区分に分けて医療費を比較し，分析する必要があるといえる．

表 1 は，実際に各都道府県が計画策定の前段階として，医療費分析を行う際に用いることのできる指標を列挙した表である．医療費の分析にあたっては，①医療費の指標，②分析単位(地域，区分)，③比較対象の 3 つをいずれに設定するのかで分析できる内容が異なる．したがって，分析目的に応じてそれらの内容を適切に選択する必要がある．

①医療費の指標については，大別すると 1 人当たり医療費をそのまま用いる方法，医療費の 3 要素[1](受診率，1 件当たり日数，1 日当たり点数)に分解して評価する方法，そして地域の人口構成の違いを考慮した形で比較できる医療費の地域差指数を用いる方法という 3 つがあげられる．また，②分析単位については，地域単位として都道府県，二次医療圏，市区町村等の 3 つのレベルで分析を行うことが事実上可能である．さらに，③比較対象については，全国レベルの比較を行う際には都道府県レベルの比較分析を，また都道府県レベルの比較を行う際には市町村レベルの比較分析を行う必要がある．また，データの時点としては，時系列データの比較あるいは単年度データの比較を行う場合の 2 つがあり得る．

このように，医療費の指標，分析単位，あるいは比較対象については，複数の組み合わせが存在する．これらは分析目的に応じて随時選択することが求められるが，特に医療費適正化計画の策定にあたっては，地域別の特性を細かな対象別(性別，年齢階層別，疾患別のグループなど)に分けて把握すること重要だと考えられる．なぜなら，1.1.1 でも述べたように，医療費適正化計画では実態に即した生活習慣病等の予防施策に力点が置かれており，実態に即した予防対策を講じるためには，地域ごとにどのような疾患，あるいは年齢階層や性別に問題があるのかを適切に評価しなければならないためである．これらの細かな分析に基づいて中長期的な予防対策を講じることにより，はじめて政策の効果が期待できると考えられ，反対に地域や対象別などの細かな分析による詳細な実態の把握がなければ，誤った対象に政策的な費用を投じることになる．したがって，これまでの医療費分析にも増して，より細

---

[1] 1 人当たり医療費は，①(医療機関への)受診率，②1 件当たり日数，③1 日当たり医療費の積で表わすことが出来る．つまり，1 人当たり医療費=1 日当たり医療費×受診率×1 件当たり日数となる．これら 3 つの要素を医療費の 3 要素と呼ぶ．

かな地域や対象別の実態把握が，医療費適正化上不可欠な要素となっている．

そして，このような分析に対しては WHO 等を含め疫学分野の政策分析で広く用いられるようになった，地理的な分布を視覚的に把握するための GIS や空間統計を用いた地図による分析も重要となろう．医療あるいは疫学分野において，GIS や空間統計を用いた代表的な分析手法に疾病地図があげられる．疾病地図は病気の発生状況等を地理的な位置関係を考慮に入れて分析できる点で有用だと考えられており，WHO や発展途上国などの疫学分野の政策研究では今やスタンダードな手法となっている．医療費の分析においても，この手法は応用できると考えられ，医療提供体制や疾病への罹患率，あるいは疾病別の死亡率等の地理的な状況を加味しながら，医療費の実態を分析することに大きな意味があるものといえる．

GIS や空間統計を用いた地図による分析の利点は，データを眺めているだけでは決して把握することができない，地理的な位置関係を考慮に入れた分析を行えることである．医療に関わるデータを取り扱うような場合には，市町村等の隣接した地域において，医療供給体制や患者特性，あるいは社会的な状況も類似した傾向にあることから，地理的な隣接関係を考慮に入れて分析を行うことが重要である．これらの要素を考慮することにより，これまで市町村や都道府県の保健師等が行ってきた分析よりも，はるかに優れた詳細な分析が可能となるであろう．[2]

表 1 医療費分析の指標，単位，比較対象の種類

医療費の指標		一人当たり医療費	一人当たり医療費
			受診率
			一件当たり日数
			一日当たり点数
		医療費指数	
分析単位	地域	都道府県	
		二次医療圏	
		市町村	
		区、村、郡など	
	区分	診療区分	入院・入院外合計
			入院
			入院外
		性別	
		年齢階層別	
		疾患別	
都道府県比較		時系列データ	
		単一年度データ	
都道府県内の二次医療圏あるいは市町村等の地域別比較		時系列データ	
		単一年度データ	

1.1.2.2. 医療費分析の実態

先に示したとおり，医療費の指標，あるいは分析単位や比較対象については，複数の組み合わせが存在する．これらは分析目的に応じて随時選択することが必要である．

しかし，現状の医療費適正化計画策定に際して用いられている医療費の指標や単位，あるいは比

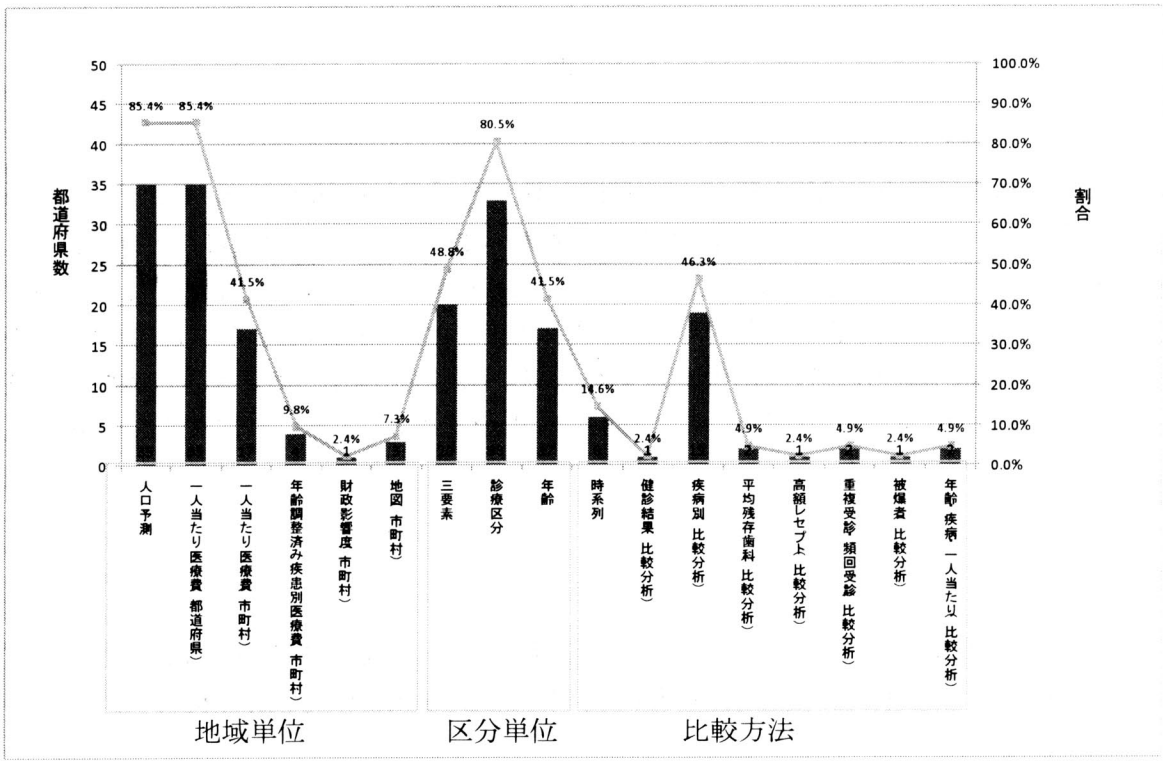
[2] より詳しい GIS や空間統計による医療分野の分析については，Waller and Gotway(2004)や Lawson(2006)を参照されたい．

較対象を含めた分析の実態は，詳細な地域特性を把握するのに十分だとは言い難い．古城・印南(2009)は，医療費適正化計画の作成状況に関する実態調査を行い，医療費適正化計画の策定に際して，各都道府県がどのような地域単位，あるいは区分単位を用いて医療費の現状を分析しているのかを明らかにした．古城・印南(2009)の調査結果を示したものが，図 3 である．古城・印南(2009)は，全国 47 都道府県のうち，該当する分析方法あるいは区分で分析を行った都道府県の割合と個数を示している．

分析結果から，地域単位をどのレベルに設定しているかということに着目してみると，市町村単位で分析しているのが41.5%となっており，17都道府県のみしか行われていないことがわかる．また，これらのいずれも 1 人当たり医療費を指標として用いて分析を行っている．

どのような区分に分けて医療費を分析しているかという点に着目すると，例えば入院と入院外などのように診療区分ごとに 1 人当たり医療費を分けて分析している都道府県は，33 都道府県であり，全体の約 80%であることがわかる．また，医療費を 3 要素に分けて分析を行った都道府県は 20 都道府県であり，全体の 48.8%であった．さらに，年齢区分に分けて分析を行った都道府県は 17 都道府県であり，全体の 41.5%であった．

さらに，疾病別の医療費分析を行った都道府県は，19 都道府県であり，全体の 46.3%であった．また，年齢・疾病・1 人当たり医療費のように，細かな対象別に医療費の高低を比較した都道府県はわずか 2 都道府県であり，全体の 4.9%であった．年齢調整済み疾患別医療費を用いたのは，4 都道府県であり，全体の 9.8%と 1 割にも満たしていない．



出典：古城・印南(2009)

図 3 医療費の分析単位の実態

上記の結果から，医療費の地域単位では都道府県単位が最も多く分析が行われており，指標としては 1 人当たり医療費が多く用いられていることがわかった．また，年齢階層別あるいは疾患別などの細かな対象別に医療費を分析した都道府県は少ないことがわかった．

しかし，都道府県が計画を策定する際にみるべきは，管轄する市町村間で医療費の高低にどの程度の差が生じているのかであり，またその水準は年齢階層や性による差があるのか，あるいは疾患ごとに異常に高い地域あるいは低い地域があるのかといった相対的な医療費水準についての実態である．このように，相対的な医療費水準の比較分析を細かな区分あるいは地域別に行うことが出来てはじめて，医療費の実態を詳細に把握したものといえ，それに即した施策立案が可能となる．さらに，医療費の指標として最も多く用いられていた 1 人当たり医療費は，地域の相対的な医療費水準を把握するのには適切でない．各都道府県や市町村の人口規模が異なる場合の比較には有用と考えられるが，他の地域や対象と比べて，相対的な水準の高低を把握し，水準の差を複数地域で同時比較することは難しい．なぜならば，1 人当たり医療費は，絶対的な額であるため，医療費の水準を評価するためには，医療費の絶対額を比較する対象別にその都度差し引き，比較することが必要だからである．

そこで，相対的な医療費水準を複数地域あるいは対象別に同時比較するのに，有用な指標が医療費の地域差指数(以下，医療費指数)である．医療費指数は，地域の医療費の利用水準を相対的に把握する方法の一つとしてあげられる．また，政策の判断材料として医療政策や健康政策のエビデンスとして用いられ，各地域の人口の年齢構成要因を除去し，相対的な医療費の高低を評価するために用いられている．なお，医療費指数は疫学分野でよく用いられる相対リスク(相対危険あるいは Relative Risk) [3]と同様の形で算出され，具体的に以下の式から算出できる．

$$\theta_i = \frac{O_i}{\sum_{j=1}^{J} \hat{\theta}_j N_{ij}} \tag{1.1}$$

ここで，θ_iは医療費の地域差指数，O_iは実績医療費，$\sum_{j=1}^{J} \hat{\theta}_j N_{ij}$は期待医療費，$\hat{\theta}$は 1 人当たり医療費，そして$N_{ij}$は地域 i の年齢階層 j における被保険者数を示している．また，i は i 地域を，j は j の区分(例えば，75 歳以上の人口など)を示す．さらに，実績医療費とは，該当地域あるいは区分の人口において，実際にかかった医療費を示す．また，期待医療費とは，該当地域あるいは区分の人口構成が平均的な集団と仮定した場合に，生じると期待される医療費を指す．つまり，期待医療費は該当地域あるいは区分の 1 人当たり医療費に該当地域あるいは区分の人口を乗じたものである．具体的には，以下の式から算出できる．

$$E(\hat{\theta}_i) = \frac{O_i}{\hat{E}_i} \ , \ \hat{E}_i = \sum_{j=1}^{J} \hat{\theta}_j N_{ij} \tag{1.2}$$

さらに，厚生労働省は次のように医療費指数を説明している．高齢者の医療費は若年者に比べ高いため，被保険者の高齢化が進んでいる市町村は1人当たり保険給付費が高くなる．そこで，各市町

[3] 相対リスク(相対危険あるいは Relative Risk)とは，疫学分野では通常罹患率の比を表わす指標として用いられている．罹患率の比とは，暴露群の患者が非暴露群の患者に対して，何倍の罹患リスクを抱えているのかを示す．つまり，暴露群の患者の罹患リスクを評価するための指標であり，
相対リスク = 評価したい群のリスク/基準とする群のリスクという形を用いて算出される．

村の年齢構成の相違による保険給付費への影響を補正し，その他の要因による保険給付費の地域格差を指数化したものが地域差指数である．また，医療費指数は，当該市町村の実績の保険給付費を基準給付費(当該市町村の5歳階級毎の被保険者数に，全国平均の5歳階級毎の1人当たり保険給付費を乗じた額の総計による保険給付費)で除して算出される．なお，地域差指数を計算する際の保険給付費とは，一般と老人の保険給付費を指しており，老人分については加入者調整をした老人保健医療費拠出金ベースの和となっている．

言い換えれば，医療費指数は医療費が各地域や区分に属する人口の特性(例えば，年齢階層や性別)によって変化するという事実に基づいて，それらの影響を除去し，人口特性の要因を気にせず比較することを可能にした指標だといえる．また，医療費は疾患によっても，施される治療が異なることから，疾患別に医療費を比較する際にも医療費指数は有用な指標となる．

指数は，1.0 を平均値として上下変動するが，厚生労働省が公表している医療費指数は，全国値を平均(1.0)として算出が行われている．ただし，都道府県全体の値を平均値とするなど，分析目的に応じて，平均値は変化させることができる．例えば，7 つの都道府県のデータが手元にある場合，それらの 7 つの平均値を基準値として，算出を行うことが可能である．この場合，7 つの都道府県内での相対的な高低を比較する指数が算出される．ただし，医療費指数はあくまで相対的な高低を示す指数であるため，基準とする平均値によって算出される値は異なることに注意が必要である．

1.1.3. 医療費指数の改良と Small Area Estiomation(SAE)

市町村や都道府県が地域の実情に合わせて政策立案を行うためには，区あるいは市町村，または二次医療圏など細かな地域ごとに分析を実行する必要があることを述べた．また，患者特性（性別，年齢など）によって疾患への罹患率は異なり，疾患ごとにかかる医療費も異なるため，性別，年齢階層別，疾患別などの細かな区分に分けて医療費を比較する必要があるとも指摘した(1.1.2 参照)．したがって，地域の実情を詳細に把握するためには，地域別かつ細かな対象別(性，年齢階層，疾患別)に医療費を分析する必要があるものといえる．

そこで，上記で示した相対的な医療費水準を比較するのに有用な医療費指数を，地域別，性・年齢階層別・疾患別とデータ区分をより細かくして算出してみることとしよう．すると，以下にあげた図 4 のような結果が得られる．図 4 は山形県国民健康保険のレセプトデータから得られた医療費データを市町村別の指数として算出したものである．グラフの縦軸は医療費指数，横軸は被保険者数を示している．

この結果から，被保険者数が少ない小さな市町村の医療費指数であるほど，推定値の値が大きくばらついていることがわかる．被保険者数が少なくなるほど，つまり左に位置するほど指数に変動が生じているものといえよう．

これは推定区分を小さくするほど，分析単位となる地域あるいは区分の人口が少なくなる，またはデータの発生率が小さい区分が生じてしまうために，データの発生頻度が必然的に小さくなり，推定上の誤差が生じるという問題である．このように推定精度が一律でない，言い換えれば誤差の生じ方が一様でない指標を政策判断材料として用いることは適切ではない．したがって，人口規模に左右されない安定した精度を有する指標を推定することが必要だといえる．

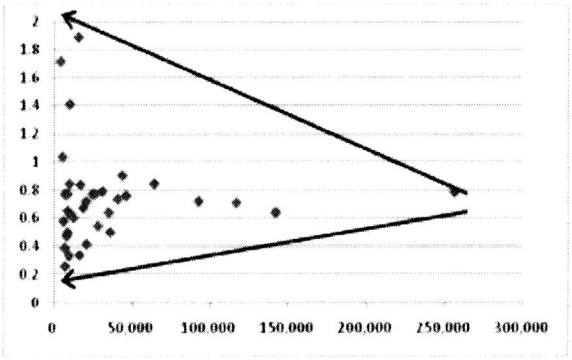
このように，地域あるいはデータの区分の仕方によって推定精度が不安定となる問題は，Small

Area Estiomation(SAEあるいは小地域推定)の問題[4]と呼ばれる．ここで，以下の例を用いてSAEの問題を具体的に考えてみよう．

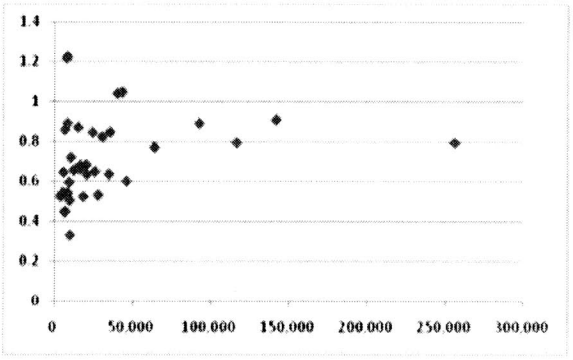
例えば，2 つの地域について比較分析を行う際に，1 人の死亡者が発生したと仮定する．人口規模が大きいA地域では，500 人の期待死亡数を持つのに対して，もともと観測された 100 人の死亡者+1 人の死亡者が発生したと仮定する．すると，SMR(標準化死亡比)は(100＋1)/500 となるため，結果的にSMRが 20%から 22%へと変化が生じる．一方，人口規模の小さな市町村であるB地域においては，5 人の期待死亡数に対して，もともと観測された 1 人の死亡者に加えて 1 人の死亡者が発生したと仮定する．すると，SMRは(1＋1)/5となるため，20%から 40%へと変化する．つまり， 同じ 1 人の死亡者数が発生したのに対して，人口規模が大きい地域と小さい地域とでは，指数へのインパクトが大きく異なる(2%の変化と 20%の変化)ことがわかる．市町村の人口規模や，データの区分に該当する対象者が少ない場合に，このような推定上の問題が生じるため，適切な補正手法を用いることにより，推定結果を補正することが必要となる．

しかし，現在行政が公表している医療費指数については，SAE の問題に対処したものは筆者が調べた限りでは見当たらず，同様に既存研究においても，医療費指数を題材に SAE の問題に対処した研究はなされていない．また，医療費指数を市町村単位で性別，年齢階層別，疾患別に推定した研究は古城，印南(2009)による研究しかみられないが，この研究においても SAE の問題を課題としてあげており，SAE の問題に対処するには至っていない．

したがって，本研究において医療費指数をより詳細な地域あるいは区分別の分析に用いることのできる，安定した精度を有する指数へと改良すべく補正を行うことに意義があるものといえる．以下では，SAE について理解を深めるため，SAE に関する既存研究のレビューを行うと共に，SAE の定義とその補正手法を整理する．



腎不全



糖尿病

データ出典:「国，都道府県の医療費適正化計画の重点対象の発見に関する研究」より山形県国民健康保険平成 18 年度 5 月診療分

図 4 人口規模による医療費指数の変動

(縦軸:医療費指数，横軸:被保険者数)

[4]日本では，小地域推定の問題等と呼ばれるが，本研究では Small area estimation として統一的に扱う．

2. Small Area Estiomation(SAE)に関する既存研究

2.1. Small Area Estiomation(SAE)の定義

Small Area Estiomation(以下, SAE)[5]とは, 補正を行わない直接推定(Direct estimation)では十分な精度を保つことのできない地域あるいは小区分のデータに対して, 安定した精度を提供する手法だと定義されている(Rao(2003), EURAREA Consortium(2004), Jiang and Lahiri(2006), Best et al.(2008)). また, McEwin and Elazar(2006)によれば, SAE とは信頼出来る推定値を生成することに加えて, より詳細な地域レベルで推定を行うことや横断分類(cross-classification)を行うことも含むものとされている.

Small Area(小地域)という言葉が示す通り, SAE は細かな地域の推定精度を向上させることと捉えられるが, 本質的には推定上の問題が生じるのは細かな地域だけではない. この点について久保川(2006)は, SAE の問題は市, 群, 町, 村などの地理的に小さい地域を指すばかりでなく, 広い地域の母集団から特定の年齢, 性, 人種などの分割された部分母集団を意味することもあると指摘している. また, 中田ほか(2008)は, 母集団の大きさに比して発生率が極めて小さい場合と母集団が小規模の場合, つまり, 小地域と小発生率のときに SAE の誤差が生じる可能性があると指摘した.

したがって, SAE の問題の核心は, 分析対象とする地域の大きさが問題なのではなく, 得られたデータが小さいことによって, 安定的に精度の高い推定が困難になることにある. したがって, "Small Area Estimation"だけではなく"Small Segment Estimation"も推定精度に関する問題を同様に生じさせることに注意しなければならない.

分析対象のデータが小さくなる原因は, 分析単位を細かくすることや 1 回の発生頻度に対してデータが与えるインパクトの大きさに関係している. 分析単位は地域だけでなく, 本研究で扱うデータのように性別や年齢, 疾病の種類というように, 分析単位が詳細であればあるほど, 統計的な変動が大きくなり, SAE の問題が生じるものといえる. 分析単位の細分化は, 特定疾患の死亡数・死亡率・受診件数のように, 様々な区分があり得るが, 所得や人口, 医療費のように分析する桁数が大きければ統計的な変動を生じないようなデータの場合にも, 小さな区分とすることで統計的な誤差変動が生じ, SAE の問題が生じる.

したがって, これらの SAE の問題の生じやすさを示す指標として, 多くの研究で用いられている人口規模ではなく, 性別や年齢階層別のデータに対して期待されるデータの値を用いる方が本質的であり, 適切だと考えられる. つまり, 発生頻度を示すデータの大小, 本研究でいえば期待される受診件数によって, SAE の補正の必要性を考えるべきだといえる. 実際に, 佐伯ほか(2005)も市町村別死亡率の補正を行う際に, 人口規模と補正結果を比較するのではなく, 期待死亡数と補正結果を比較しており, 補正によって生じる値の差が期待死亡数と関係することを示した. そこで, 本研究においても以下の分析で SAE の補正の必要性や補正結果の比較を行う際には, 人口規模の大小ではなく期待される受診件数との比較を行うこととする.

2.2. 既存研究

SAE について解説を行った主要な文献として, Rao(2003), Waller and Gotway(2004), Satorra and Ventura(2006), SAE package developers(2007), Best et al.(2008), Chandra et al(2009)

[5]わが国ではの研究では小地域推定の問題等と呼ばれているが, 本研究では Small area estimation(SAE)として統一的に扱うこととする.

があげられる. Rao(2003)は, SAE の様々なモデルについて非常に詳しく, かつ網羅的に解説している点で有用な文献であり, SAE の問題を概観する上で役立つ文献だといえる. SAE package developers(2007)は, SAE のベイジアン[6]以外の手法について簡潔に説明を行っており, サンプルデータを用いて例示を行っている点で有用な研究だといえる. また, SAEのベイジアンによる手法に特化した文献として, Lawson et al.(2003), Carlin and Louis(2009)があげられる. 国内の研究としてSAE の説明を行ったものに, 久保川(2006), 中田ほか(2008)がある.

SAE は 1980 年代以降研究が盛んとなり, それ以降, SAE の問題を応用した研究や手法自体の発達が加速したものと考えられる. その理由として, 精度の高い小地域データへのニーズが高まったことがあげられる(Rao(2003)). 以前は, 国レベルのデータや主要な領域, そして大きな領域のデータで満足していたものの, 現在は政策決定のため, 政策担当者がより小さな地域の推定値を必要とするようになったと指摘されている(Rao(2003)). このような状況はわが国についても当てはまり, 地方分権化が進展し, 地方の実態をより詳細に把握するニーズが高まっていると考えられるため, 国内においてもSAE の研究が様々な分野でなされ, より多くの研究事例や解説書が増えることが望まれよう.

SAE の応用領域としては, 疫学分野が多く, その他では所得や貧困問題(Income and Povety), 雇用問題(Monthly Estimates of Employment and Inemployment), 農作物に関する問題(County Estimates of Crop Yeild), アルコール問題(Substance Abuse in States and Metropolitan)などに応用されている(Rao(2003)). SAE の応用領域は多岐に渡るため, その結果数多くの手法が提案されるに至っている. しかし, その一方で各種のモデルは, より一層複雑となっていることも事実であり, 著者によって分類の仕方が異なることから, SAE の手法にどういったものがあり, どのような場面でどの手法を用いれば良いのかが分かりづらい現状となっている. そのため, SAE の問題に精通した者でなければ, 適切な手法を選択しづらくなっていることが, 実用上の壁となっている可能性がある.

そこで, 以下では Rao(2003), McEwin and Elazar(2006), SAE package developers(2007), Saei and Chamber(2003), Best et al.(2008)による SAE の分類に基づき, それらを統合して SAE の手法を大まかに分類し, その特徴について述べることとする.

2.3. SAE の手法分類

SAE の手法は, 大きく分けて①Direct Estimator(Design Based Estimator の一部)と②Model Based Estimator(Indirect Estimator とも呼ばれている)の 2 つに分けられる[7]. これら 2 つの方法は, 補助データ(Auxiliary data)を用いるか否かによって分けられる(Saei and hamber(2003)). ①Direct Estimator は補助データを用いずに推定を行う手法であり[8], ②Model Based Estimator(Indirect Estimator)は補助データを用いて推定を行う手法である. ここでいう補助データとは, 近隣地域のデータ, あるいは異なる時点における同地域のデータを指しており, 分析者が有するデータに依存して, どちらの補助データを用いるかを決定する. 補助データを用いる②Model Based Estimator では, 時系列データを用いる場合と単年度のみのデータ(近隣地域等の共変量によるデー

---

6 ベイズ統計学の考え方を応用した手法のことを意味する.

7 ただし, 著者によって分類の軸が異なるため, SAE の手法分類には違いがある. そのため, ここにあげる分類はあくまで絶対的なものではないことに注意されたい.

8 ただし, Rao(2003)は Design Based Estimator の場合には補助データを用いる場合もあると指摘している.

タ)を用いる場合の二つがあり得る.

さらに, Model Based Estimator は①Fixed Effect Model, ②Random Effect Model, ③Linear Mixed Model の 3 つに分類することでき(Saei and Chambers(2003)), 精度がより高く, 最適な推定を行うことができる手法(Rao(2003))とされている. また, データ構造の複雑さと応答変数の性質に依存して様々なモデルを用いることができる(Rao(2003)). ①Fixed Effect Model は, 関心のある変数がそれに対応する変数で完全に説明されることを仮定するモデルであり, ②Random Effect Model は補助データ(他の変数)ですべては説明されず Random Effect(混合効果)を伴うモデルとされている(Saei and Chamber(2003)). さらに, ②Random Effect Model は Empirical Best Linear Unbiased Model(EBLUP)や Empirical Bayes Method(経験ベイズ法), Hierarhical Bayed Method (階層ベイズ法)の 3 つを含んでいる. 経験ベイズ法と階層ベイズ法[9]とは, ベイズ統計の考え方を SAE の手法に応用したものであり, 事後分布の近似を求めるものを経験ベイズ法, またそれ以外のすべての手法をフルベイズ(Full bayes)法として分類できる(Lawson et al.(2003)). ベイジアンによる手法は, 他の手法に比べて, 柔軟なモデリングが行える[10]という点で優れていると考えられ, Gomez-Rubio et al.(2009)が指摘しているように, 首尾一貫したフレームワークのもと分析を進めることが出来る点で優れている. また, 階層ベイズ法は SAE の中でも最も新しい手法(Trevisani and Torelli(2004))とされており, 簡単なモデルから複雑なモデルまで多様なモデルが提案されている.

2.4. 手法の精度と簡便性

このように, 大まかにみても SAE の手法は多種多様な手法とモデルが提案されている. 各々の手法やモデルには利点と欠点があり, 補正に用いる際には十分な議論が必要だといえる. 特に, 手法の精度と簡便性については, 各々の手法で一長一短があげられ, 政策現場で用いることを念頭に置けば, 十分な考慮が必要だといえる. McEwin and Elazar(2006)が指摘したように, SAE のモデル化された推定値の質と手法の簡便性の間には, トレードオフの関係が存在している. したがって, エビデンスとしての利用のしやすさを考慮に入れると, いずれの手法が適切かについては難しい選択である. 指数についてより良い精度を重視するか, 政策実務家が利用することを考え, 計算上の複雑さを回避し, ある程度の精度の安定性で妥協するかについては議論の余地が残るであろう. 手法によって補正結果となる推定値に大きな差がみられる場合には, 精度を重視した手法選択を行うべきであるし, 反対にあまり差がないならば, 実用性という観点から使いやすさを重視した手法選択が望まれる.

しかしながら, 国内の研究では, 一部の手法のみが用いられている現状にあり, 手法選択について十分な議論がなされていない. 応用研究として, 平子ほか(1999), 佐伯ほか(1999), 相田ほか(2004), 佐伯ほか(2005)等があげられるが, いずれも経験ベイズ法による手法を用いて補正が行われている.

経験ベイズ法はこれまでに提案されてきた Model Based Estimator の中でもより精度が高いとされるベイジアンの手法に属するため, 補正結果の精度がある程度改善させるものと考えられるが, 一方で補正結果が過度に収束しすぎる(Overshrinkage)ことや推定値の分布が真のリスクの分布よりも狭い傾向にあることが欠点として指摘されている(Devine et.al(1994)). また, 経験ベイズ法よりも優れた手

---

[9] 両手法の詳細については, 4.3 分析手法にて説明を行う.

[10] この点については, Clayton et al.(1993), Cislaghi et al.(1995), Clayton and Bernardinelli(1992), Olsen et al.(1996)を参照のこと.

法として，階層ベイズ法による補正が既に提案[11]されており，経験ベイズ法と階層ベイズ法のいずれを用いるべきか，政策的な観点から検討することに意義があるものといえよう．経験ベイズ法は，精度という観点では階層ベイズ法に劣るが，実用性という観点で有用であるとも指摘されている(Aylin et al.(1999))ため，学術的に最も推定精度が高いとされる階層ベイズ法とどちらを選択すべきか実証的に示すことが重要だといえる．

階層ベイズ法が経験ベイズ法よりも優れている点として，パラメータの不確実性を考慮に入れて推定する点があげられる(Gomez-Rubio and Lopez-Quilez(2006))．経験ベイズ法はデータから事前分布の値を一意に定めるが，階層ベイズ法は MCMC 法によるサンプリング手法を用いて分布を形成するため，パラメータの不確実性をより考慮している(Bernardinelli and Montomoli (1992))と考えられるためである．経験ベイズ法は，スムージングに用いるパラメータを点推定により求めるため，真に正確な値を算出出来ず，MCMC 法というサンプリング手法を用いた階層ベイズ法は，推定に影響を与えるパラメータの不確実性をより表わすことができるという点に起因している．

しかし，階層ベイズ法による推定では，上記でも記述したように WinBUGS 等のソフトウェアにより MCMC 法を用いて推定を行うため，コンピュータによる計算負荷が高い．この点については，Rao(2003)も階層ベイズ法による計算は正確だが，MCMC 法を用いるため，計算が非常に複雑だと指摘している．

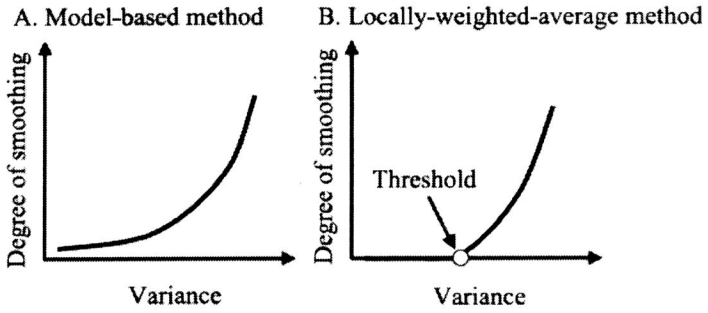
そこで，本研究では，医療費指数の改良を行うため上記 2 つのベイジアンによる方法(経験ベイズ法と階層ベイズ法)を用いて SAE の補正を行うこととし，推定精度と簡便性のどちらを重視して，手法を選択を行うべきかについて検討を行う．さらに，階層ベイズ法については，非常に多くのモデルが提案されており，実際に用いる場合にどのモデルが適切かを検討する必要があるため，今回の分析では，ポアソンガンマモデルとポアソン対数正規モデルの 2 つのモデルを用いることとする．なお，既存研究においては，ポアソンガンマモデルよりもポアソン対数正規モデルの方が空間的な相関を扱えるため，より優れたモデルとして提案されている(Richardson and Spiegelhalter(1996), Lawson(2003))．

上記にあげた 2 つの手法はベイジアンによる手法である．そのため，ベイズ統計学を知らない者にとっては敷居が高い．実際に Duell et al.(2007)においても，ベイジアンによる手法は高度な統計的知識や技術を要することから，利用が遅れていることも指摘されている．また，ベイジアンによる推定は，事前分布のロバストネスやコンピュータ計算の複雑さ(モデルの複雑さ)という観点から批判もなされている(Saei and Chamber(2003))．

したがって，非ベイジアンによる補正手法を分析に組み込むことにより，ベイジアンによる手法と比べて，どの程度推定結果に差が生じるのかを明らかにすることとした．具体的には，Locally Weighted Average Method(以下，LWA)という公衆衛生の分野で提案される補正手法(Waller and Gotway(2004))を用いる．LWA は，より単純な SAE の手法として提案されており，ベイジアンのように分布を仮定せず，複雑なモデルやコンピュータ計算を用いずに補正を行うことができる．また，LWA の方が，ベイジアンによる手法が属する Model-baesd estimator よりも計算が容易で，利用しやすいということも指摘されている(Duell et al.(2007))．さらに，Kafader(1994)による研究結果では，LWA による補正結果の方が，経験ベイズ手法による補正結果よりも真の変化(True Variation)を保ち，不自然な率の上昇を是正するという意味において，より正確な手法だとも指摘されていることから，ベイジア

[11] Bernardinelli and Montomoli (1992), Gomez-Rubio and Lopez-Quilez(2006), MacNab et al.(2004)

ンによる手法との比較に用いることに意義があるものと考えられる. ただし, LWA による補正では, 隣接関係等を用いて加重平均をとるため, 隣接市町村数が多いほど, 補正が大きくなされることが欠点の 1 つとしてあげられる. また, 都道府県等の外側に位置する市町村は必然的に隣接市町村が少なくなるため, 補正があまりなされないという特徴を持つ. また, Duell et al.(2007)においても, LWA は統計的には厳密な手法ではないと指摘されている. 図 5 は, Model-based Method と LWA の補正の効果を比較した Duell et al.(2007)による図である. この図は, 補正の効果がどの地点から始まるかを示した図であり, Model-Based Method の場合にはすべてのデータに対して補正が行われることを意味する. 反対に, LWA ではある一定以上の地点から補正が開始されることを意味しており, この地点を人口規模や地理的な関係により, 分析者が主観的に設定するという点において統計的に厳密ではないと指摘されている. したがって, LWA は簡便性を有する半面, 隣接関係によって補正の程度が異なるという欠点を持つものといえる.



出典:Duell et al.(2007)

図 5 Model-based Method と Locally Weighted Average Method の効果の比較

以上の既存研究レビューの結果から, 本研究で用いる 3 つの手法[12](LWA, 経験ベイズ法, 階層ベイズ法)の利点と欠点を整理すると, 表 2 のように示すことができる. また, 階層ベイズ法による 2 つのモデル(ポアソンガンマモデルとポアソン対数正規モデル)を含めた計 4 つの手法について, 手法の複雑さと推定精度により, 位置づけを行ったものが図 6 である. 本研究では, 以下で行う分析と考察において, これら手法の利点と欠点を踏まえつつ, 医療費適正化上, 指数を改良する際にいずれの手法を用いることが適切といえるのか検討することとする.

[12] 補正に際しては, 以下の 3 つの手法 LWA, 経験ベイズ法, 階層ベイズ法を用い, そのうち階層ベイズ法については 2 つのモデル(ポアソンガンマモデルとポアソン対数正規モデル)を用いるため, 結果として合計で 4 つの方法を用いて補正を行うこととする.

表 2 各手法の概要と利点・欠点

<table>
<tr><th>手法</th><th colspan="2">補正方法</th><th>利点</th><th colspan="2">欠点</th></tr>
<tr><td>Locally weighted average method</td><td colspan="2">①近隣地域の値との差を減らそうとする
②周辺地域の値を用いて、加重平均をとる</td><td>①四則演算のみで計算可能
②ベイズ統計学を知らなくても補正が出来る
③一定以上の分散を持つ地域のみを補正する</td><td colspan="2">①隣接市町村数が多いほど、補正がなされる
② 都道府県の外側にある市町村は必然的に隣接市町村が少ないため、補正があまりされない傾向にある
③隣接関係によって補正の程度が異なってしまう</td></tr>
<tr><td>Empirical bayes
(経験ベイズ)</td><td rowspan="2">①ベイズ統計の考え方を用いて、指数に事前分布を仮定。
②パラメータが分布を持つと仮定
③ θ iというリスクのパラメータ、つまり医療費指数がどのような地理的分布に従うのかという考えを補正に反映できる</td><td>①事前分布の形状を決定するパラメータをデータから計算する手法</td><td>①ベイジアンの手法の中でも計算方法がより簡便である</td><td rowspan="2">①高度な統計的知識や技術を要することから、利用が遅れているといわれる
②ベイズ統計学の知識がないと利用できない
③精度の高い手法を用いようとすると、複雑なモデルになってしまう</td><td>①階層ベイズに比べて、補正結果が過度に収束しすぎる
②推定値の分布が真のリスクの分布よりも狭い傾向にある</td></tr>
<tr><td>Hierarchical bayes
(階層ベイズ)</td><td>①SAEの最も新しいの手法
②精度という観点から、最も優れている
③事前分布の形状を決定するパラメータをサンプリングの手法を使って計算</td><td>①パラメータの不確実性を考慮に入れ、限られたデータから、より精度の高い推定結果を導くことが出来る</td><td>①計算するためにMCMC法を使うため、RやWinBUGS等のソフトウェアを用いる必要があるため、ソフトウェアに慣れる必要がある</td></tr>
</table>

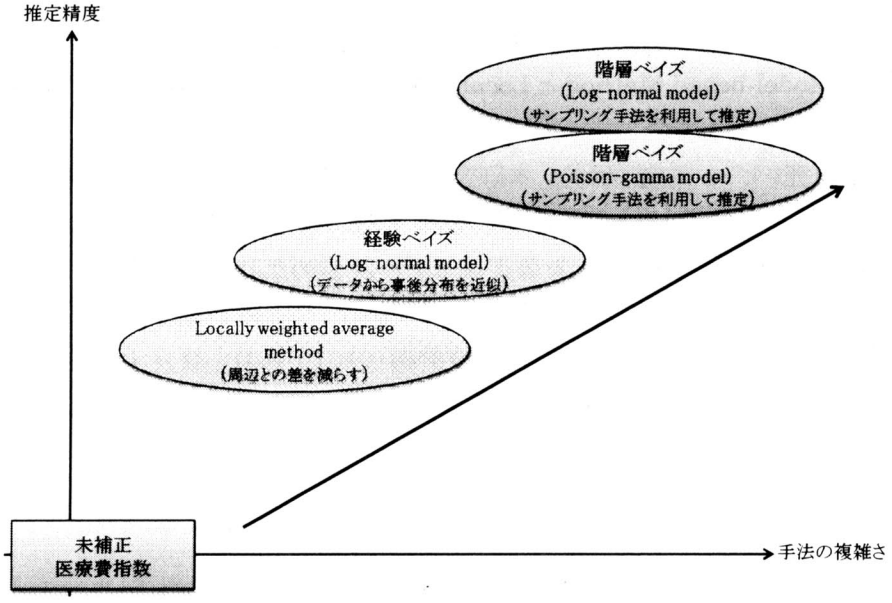


図 6 推定精度と手法の複雑さによる手法の位置づけ

3. 本研究の目的と意義

本研究の目的は，大きく分けて2つある．第1は，ベイズ統計を用いて医療費指数の改良を行うこと，第2は，小地域推定の方法選択の知見を獲得することである．

地域の医療費の利用水準を相対的に把握する方法の一つとして，医療費の地域差指数による分析が挙げられる．しかし，現在の医療費の地域差指数の算出方法では，地域別あるいは性，年齢，疾患別等の細かな区分に分けて推定した際に，分析単位の小ささから Small Area Estiomation(SAE)の問題が生じてしまう．SAE とは，分析地域の人口が少ない，あるいは分析単位が小さいために誤差変動が大きくなり，精度の高い推定結果を得られない問題である．

SAE に関する国内の既存研究では，市区町村別の死亡率等を経験ベイズ法により補正を行う研究がなされているが，医療費の分析では未だ実施されていない．また，補正の方法には，経験ベイズ法を用いた方法以外にもいくつかあるが，適切な手法選択については十分な議論がなされていない．一方，国外の研究では，精度を高めることを念頭に複数の補正方法が開発された結果，補正方法が高度に複雑化し，精度と簡便性の間にトレードオフの関係が生じている．

そこで，本研究では，既存研究のレビューから有用と判断した 4 つの SAE の手法(Locally Weighted Average(LWA)，経験ベイズ法(Empirical Bayes Method)，階層ベイズ法(Hierarchical Bayes Method)によるポアソンガンマモデルとポアソン対数正規モデル)を用いて，医療費指数の補正を行う．これらの補正を実際に行うことにより，より精度の優れた指数を算出する．さらに，これら 4 つの手法により得られた結果から，医療費適正化上，いずれの手法で改良すべきかを検討し選択する．最後に，補正前と補正後の指数から，適正化重点対象地域として選択される地域に変化が生じるのかを考察することにより，指数の補正の必要性を確認する．

本研究を通じて， 医療費指数の改良を行うことにより，医療費適正化計画の策定に際して，各都道府県が医療費の実態を細かな地域あるいは対象別に把握することを手助けすることができると考えられる．

また，精度と簡便性という2つの視点から多様な手法が提案される SAE の手法について，データ特性に応じて適切な手法を選択するという手法選択に関する知見を獲得することにより，政策現場において補正手法を用いる際の判断基準に示唆を与えるものとなろう．そして，これは医療費データに限らず他のデータに対して SAE の手法選択を行う場合にも参考となるものと考えられる．また，SAE の手法を適用する際には，共通してデータ区分やデータ特性に応じて適切な手法を選択すべきと考えられるため，医療分野に限らず SAE の手法が政策現場で応用されることをより手助けすると考えられる．

4. 分析方法

4.1. 分析データ

平成 18 年度 5 月診療分の山形県国民健康保険の市町村別レセプト集計データを用いて，高血圧疾患の受診件数に関する分析を行った．これは，平成 21 年度厚生労働科学研究「国，都道府県の医療費適正化計画の重点対象の発見に関する研究」において協力自治体から提供されたデータの一部[13]である．

本研究の目的は，第 1 にベイズ統計を用いて医療費指数の改良を行うこと，第 2 に，小地域推定の方法選択の知見を獲得することであった．したがって，SAE の問題が発生しやすい人口規模が小さい市町村や細分化した医療費（性別，年齢，疾病別など）において，ベイズ統計の有用性を実証的に明らかにすることが課題である．そのため，複数の市町村の地域差指数を改良する必要がある．また，この研究は学術的な精度の向上だけでなく，医療費適正化計画の政策資料の充実を目指しているため，医療費適正化政策上重要な疾病を分析することも必要である．

そこで本研究では，小規模自治体から比較的大きな自治体まで兼ね備えている山形県を対象にし，また予防の観点から最重要と考えられている高血圧疾患のデータを分析対象に取り上げた．たしかに，人口規模が異なる市町村が所属し，詳細なデータが入手できるのであれば，他の都道府県を分析対象に行ってもよい．条件を満たすならば他の任意の都道府県を対象にしてよく，今回はデータの入手可能性の観点から山形県を分析対象とした．また，医療費適正化の観点からいえば，一つの都道府県だけでなく，複数の都道府県，できれば全都道府県を対象に分析する方が望ましいとの指摘もありえよう．しかし，研究の目的は，政策資料としての分析手法の改良を，理論的だけでなく実証的に吟味することにあり，改良した手法を包括的に適用し，政策資料の改善を目指す部分までは，研究の射程を超えており，今後の課題とすることにした．

次に，医療費を直接分析せず，受診件数を分析対象としていることについて説明する．

通常，一人当たり医療費を分析することが多いが，一人当たり医療費は医療費の 3 要素[14]（受診件数，受診日数，点数）に分解できることが知られている．しかも，一人当たり医療費は，3要素の中でも受診件数によって大きく影響を受けることが，これまでの研究により実証されている．また，保健予防の観点から見ても，患者の発症数を間接的に把握できる受診件数が最も重要な指標にあたる．さらに，既存研究において，国内外を含め受診件数と点数共に医療費を題材として SAE の補正を行った研究は筆者が調べた限りでは見当たらない．したがって，これら 3 つの理由から，受診件数を用いた医療費指数を補正することに政策的意義があるものと考え，本研究では期待件数を改良することとした．したがって，以下で用いる医療費指数は，受診件数を代替的な指標として用いた医療費指数を指す．

最後に高血圧疾患を分析対象に選択した理由について簡単に説明する．高血圧疾患のデータを用いた理由は，①財政上，また医療費適正化上も重要な疾患であること，②性別・年齢階層別に受診傾向，期待件数が大きく異なることの 2 つである．期待件数とは，被保険者数に 6 府県（山形県，静岡県，長野県，大阪府，広島県，高知県）の平均受診率[15]を乗じた値であり，性・年齢階層・疾患別に計

---

[13] 本研究は，平成 21 年度厚生労働科学研究「国，都道府県の医療費適正化計画の重点対象の発見に関する研究」の一部として行った研究であるため，平成 21 年度報告書に掲載予定である．

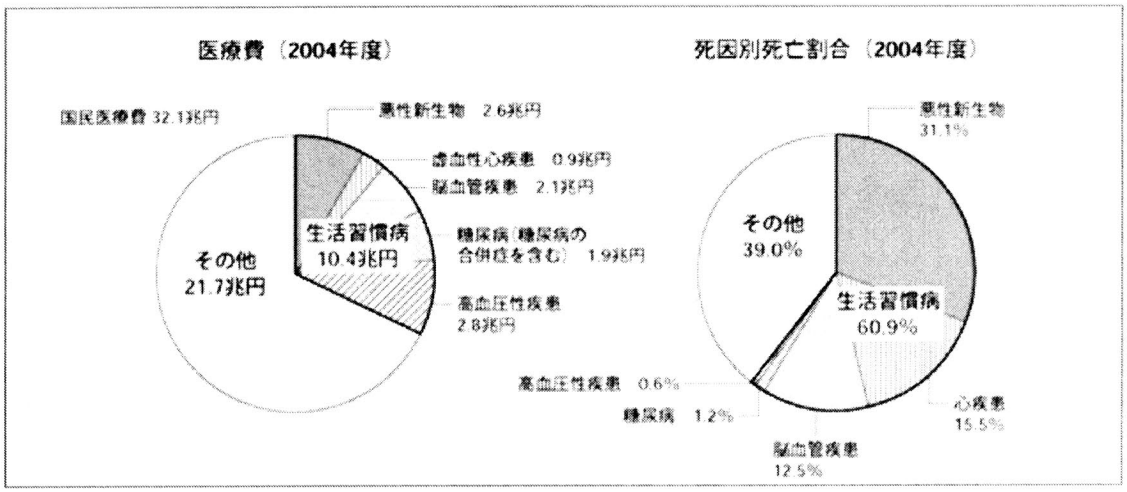
[14] 1 人当たり医療費は，①(医療機関への)受診率，②1 件当たり日数，③1 日当たり医療費の積で表わすことが出来る．つまり，1 人当たり医療費=1 日当たり医療費×受診率×1 件当たり日数となる．これら 3 つの要素を医療費の 3 要素と呼ぶ．

[15] 一人当たり受診件数の算出にあたっては，本研究のデータ出典である「国，都道府県の医療費適正化計画の重点対

算することができる.

高血圧疾患は，生活習慣病の代表疾患であるとともに，予防効果の高さから医療費適正化上重要な疾患として位置づけられている．実際に，医療費に占める割合を疾患別に確認すると，高血圧疾患は生活習慣病の医療費総額10．4兆円のうち総額で2．8兆円と最も大きな割合を占めていることがわかる(図 7).

さらに，死因別死亡割合においても，悪性新生物や心疾患あるいは脳血管疾患などの重症疾患の前段階の疾患として高血圧疾患は捉えられており，早期に予防に取り組むことで医療費を大幅に適正化できる可能性がある疾患として重要な疾患だといえる.



出典：平成19年度厚生労働白書

図 7 生活習慣病の医療費の割合

また，高血圧疾患の受診傾向と期待件数は，性別，年齢階層別に大きく異なっており，他の疾患に比べデータ区分(性・年齢階層区分)による違いが大きい．実際に補正を行う際には，どの程度データ区分を細かくするのかにするかによって補正の必要性が異なる．図 8，図 9 より，高血圧疾患の受診率は性別・年齢階層別に大きく異なっていることが確認できる．他の疾患においても差が生じていることが確認出来るが，高血圧疾患ほど大きく異なる疾患は見当たらない．図 10 は，医療費適正化上重要な疾患として位置づけられる 2 つの疾患(高血圧疾患と脳卒中(くも膜下出血，脳内出血，脳梗塞の3疾患の合計)の性・年齢階層別の受診率を比較したグラフであるが，脳卒中の受診率に高血圧ほどの性差あるいは年齢階層による差はないことが確認出来る.

さらに，高血圧疾患の医療費データを，高血圧(合計)，男性高血圧，女性高血圧，0-39 歳高血圧，40-49 歳高血圧，50-59 歳高血圧，60-69 歳高血圧，70 歳以上高血圧の 8 つのデータ区分別に，実績件数または期待件数の被保険者数との関係を示したグラフが図 11 である．また，それらの期待件数の要約統計量をデータ区分別に示したものが図 12 と表 3 である.

全体的な傾向として，被保険者数が少ないほど，実績件数と期待件数ともにデータがばらついていることがわかる．さらに，同一の被保険者数でみると，期待件数の方が実績件数を下回っている場合が

象の発見に関する研究」の研究協力自治体である 6 府県(山形県，長野県，静岡県，大阪府，広島県，高知県)のデータを用いている.

多いことがわかる．被保険者数が少ない方が受診件数が多くなるとは考えにくいため，被保険数が小さいほど件数のばらつきが大きいことは，SAE の問題，つまり誤差を生じていることを示唆する結果だといえる．

また，図 12 と表 3 からは，期待件数と期待件数の平均，最大値，最小値，標準偏差共に，高血圧(合計)のデータ区分が最も高い値を示していることがわかる．高血圧(合計)は，対象となる被保険者数が最も多い区分であるため，当然の結果だといえる．

また，高血圧についで高い値を示しているのが，70 歳以上高血圧，女性高血圧，男性高血圧のデータ区分である．その他の区分においては，60-69 歳高血圧，50-59 歳高血圧，40-49 歳高血圧，0-39 歳高血圧のデータ区分の順に，期待件数の平均，最大値，最小値，標準偏差の 4 つが共に小さい値を示している．これは，高血圧疾患が年齢階層別にみると高齢であるほど罹患しやすく，性別でみると女性であるほど罹患しやすい，あるいは受診頻度が高いという傾向を示しているといえる．このようなデータ区分による受診率の違いあるいは期待件数の違いを有するデータ区分では，佐伯ほか(2005)が期待件数の大きさによって補正の度合が異なっていることを示したように，補正の必要性が大きく異なる可能性が高い．

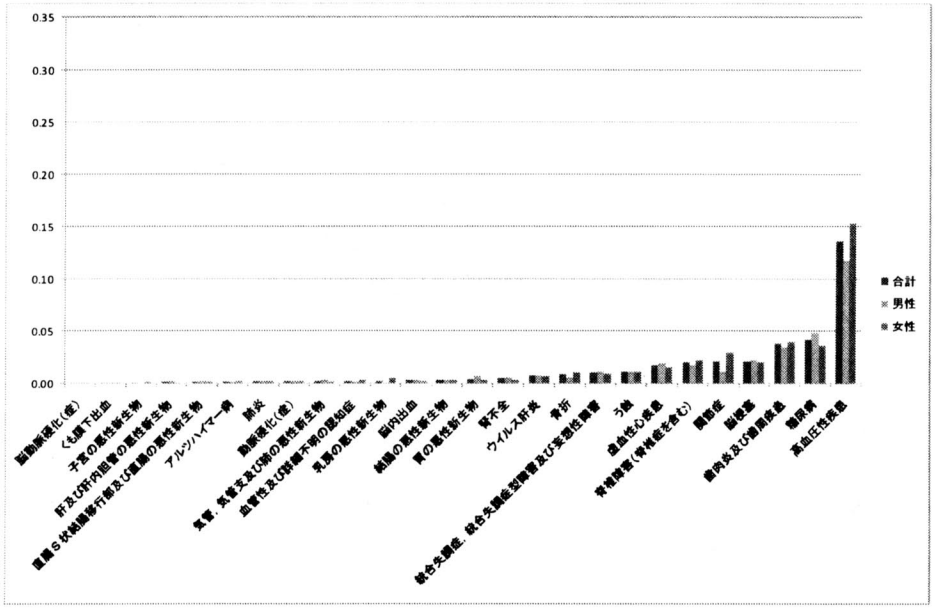


図 8 疾患別受診率(合計・性別)

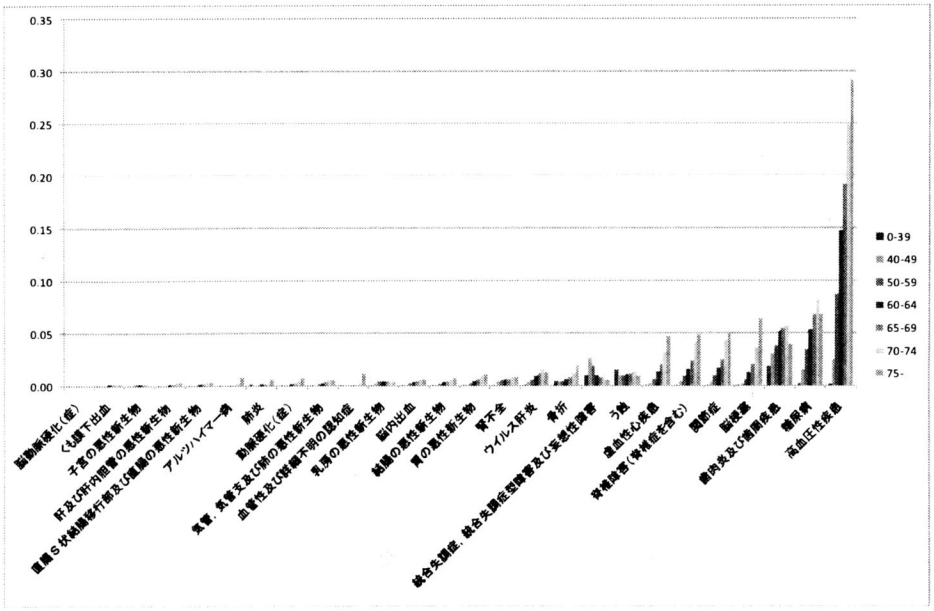


図 9 疾患別受診率(年齢階層別)

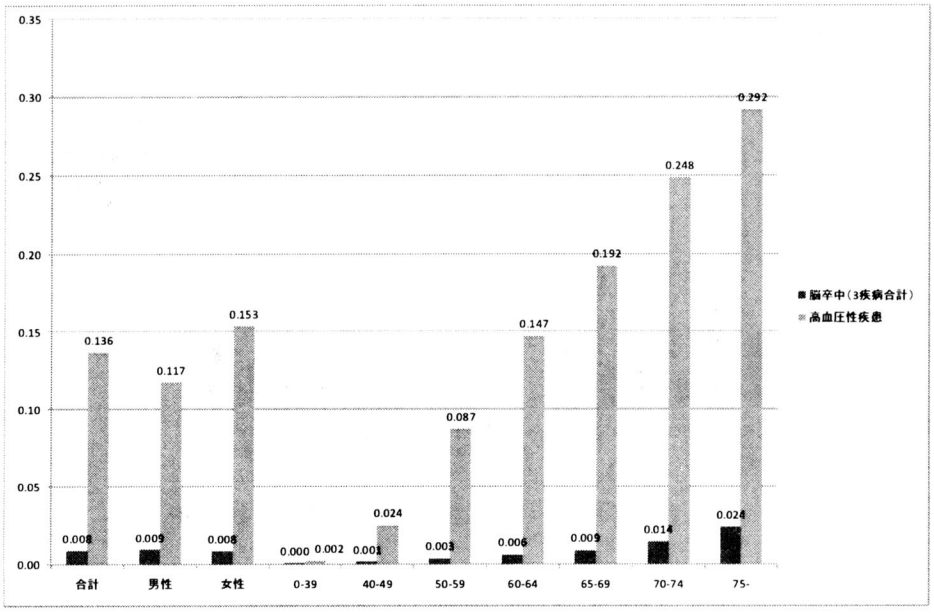


図 10 高血圧と脳卒中の性・年齢階層別受診率

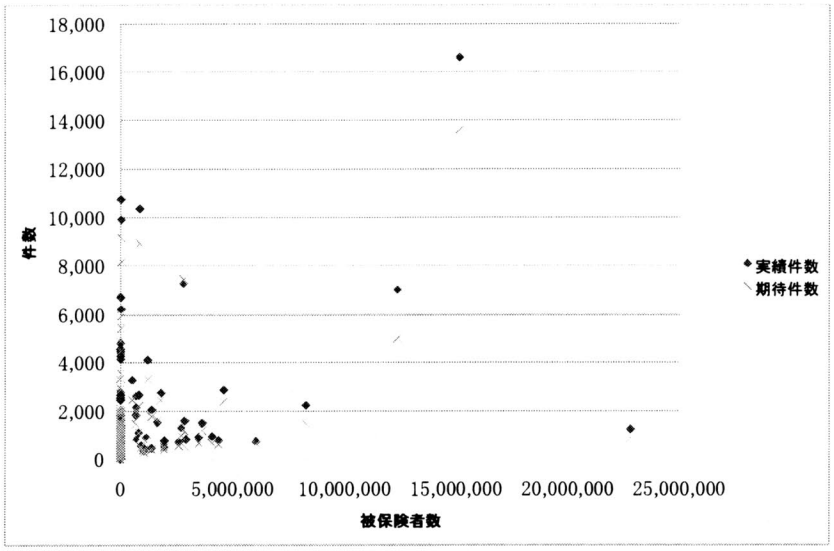


図 11 実績件数, 期待件数と被保険者数

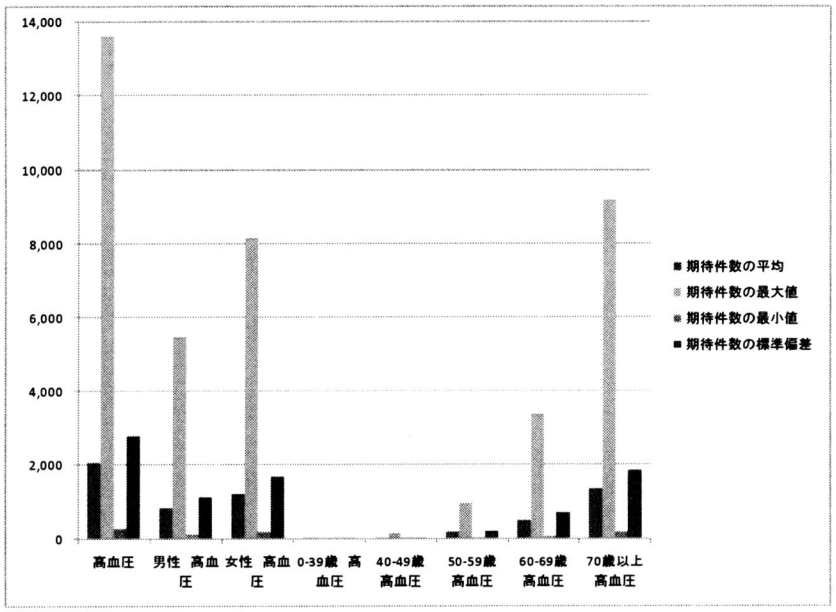


図 12 各データ区分の期待件数の要約統計量

表 3 各データ区分の要約統計量

	期待件数の平均	期待件数の最大値	期待件数の最小値	期待件数の標準偏差
高血圧	2,036	13,618	270	2,775
男性 高血圧	827	5,462	108	1,106
女性 高血圧	1,209	8,156	162	1,670
0-39歳 高血圧	5	36	1	7
40-49歳 高血圧	22	132	4	28
50-59歳 高血圧	160	921	23	200
60-69歳 高血圧	494	3,354	61	687
70歳以上 高血圧	1,354	9,175	174	1,855

以上より，本研究では，分析データとして，平成 18 年度 5 月診療分における山形県国民健康保険市町村別レセプト集計データの高血圧疾患に関する受診件数を用いる．

また，高血圧疾患を分析対象とする理由としては，財政上，または医療費適正化上重要な疾患であること，そして性別・年齢階層別に受診傾向，期待件数が大きく異なることがあげられる．

さらに，データ区分の違いによる補正結果を比較するため，8 つのデータ区分を用いて補正を行う．

補正に際しては，既存研究のレビューから妥当と考えた 4 つの手法(Locally Weighted Average, 経験ベイズ法，階層ベイズ法による 2 つのモデル(ポアソンガンマモデル，ポアソン対数正規モデル))を用いる．これらの分析を通じて，精度という面で最も良いと考えられる手法(階層いベイズ法)をデータ区分によらず常に適用すべきか，あるいはデータ区分によっては簡便な手法(LWA や経験ベイズ法)で代替可能となるのかを明らかにする．

4.2. 分析の流れ

分析の流れを図 13 に示した．はじめに，1. 補正手法の選択に関する分析として，医療費指数の改良を行う．改良にあたっては，補正結果の比較を行うことにより，データ区分別に適切な手法を選択する．つぎに，2. 適正化重点対象地域の選定に関する分析として，補正された医療費指数を用いて，政策的に重要な地域を選定する．これら 2 つの分析を通じて，本研究の目的である 1. 医療費指数の改良と 2. SAE の手法選択に関する知見の獲得を目指す．

1.補正手法の選択に関する分析では，既存研究レビューの結果選択した 4 つの SAE の手法(① Locally Weighted Average Method, ②経験ベイズ法(モーメント法を用いたポアソンガンマモデル)，階層ベイズ法(③ポアソンガンマモデル，④ポアソン対数正規モデル))を，次の 8 つのデータ区分に当てはめて，医療費指数の補正を行い比較分析した．8 つのデータ区分とは，高血圧(合計)，男性高血圧，女性高血圧，0-39 歳高血圧，40-49 歳高血圧，50-59 歳高血圧，60-69 歳高血圧，70 歳以上高血圧である．

なお，補正の比較分析の際には，補正の精度と簡便性のバランスの観点から考察した．学術的には補正の精度が高いことが重視されるが，政策的には行政官や保健師が計算できるよう簡便性も重視される．したがって，精度に大差が出ないのであれば，より簡便な方法が望ましい．したがって，4つの補正方法を 8 つのデータ区分に適用し，補正の精度を実証的に確かめたうえで，8 つのデータ区分ごとに有用な補正手法を検討した．

2.適正化重点対象地域の選定では，補正前と補正後の指数をマッピングすることにより，結果の違いを考察した．重要地域の選定にあたっては，標準偏差分類[16]という手法を用いて地域の選定を行う．標準偏差分類とは，GIS や空間統計の分野において地理的な情報を視覚化する際に用いられる手法の 1 つであり，平均値からの乖離度に基づいてクラス分けする方法である(村山，柴崎(2008))．この手法を用いることで，それぞれの県内で医療費が著しく高いあるいは低い地域を客観的に識別し，選定することができる．分類された結果より，重点的に施策を行うべき地域が，補正手法により異なるのか，あるいは異ならないのかを明らかにする．

[16] 詳しくは， 村山，柴崎(2008)あるいは Arc Map ユーザーガイドを参照されたい．

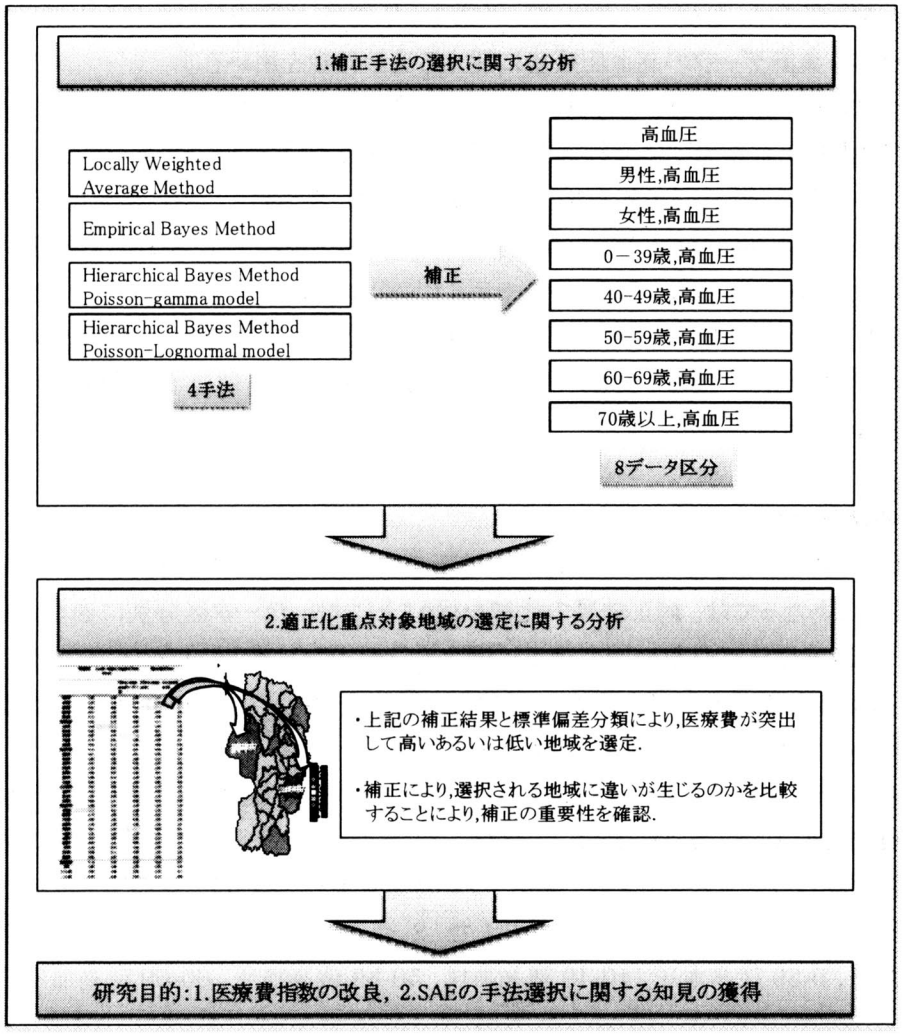


図 13 分析の流れ

4.3. 補正手法

以下では，既存研究のレビューから選択した 4 つの補正手法について説明を行う．Locally Weighted Average Method (LWA)は，非ベイジアンの手法として選択した手法であり，複雑なモデルやコンピュータ計算を用いずに補正を行うことができる．また，経験ベイズ法と階層ベイズ法は，これまで提案されてきた SAE の手法よりも優れた手法として提案されるベイジアンによる手法である．一方の経験ベイズ法は，ベイジアンの手法の中でもより簡便な手法として考えられている．階層ベイズ法は，ベイジアンの手法の中でも最も新しく，より精度が高い手法として提案されている．また，階層ベイズ法については，多様なモデルが提案されているため，その中でもより単純なモデルとしてポアソンガンマモデルとポアソン対数正規モデルの 2 つを用いる．ポアソンガンマモデルよりもポアソン対数正規モデルの方が，空間的な相関を扱えるためより精度の高いモデルだとされている．

4.3.1. Locally Weighted Average (LWA) Method

Locally Weighted Average Method(以下，LWA)とは，地理的な領域にもとづいて加重平均をとり，その値を推定値とする方法である．この手法は，近接する領域であれば，類似した傾向を示すという前

提に基づいており，近接地域の加重平均を算出することにより，周囲の値に推定値を近づける手法である．LWA について，解説を行ったものとして Waller and Gotway(2004)があげられる．Waller and Gotway(2004)は，LWA を近隣地域から情報を借りることで良い推定値(精度が安定しており，ノイズの少ない推定値)を生成する手法だと述べた．また，回帰分析における Scatterplot Smoothers の手法や，時系列分析における移動平均に関する手法に類似しているとも指摘した．LWA の適用を行った研究としては，Duel et al. (2007)[17]があげられる．

図 14 は，LWA が加重平均を算出する際に，重みを付与する範囲を円で表わしたイメージ図である．補正対象によって円が移動しており，重み付けする範囲が円盤の形であることから Disk smoothing という名前でも呼ばれる手法である(Waller and Gotway(2004))．

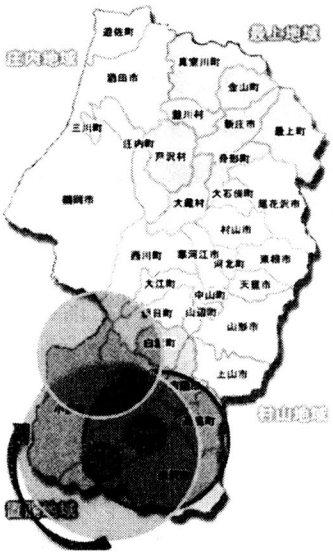


図 14 Locally Weighted Averages Method による重み付け

LWA は，加重平均を算出することにより，補正を行う手法であるため，近接地域の設定方法など加重の加え方が推定上の重要な要素となる．加重の仕方にはいくつかの方法が存在しており，Waller and Gotway(2004)によれば，①距離による重み付け，②人口規模による重み付け，③境界線の有無による重み付け，④近隣地域というグループ属性による重み付けなどが可能となっている．これらの加重の加え方は補正対象であるデータの特性に応じて選択する必要があり，複数の要素を用いて同時に重みづけすることも可能である．

医療費指数の補正を行う場合には，同一の二次医療圏でかつ隣接地域の中では医療供給体制や患者の諸条件がある程度同質だと仮定することが出来るため，本研究では同一二次医療圏という要素に加えて隣接地域として境界線を接するという要素で重み付けを行うこととした．しかし，近接地域の設定方法は推定結果を大きく変化させる可能性があるため，より妥当な設定方法は如何なる条件かについては別途研究が必要であろう．特に，二次医療圏の設定方法についてはいくつかの問題点が指摘されているため，真に同質な範囲とは如何なるものかについて明らかにする必要がある．さらに，実績

[17] Duel et al. (2007)は，ポリゴンごとにスムージングを行う Locally Weighted Average Method を新たに提案している．

件数の補正の際には，人口規模の影響を除去していないため，人口規模を補正の要素に入れ込むこととした．

したがって，本研究では①同一の二次医療圏内にあるか否か，②境界線を接するか否か，③人口規模という 3 つの要素を加味して補正することとした．これらの条件を用いて LWA を定義すると，次式のように示すことができる．

$$\hat{\theta}_i = \frac{\sum_{j=1}^{N} W_{ij}{}^{*} r_j \, I[j \in N_i] + r_i}{\sum_{j=1}^{N} W_{ij}{}^{*} p_j \, I[j \in N_i] + p_i}$$

$$W_{ij} = \begin{cases} 1 & \text{領域 i と j が境界を共有するとき} \\ 0 & \text{それ以外} \end{cases}$$

$$I[j \in N_i] = \begin{cases} 1 & \text{領域 i と j が同一の二次医療圏内} \\ 0 & \text{それ以外} \end{cases}$$

r: 単純医療費指数(未補正の医療費指数)

p_i: 人口 (4.1)

ここで，y_iは観測された医療費指数，$\hat{\theta}_i$は LWA による補正後の指数を示している．また，N_iは地域 j が地域 i と同一の二次医療圏であることを指す．

4.3.2. Empirical Bayes(EB) Method

Empirical Bayes(EB) Method(以下，経験ベイズ法)とは，ベイズ統計の考え方を SAE の手法に応用したものであり，事前分布に関するパラメータをデータから計算するという特徴を有する．このようなベイジアンによる手法は，1980 年代以降発達した手法であり(Lawson et al. (1999))，SAE の新たなアプローチとして注目され，より精度の優れた手法として広く用いられるようになってきている．このようなベイジアンによる推定では，通常 SAE の手法に限らず，データに事前分布を想定し，尤度と事前分布をかけあわせることにより，事後分布を得る．この事後分布が推定結果となる．

ベイジアンの手法では，パラメータが分布を持つという仮定をモデルに組み込むことができるという点が特徴的である．この分布とは，分析者が事前に有している分析対象のデータに関する信念あるいは考えを指している．つまり，本研究でいえばθ_iというリスクのパラメータ，つまり医療費指数がどのような分布に従うのかという考えを反映したものである．

経験ベイズ法は，事前分布の形状を決定するパラメータに対して，更なる事前分布(パラメータ)を仮定し，推定を行う手法である．なお，更に設けた事前分布のパラメータをハイパーパラメータと呼ぶ．また，ハイパーパラメータを設定するという点でいえば，経験ベイズ法は Hierarchical Bayes Method[18](以下，階層ベイズ法)の一部だといえる．しかし，たいていの場合には事前分布のパラメータは未知であるため，今手元に有しているデータを用いてパラメータの値を設定し，推定値を算出しようとするのが経験ベイズ法である．つまり，経験ベイズ法による推定では，データに基づいて事前分布を

[18] 階層ベイズ法は，事前分布のパラメータに更なる事前分布(ハイパーパラメータ)を仮定し，モデルを階層構造として推定する手法である．

設定しているといえ，その意味でより頻度論に近い手法だといえる(Rao(2003)).

また，後に詳述するが，これらのパラメータを数値計算(MCMC 法(マルコフ連鎖モンテカルロ法))により与えようとする手法が階層ベイズ法である.

ここで，m 個の地域に対する医療費指数(θ_i)を補正するため，ポアソンガンマモデルを仮定するとき，医療費指数は以下のように表わすことができる.

$$\hat{\theta}_i^{eb} = (O_i + \hat{\alpha})/(e_i + \hat{\beta}) = (1 - \hat{B}_i)\hat{\theta}_i + \hat{B}_i\hat{\mu} \tag{4.2}$$

ここで，$B_i = \hat{\beta}/(e_i + \hat{\beta})$, $\hat{\mu} = \hat{\alpha}/\hat{\beta}$, $\varphi^2 = \beta/\alpha^2$であり，$i=1, \ldots, m$を指す.

$$\begin{aligned} &\theta_i \sim Gamma(\beta/\alpha, \beta/\alpha^2) \\ &y_i \sim Poisson(\theta_i e_i) \end{aligned} \tag{4.3}$$

θ_iは平均β/α，分散，β/α^2を持つガンマ分布に従うと仮定し，観測された受診件数(y_i)は平均と分散共に$\theta_i e_i$であるポアソン分布に従うと仮定している.

上式において，ハイパーパラメータとはαとβを指し，経験ベイズ法はこれらの値をデータから計算しようとしている. また，Devine et. al(1994)は，ベイズ推定値を観測値であるθ_iと事前分布の平均である$\hat{\mu}$のそれぞれについて重みづけを行い，足し合わせたものだと説明した. また，V. Gomez-Rubio and A. Lopez-Quilez(2006)は経験ベイズ法を Direct estimation と Relative risk(相対リスク)の事前平均との妥協点だと説明している. 推定値$\hat{\theta}_i^{eb}$は，結果として人口規模が小さいとき$\hat{\mu}$に近づき，反対に人口規模が大きいときには，$\hat{\theta}_i$に近づくこととなる. つまり，人口規模が大きいときには，観測された指数が安定した精度を持つと考えるため補正を行わず，観測値$\hat{\theta}_i$に収束させるということを意味している.

さらに，経験ベイズ法におけるハイパーパラメータの算出方法については，複数の方法が提案されている. 本研究では，より簡便な方法として Marshall(1991)によって提案されたモーメント法[19]による算出方法を用いる. 実際に，自治体の政策担当者が計算を行う場合には，ベイジアンの手法を用いることによって推定精度を上げつつも，より簡便な方法であることが望ましいと考えたためである. その他の手法(最尤法等)では収束計算を行う必要があるため，単純な四則演算のみで計算を行うことが難しい[20]ため，より単純なモーメント法を選択した.

モーメント法によるハイパーパラメータの値は，次式から得ることができる.

$$\begin{aligned} \hat{\mu} &= \frac{\sum_i O_i}{\sum e_i} \\ \varphi^2 &= S^2 - \frac{\hat{\mu}}{\frac{1}{m}\sum_{1=1}^{m} e_i} \end{aligned} \tag{4.4}$$

---

19 この点について，国内の文献として中谷ほか(2004)や佐伯ほか(2005)が説明を加えているので参照されたい.

20 ただし，経験ベイズ法については，R コマンドとしていくつかの関数が既に定義されている.

ここで，$S^2 = \frac{\sum_{i=1}^{m} e_i(\hat{\theta}_i - \hat{\mu})^2}{\sum_{i=1}^{m} e_i}$である．また，$\varphi^2$が負値となる場合には，ゼロで切り捨てを行う．

また，$\hat{\mu}$の値の設定の仕方には 2 つの方法があり，Global mean(全体の平均値)に収束させる場合と Local mean(限定された地域の平均値)に収束させる場合とが考えられる．医療費指数の場合には，地理的な条件やその他の状況により二次医療圏が設定されているため，二次医療圏内の値に収束させるこが Global mean を用いることよりも妥当と考え，Local mean を用いることとする．同一の二次医療圏内においては，医療供給体制や社会的状況，あるいは患者の生活習慣等がある程度似ているであろうという前提に基づいている．つまり，医療費の発生に関わる諸要因がある程度同じだと仮定している．経験ベイズ法を用いて合計特殊出生率や標準化死亡比のスムージングを行った佐伯ほか(1999)，平子ほか(1999)においても，二次医療圏の平均値を Local mean と設定し補正を行っている．Global mean として県全体の平均値に収束させるよりも，二次医療圏等の近接地域あるいは諸条件が同じと考えられる Local mean への収束の方が，地理的状況を加味しているため，より良い推定値だと考えているためである．ただし，Local mean をどの範囲に設定すべきかについては，LWA の近隣地域の設定方法と同様，議論の余地が残る．そもそも医療圏の設定自体が恣意的に行った設定であり，二次医療圏の実態についても人口や面積にばらつきが生じていることや日常生活と住民の受診行動が一致しないなどの問題点が指摘されているため，詳細な分析が必要であろう．つまり，経験ベイズ法として Global mean ではなく Local mean に収束させたほうがより地理的な条件を加味できる(Martuzzi and Elliott(1996))ということに異論はないが，Local mean をどの範囲に設定すべきかについては別途詳細な研究が必要であろう．

4.3.3. Hierarchical Bayes(HB) Method

Hierarchical Bayes(HB) Method(階層ベイズ法)は，SAE の手法の中でも最も新しい手法だといえる(Trevisani and Torelli(2004))．また，事後分布を得る際に異なるタイプの効果(例えば，固定効果(Fixed Effect)や混合効果(Random Effect)，あるいは変化の源となる変数等をモデルに組み込むことができ，モデルパラメータの不確実性を統一的に扱うことができる手法である(V. Gomez-Rubio and A. Lopez-Quilez(2006))．

階層ベイズ法による推定値は以下のように表わされる．

$$\theta_i{}^b = (O_i + \alpha)/(e_i + \beta) = (1 - B_i)\hat{\theta}_i + B_i\mu \tag{4.5}$$

推定式は，経験ベイズ法によるものと同様，医療費指数θ_iと平均μの加重平均の形をとる．

階層ベイズ法による SAE の補正では，単純なモデルから，非常に複雑なモデルまで数多くのモデルが存在している[21]．前述したように固定効果や混合効果を同時に入れるタイプのモデルや空間的な自己相関を加味できる複雑なモデルまで非常に幅広いが，本研究では，政策判断材料として用いることを念頭に置き，複雑過ぎないモデルとして，以下にあげるポアソンガンマモデル(HBP)とポアソン対数正規モデル(HBL)を用いることとする．

ポアソンガンマモデルは次式のように示される．

[21]詳しくは，Rao(2003)，Lawson et al. (2003)を参照されたい．

$$\theta_i \sim Gamma(\beta/\alpha,\ \beta/\alpha^2)$$
$$y_i \sim Poisson(\theta_i e_i) \tag{4.6}$$

また，ポアソン対数正規モデルは，以下のように示される．

$$y_i \sim Poisson(\theta_i e_i)$$
$$Log\theta_i = \alpha + v_i$$
$$v_i \sim N(0,\ \tau_v{}^2) \tag{4.7}$$

両者のモデルの違いは，空間的な相関を扱えるか否かという点である．Lawson et al. (2003)によれば，空間的な相関を扱えないことがポアソンガンマモデルの欠点だと指摘されている．つまり，指数の分布を補正する際に，ポアソンガンマモデルを仮定した階層ベイズ法では，医療費指数などのリスクが近隣地域で同じような傾向を持つという，空間的な相関をモデルに組み込むことが出来ない．それに対し，ポアソン対数正規モデルでは指数に関わる空間的な相関関係をモデルに組む込むことができ，より柔軟な推定を行うことができる．2 つのモデルを医療費指数の補正に用いることで，医療費指数の補正にあたってどちらが適切かという点について，示唆を与えることができるといえよう．

なお，階層ベイズ法による補正に際しては，すべてMCMC 法(ギブズサンプラー)を用いて推定を行い，各々のシミュレーション結果についてはマルコフ連鎖が収束しているか否か(定常状態に達したか)を確認するため，Gelman-Rubin 統計量により収束診断[22]を行った．

[22] 詳しくは，Lawson et al. (2003)，古谷(2008)を参照されたい．

5. 分析結果と考察

以下では，補正手法の選択に関する分析，適正化重点対象地域の分析の順に，分析と考察を行う．補正手法の選択に関する分析では，高血圧疾患に関する 8 つのデータ区分[23](性別，年齢によってデータを分割した)に対して，既存研究レビューの結果選択した 4 つの補正手法(LWA，経験ベイズ法(EB)，階層ベイズ法の 2 つのモデル(ポアソンガンマモデル(HBP)とポアソン対数正規モデル(HBL))を適用し，補正を行う．

補正にあたっては，データ区分によって適切な手法は異なるのか，あるいは全てのデータ区分において適切な手法は同じなのかを明らかにするため，補正結果の比較を行う．その際，データ区分によって手法選択で重視すべき点(精度あるいは簡便性など)を明らかにする．医療費適正化という政策目的を念頭に置いて医療費指数の補正を行うため，実際に SAE の手法を用いる主体として想定する国あるいは都道府県等の政策担当者が使うためには，ある程度の推定精度を改善しつつも，複雑過ぎない手法であることが求められるためである．

適正化重点対象地域の分析では，補正された指数を用いて，医療費適正化上重点的に施策を行うべき地域を選定する．選定に際しては，平均値からの乖離度によって突出して高いあるいは低い地域を選定するため，標準偏差分類という手法を用いて地域を選定する．補正前と補正後の指数による地図を視覚的に確認することにより，選定される地域に違いが出るのかを明らかにする．最後に，これらの分析を通じて，政策的に補正を行うことが有用であるか否かを考察する．

5.1. 補正手法の選択に関する分析

5.1.1. 分析結果

5.1.1.1. 4つの補正手法の比較分析

4 つの手法を用いた補正結果の一覧を，本節の最後にある表 7 から表 14 に示した．また，手法別の補正結果の全体の傾向を把握し，比較するため，図 15，図 16 に各手法による補正結果の最大値，第 3 四分位点，第 1 四分位点，最小値を箱ひげ図の形で表した．図では単純指数，LWA(Locally Weighted Average Method)，Empirical Bayes Method(経験ベイズ法)，Hierarhical Bayes Method (Poisson-gamma model)(階層ベイズ法)，Hierarchical Bayes(Log-normal model) (階層ベイズ法)の順に示している．

はじめに，データ区分別に箱ひげ図に示した値と全体の傾向を図 15，図 16 より確認すると，高血圧では，手法による差異はあまり生じていない．最大値，第3四分位点，第1四分位点は大きな変化を生じていない一方で，LWA による最大値のみ小さくなっていることがわかる．

男性高血圧では，階層ベイズ法の 2 つの手法による結果と単純指数の結果が類似した傾向にあることがわかる．一方で，LWA と経験ベイズ法による結果の最大値が大きく縮みこんでいることが分かる．最大値，第 3 四分位点，第 1 四分位点は大きな変化を生じていない．

女性高血圧では，男性高血圧とほぼ同様の傾向にあり，LWA と経験ベイズ法による結果が類似しており，最大値のみ大きく縮みこむ結果となった．

0-39 歳高血圧では，LWA から階層ベイズ法へと精度が高い補正方法になるほど補正後の指数の変動が縮小する傾向にあった．このような顕著な推移は，上記の高血圧，男性高血圧，女性高血圧等

[23]高血圧(合計)，男性高血圧，女性高血圧，0-39 歳高血圧，40-49 歳高血圧，50-59 歳高血圧，60-69 歳高血圧，70 歳以上高血圧の 8 つの区分である．

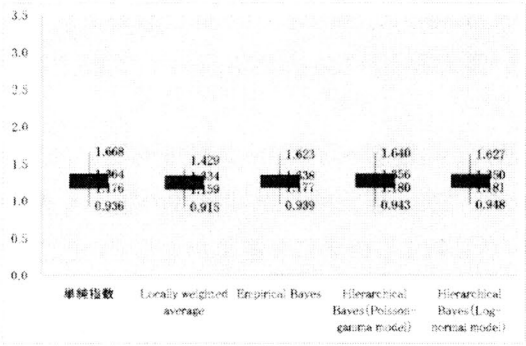
のデータ区分にはみられず，特徴的な結果だといえる．さらに，補正前の単純指数の値も大きく広がっていることも読み取れる．また，他の区分では著しい差のみられなかった階層ベイズ法と経験ベイズ法の補正結果に差が生じていることがわかる．

40-49歳高血圧では，0-39歳高血圧と同様，LWAから階層ベイズ法へと精度が高い補正方法になるほど補正後の指数の変動が縮小する傾向にあった．また，補正前の単純指数の値が大きく広がっていることがわかる．さらに，経験ベイズ法と階層ベイズ法による結果にも差異が生じていることがわかった．

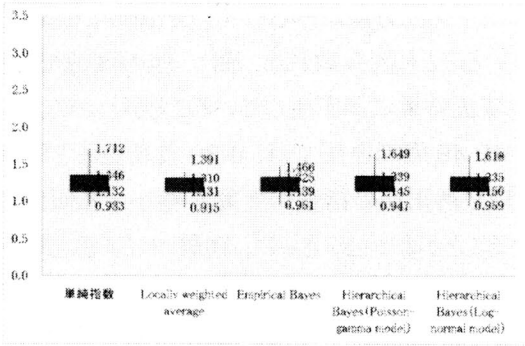
50-59歳高血圧では，LWAの補正結果は，最大値が単純指数よりも大きくなるという結果になった．また，単純指数と階層ベイズ法の結果に著しい差はみられない．一方，経験ベイズ法による第 3 四分位点と第 1 四分位点は他の手法と比べ，その幅が小さくなっていることがわかる．また，最小値に大きな差は見られなったものの，最大値は手法ごとに差を生じる結果となった．

60-69歳高血圧では，経験ベイズの補正結果が他の手法と比較して，大きく縮みこむ結果となった．経験ベイズ以外の手法では，単純指数の全体的な傾向と比較して著しい差は見られなかった．

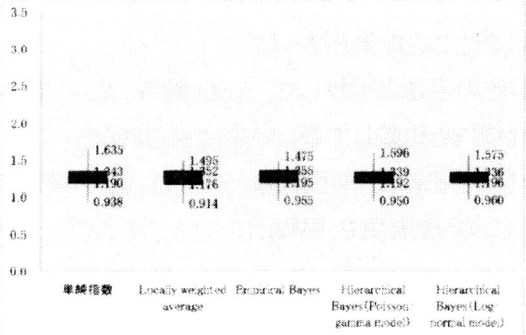
70 歳以上高血圧では，LWA による最大値が他の手法に比べ小さくなっていた．最小値に関しては大きな変化はみられなかった．また，経験ベイズ法による最大値も LWA についで小さくなっていることがわかる．



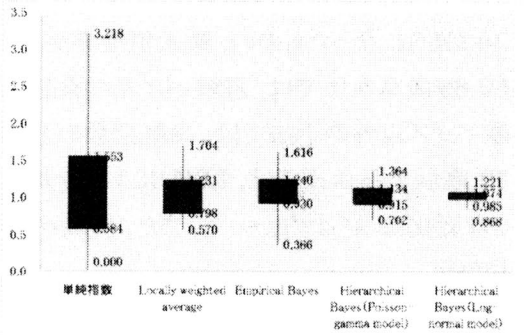
高血圧



男性 高血圧

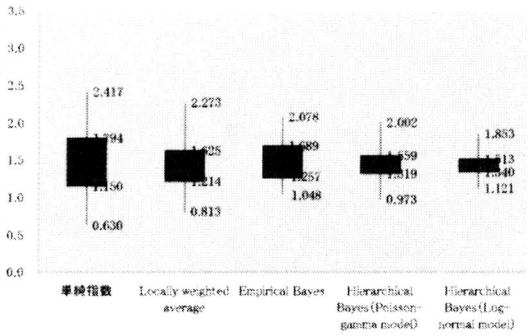


女性 高血圧

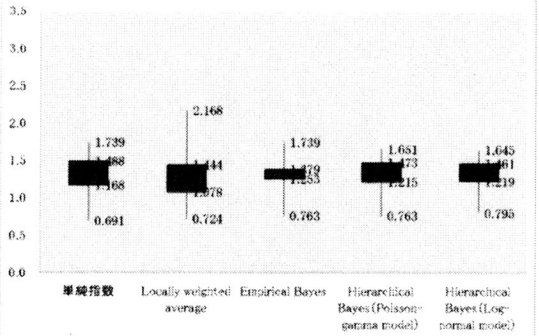


0-39歳 高血圧

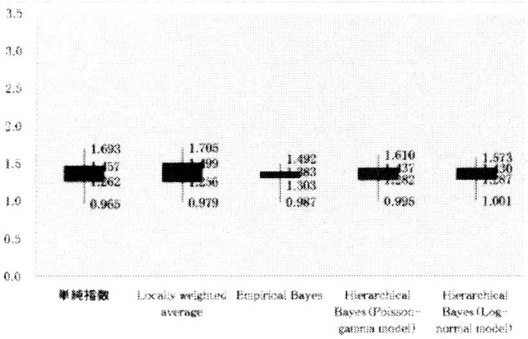
図 15 データ区分の違いによる各手法の指数の補正結果



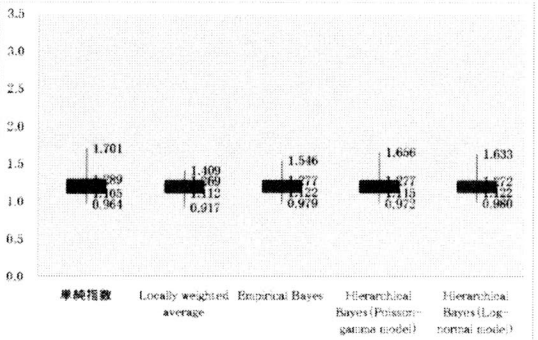
40-49歳 高血圧



50-59歳 高血圧



60-69歳 高血圧



70歳以上 高血圧

図 16 データ区分の違いによる各手法の指数の補正結果

5.1.1.2. 補正差の傾向

全てのデータ区分を統合し，補正差(各手法による補正結果の単純指数との差)と期待件数を同時に示したものが図 17 である.

図 17 から，全体の傾向を観察すると，4 つの手法すべてにおいて，期待件数が小さいほど補正差が大きくなっていることがわかる. 補正差は 0.5 から－0.5 の範囲に集中しており，補正差が絶対値で 0.5 以上の指数は期待件数が 500 以下となる傾向にあった. また，期待件数が 2,000 以上の指数では，補正差がほぼゼロに近いところに位置しており，あまり補正がなされていないことがわかる. さらに，期待件数が非常に小さい指数では，補正差の絶対値が最大で約 2.7 と期待件数が大きい指数の補正差がゼロであることと対象的な結果となっている. これらの結果から，期待件数が小さい場合ほど，SAE の問題を生じる誤差が大きく生じていることが示唆された. さらに，0.5 以上の補正差を生じた指数は，4 つの手法すべてにおいて存在していることがわかった.

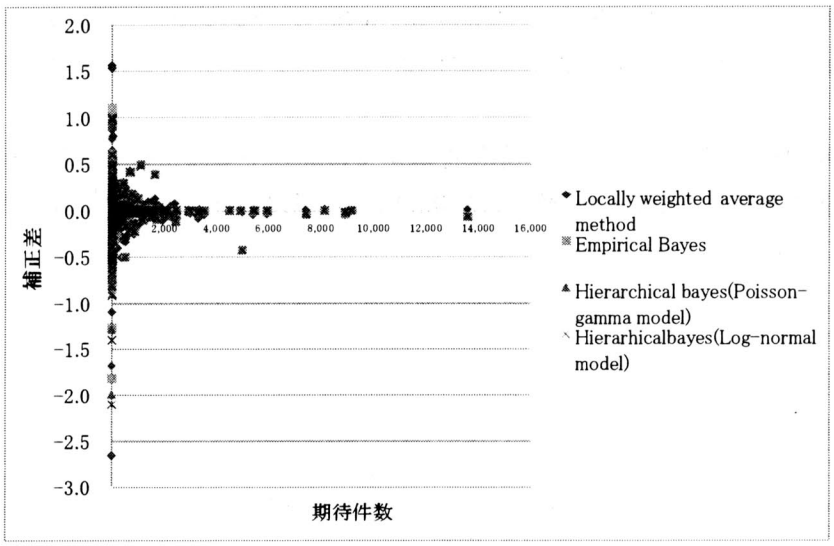


図 17 すべてのデータ区分における期待件数と補正差

5.1.1.3. データ区分別にみた補正差の傾向

つぎに，図 18と図 19より，データ区分別に補正差と期待件数の関係を確認した. 全体の傾向と同様，すべてのデータ区分において期待件数が小さいほど，補正差が大きくなっていることがわかる. 高血圧，0－39 歳高血圧，40－49 歳高血圧，50－59 歳高血圧の 4 つのデータ区分では，補正差のばらつきが大きく生じると同時に，0.5 以上の補正差を伴った指数が複数存在している.

反対に，男性高血圧，女性高血圧，60－69 歳高血圧，70 歳以上高血圧の 4 つのデータ区分では，0.5 以上の補正差を 1 つも生じていない.

個別に補正差と期待件数の関係をみてみると，高血圧のデータ区分では，期待件数が約 3,000 以下の範囲で補正差のばらつきが生じており，補正差のほとんどが 0.5 から-0.5 の範囲に収まっている.

男性高血圧のデータ区分では，期待件数が 1,000 以下の範囲で補正差がゼロから多少乖離している指数が確認できるが，すべての指数で補正差は絶対値で 0.5 以下となっており，あまり大きな補正は全体としてなされていないことがわかる.

女性高血圧のデータ区分では，男性高血圧のデータ区分と同様に期待件数が約 1,000 以下の範

囲で補正差のばらつきが生じている．ただし，補正差はすべて絶対値で 0.5 以下の範囲に収まっており，あまり補正がなされていないものといえる．

0－39 歳高血圧のデータ区分では，補正差の大きい指数が目立っており，補正差が 1.0 以上を示す指数が複数存在している．0－39 歳高血圧以外に，1.0 以上の補正差を複数生じたデータ区分は他になく，その他の区分では殆どの指数が絶対値で 0.5 以下の補正差に留まっている．0－39 歳高血圧のデータ区分では，殆どの指数が絶対値で 0.5 以上の補正差を生じさせている．

40－49 歳高血圧のデータ区分では，期待件数が 50 以下の範囲で補正差にばらつきが生じている．また，補正差が絶対値で 0.5 以上の指数も複数存在しており，0－39 歳高血圧についで補正が大きくなされていることがわかる．

50－59 歳高血圧のデータ区分では，期待件数が 400 以下で補正差にばらつきが生じている．しかし，ほとんどの補正差は絶対値で 0.5 以下の範囲に収まっている．ただし，補正差が絶対値で 0.5 以上の指数もいくつか存在している．

60－69 歳高血圧のデータ区分では，期待件数が 1,000 以下の範囲で，補正差にばらつきが生じている．補正差のすべては絶対値で 0.5 以下の範囲に収まっており，あまり大きな補正はなされていない．

70 歳以上高血圧のデータ区分では，期待件数が 2,000 以下で補正差にばらつきが生じている．ただし，ほぼすべての指数は補正差が絶対値で 0.5 以下の範囲に収まっており，ほぼゼロに近い範囲に収まっていることもわかる．

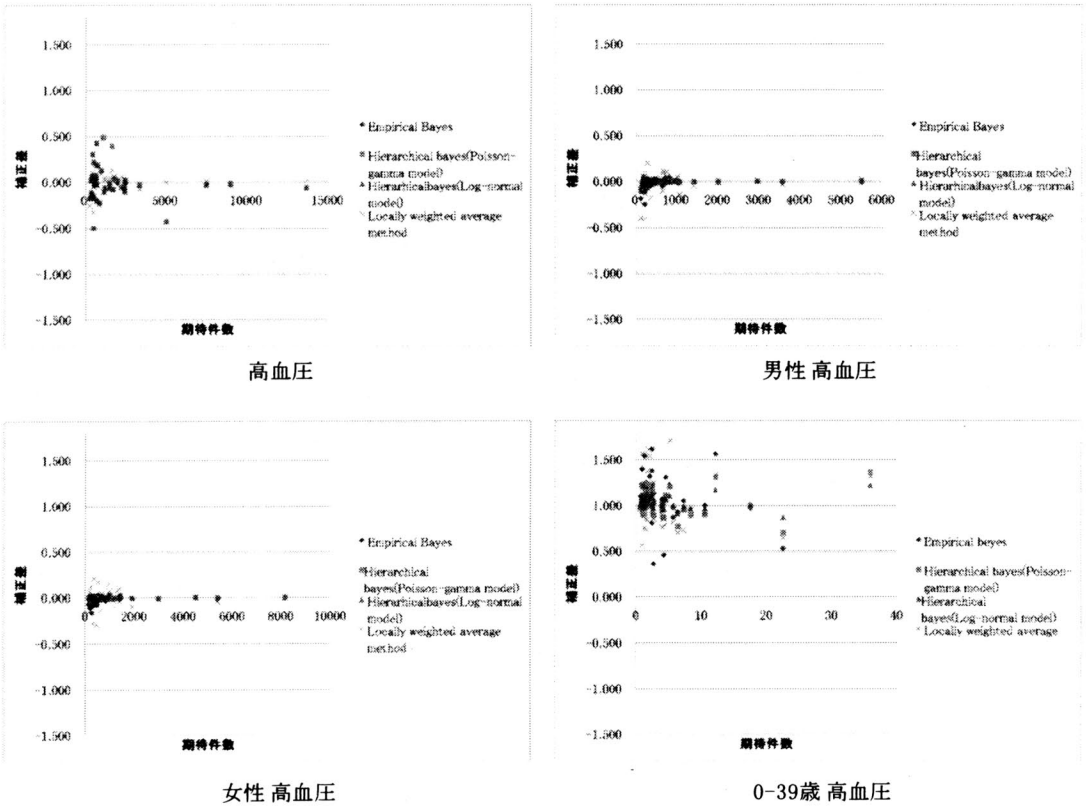
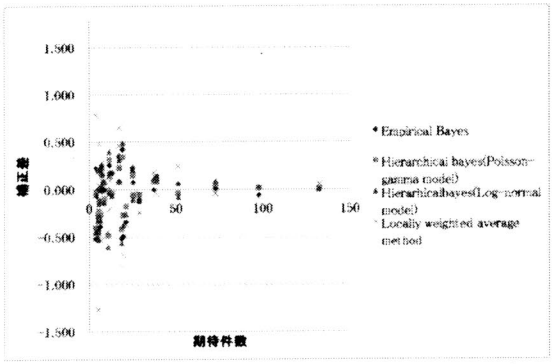
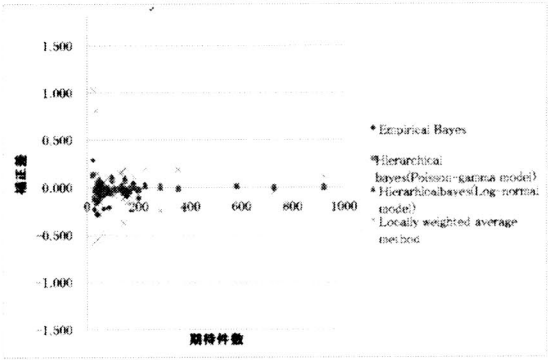


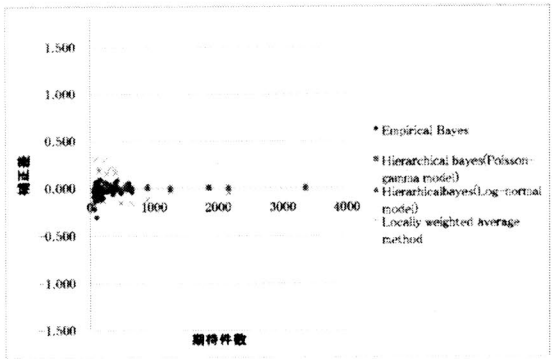
図 18 データ区分別の期待件数と補正差



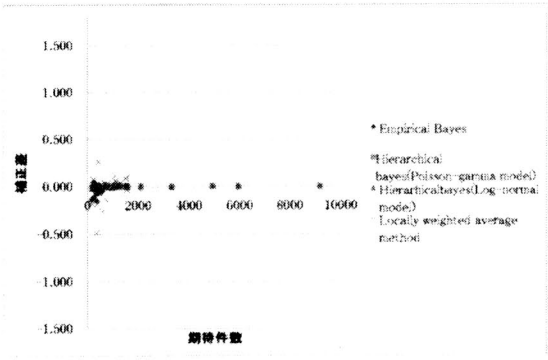
40-49歳 高血圧



50-59歳 高血圧



60-69歳 高血圧



70歳以上 高血圧

図 19 データ区分別の期待件数と補正差

図 20 は，データ区分別の補正差の標準偏差を示している．高血圧，0－39 歳高血圧，40－49 歳高血圧，50－59 歳高血圧の 4 つのデータ区分では，補正差の標準偏差がその他のデータ区分に比べ非常に大きな値となっている．特に 0－39 歳高血圧，40－49 歳高血圧，50－59 歳高血圧の 3 つのデータ区分の標準偏差は非常に大きな値を示しており，補正により指数の結果に差が生じているものといえる．

つぎに，補正差の標準偏差を手法別に比較すると，同じデータ区分においても異なった結果を示していることがわかる．たとえば，全体として大きな標準偏差を示している 0－39 歳高血圧，40－49 歳高血圧，60－69 歳高血圧においても，手法によってその値に差が生じている．特に，LWA とベイジアンによる 3 つの手法間では大きな差が生じている．LWA の標準偏差は，高血圧を除いてすべてのデータ区分で他の 3 つの手法よりも高い値を示しており，その差が大きい．データ区分によって，各種手法による補正の度合が変化するものといえる．

また，多くのデータ区分において，経験ベイズ法，階層ベイズ法(ポアソンガンマモデル)，階層ベイズ法(ポアソン対数正規モデル)による 3 つの結果は，比較的類似した傾向を示しているが，0－39 歳高血圧，40－49 歳高血圧のデータ区分では，経験ベイズ法と階層ベイズ法の間に差が生じていることがわかる．

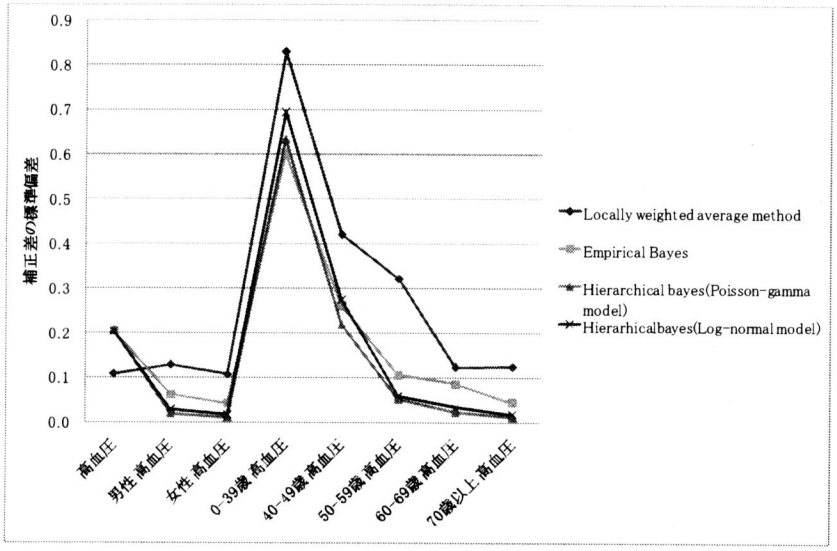


図 20 データ区分別の補正差の標準偏差

続いて，図 21 は，各手法による補正結果が補正前と比較してどの程度異なるのか(単純指数との乖離度)を示したグラフである．また，手法間の補正結果の関係をみるため，補正された指数を 2 つの手法ごとに散布図の形で比較したものが，図 23 である．

ここで，単純指数との乖離度，また手法間の補正結果の乖離度を示すために用いる指標は，平均二乗誤差とよばれるものであり，次式のように示される．

$$\text{平均二乗誤差}=\frac{1}{m}\sum_{i=1}^{m}(\hat{\theta}_i-\theta_i)^2 \tag{5.1}$$

ここで，m は地域の数を示し，$\hat{\theta}_i$は補正後の指数，θ_iは補正前の単純指数である．元のデータから乖離しすぎることなく補正を行うという考えのもとでは，この平均二乗誤差の値はより小さい方が良いということになる．

① 単純指数との比較

データ区分別に平均二乗誤差の値を比較すると，0－39 歳高血圧のすべての手法において最も高い値となっている．ついで，40－49 歳高血圧，50－59 歳高血圧，高血圧のデータ区分で高い値を示す結果となった．男性高血圧，女性高血圧，50－59 歳高血圧(LWA を除く)，60－69 歳高血圧，70 歳以上高血圧においては，ほぼゼロに近い値となっており，全体としてはあまり補正がなされていないといえる．つまり，全体として補正が大きくなされたデータ区分は，0－39 歳高血圧，40－49 歳高血圧，50－59 歳高血圧であった．

さらに，手法別に平均二乗誤差の値を比較すると，ほぼすべてのデータ区分において高い値を示したLWAの平均二乗誤差は，40－49歳高血圧と50－59歳高血圧において著しく高い値を示している．また，0－39歳高血圧，40－49歳高血圧では，ベイジアンによる3つ手法の平均二乗誤差も高い値を示しているのに対し，50－59 歳高血圧ではほぼゼロに近く，LWA のみ高い値を示している．また，LWA による平均二乗誤差が高血圧を除くすべてのデータ区分において，大きいことが分かる．唯一，

LWA の平均二乗誤差が他の手法に比べて低い値を示した高血圧では，LWA が最も小さい値を示しており，他のベイジアンの手法よりも低い値となっていた．このような傾向は他のデータ区分にみられないため，特徴的な結果だといえる．

また，LWA に次いで平均二乗誤差の値が大きい傾向にあったのは，経験ベイズ法による補正結果である．ただし，男性高血圧，女性高血圧，50－59 歳高血圧，60－69 歳高血圧，70 歳以上高血圧の5つのデータ区分においては，経験ベイズ法の平均二乗誤差の値は大きいとはいえず，階層ベイズ法との差もあまりないものといえる．

階層ベイズ法による 2 つのモデル間の比較では，高血圧を除きポアソンガンマモデルよりもポアソン対数正規モデルの方が平均二乗誤差の値が大きい傾向にあることがわかった．

② 手法間の補正結果の比較

上記で記述した平均二乗誤差は，全体としての単純指数との乖離度を示したグラフであった．ここでは，手法間の乖離度を示す平均二乗誤差を算出し，グラフ化した図 22 と，2 つの手法ごとに比較した散布図(図 23)より，手法間の補正結果を比較する．

図 22 より，最も精度が高いと考えられる階層ベイズと LWA との差は，40－49 歳高血圧，0－39 歳高血圧，50－59歳高血圧，高血圧の4つのデータ区分において大きな値を示している．これらのデータ区分においては，補正結果に大きな差が生じていることがわかる．

また，反対に，男性高血圧，女性高血圧，60－69 歳高血圧，70 歳以上高血圧では平均二乗誤差が約 0.01 と非常に小さな値となっており，最も精度が良い手法と考えられる階層ベイズ法と LWA の補正結果に大きな差が生じていない．

さらに，3 つのベイジアンの手法を比較してみると，ベイジアンの手法の中では比較的簡便な方法である経験ベイズ法と最も精度が良いと考えられる階層ベイズの手法は高血圧，男性高血圧，女性高血圧，50－59 歳高血圧，60－69 歳高血圧，70 歳以上高血圧の 6 つのデータ区分において，あまり大きな差は生じていないことがわかる．反対に，0－39 歳高血圧，40－49 歳高血圧の 2 つのデータ区分において，両手法による平均二乗誤差の値は大きく，またその差も大きいことがわかる．これら 2 つのデータ区分においては，ベイジアンの手法においても補正結果に差異が生じているといえる．なお，経験ベイズ法と 2 つの階層ベイズ法の比較では，特に経験ベイズ法とポアソン対数正規モデルを仮定した階層ベイズ法による平均二乗誤差に差が生じており，補正結果に差が生じるものといえる．

図 23 より，手法間の補正結果の関係をみることができる．手法の精度が高いと考えられる階層ベイズ法と LWA の間では，補正結果に最もばらつきが生じていることがわかる．また，経験ベイズ法と階層ベイズ法の間における，補正結果の方が LWA と階層ベイズ法よりもばらつきが少ないといえる．さらに，階層ベイズ法による 2 つのモデル間(ポアソンガンマモデルとポアソン対数正規モデル)では，補正結果に大きな差異はみられず，ほぼ一致しているようにみえる．

そこで，これら 3 つのベイジアンによる手法の結果をさらに詳細に比較するため，補正結果の差を期待件数との関係でみたものが，図 24 である．グラフより，期待件数が約 140~1，000 程度の範囲において両手法による補正差が 0．2 以上となっており，補正結果の間に差が生じることがわかる．

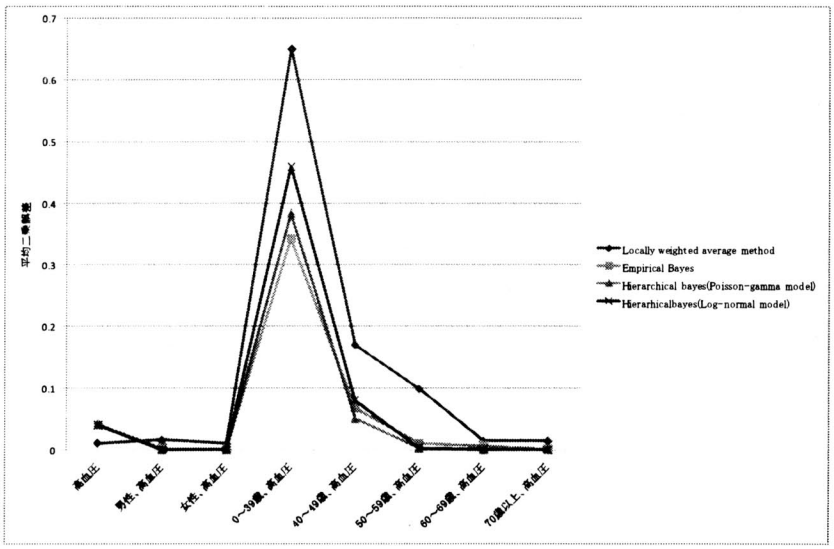


図 21 手法別の単純指数との平均二乗誤差

表 4 手法別の単純指数との平均二乗誤差

	Locally weighted average method	Empirical Bayes	Hierarchical bayes(Poisson-gamma model)	Hierarhicalbayes (Log-normal model)
高血圧	0.012	0.041	0.041	0.040
男性、高血圧	0.017	0.004	0.000	0.001
女性、高血圧	0.012	0.002	0.000	0.000
0～39歳、高血圧	0.651	0.343	0.383	0.460
40～49歳、高血圧	0.171	0.067	0.051	0.080
50～59歳、高血圧	0.099	0.011	0.003	0.004
60～69歳、高血圧	0.016	0.008	0.001	0.001
70歳以上、高血圧	0.016	0.002	0.000	0.000

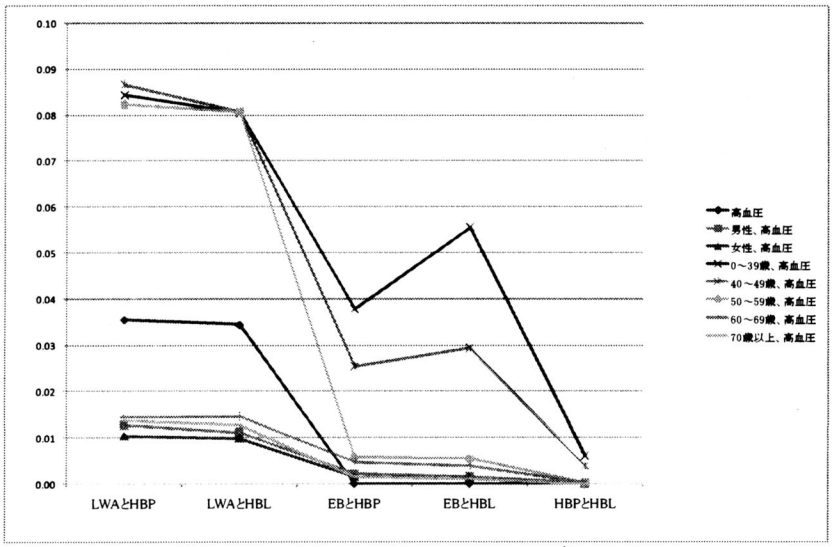


図 22 手法間の平均二乗誤差

表 5 手法間の平均二乗誤差

	LWAとHBP	LWAとHBL	EBとHBP	EBとHBL	HBPとHBL
高血圧	0.036	0.034	0.000	0.000	0.000
男性、高血圧	0.013	0.011	0.002	0.001	0.000
女性、高血圧	0.010	0.010	0.001	0.001	0.000
0～39歳、高血圧	0.084	0.080	0.038	0.056	0.006
40～49歳、高血圧	0.087	0.081	0.025	0.029	0.004
50～59歳、高血圧	0.082	0.081	0.006	0.005	0.000
60～69歳、高血圧	0.014	0.015	0.005	0.004	0.000
70歳以上、高血圧	0.014	0.013	0.001	0.001	0.000

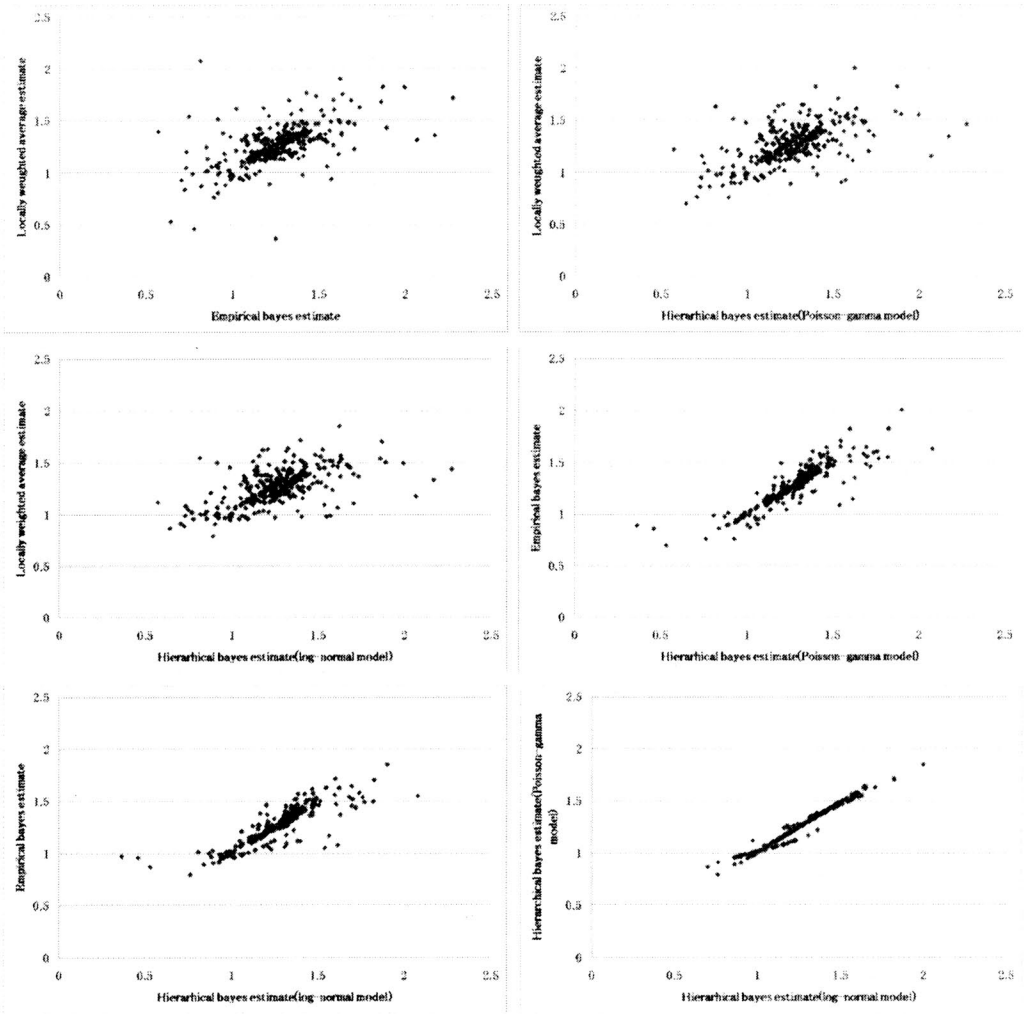


図 23 手法間の補正結果の比較

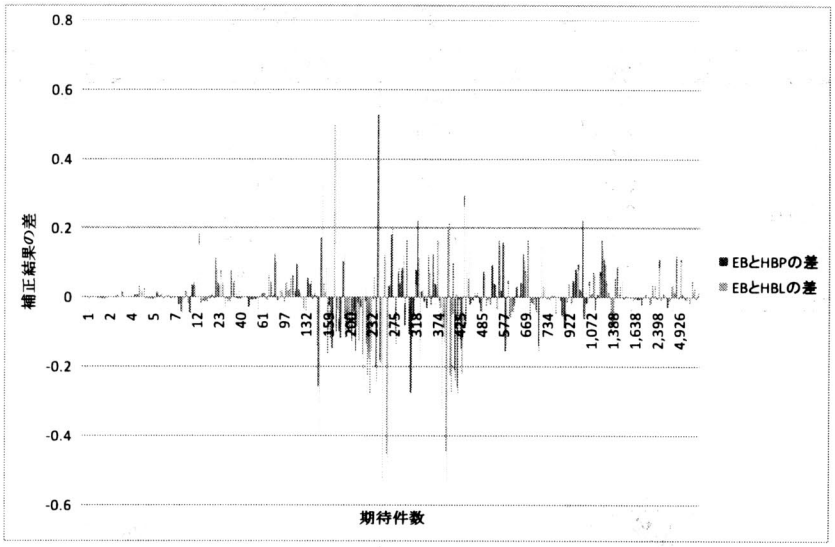


図 24 階層ベイズと経験ベイズの補正差と期待件数

また，階層ベイズ法による 2 つのモデルの比較では，DIC(Deviance Information Criterion)[24]による比較が可能である．V.G´omez and Rubio et al.(2009)によれば，ベイジアンモデルを用いて推定を行うとき，最も良いモデルを選択するために DIC 値を計算する．DIC は同様のデータを用いて推定したときのみ用いることができる．また，MCMC 法による事後分布のサンプリングから得られる値であるため，今回用いた手法の中では，階層ベイズ法による 2 つの手法(ポアソンガンマモデルとポアソン対数正規モデル)のみ比較が可能となる．つまり，DIC は事後分布のサンプリングを用いて計算されるため，モーメント法を用いた経験ベイズ法の場合には算出が難しい．なお，経験ベイズ推定値の不確実性の評価については，別途 Butar and Lahiri(2003)によって提案されているが，本研究では複数の経験ベイズ推定値を算出している訳ではないため，それを行うには至らなかった．また，モデルの比較において，DIC の値は最も低い値を示すモデルが良いものとされることから，相対的なモデルの比較を可能とするものであることに注意が必要である．

表 6 より，データ区分別に DIC の値を確認すると，高血圧，60－69 歳高血圧，70 歳以上高血圧の 3 つのデータ区分においては，ポアソン対数正規モデルの方が DIC の値が小さく，モデルの当てはまりが良いといえる．また，男性高血圧，女性高血圧，0－39 歳高血圧，40－49 歳高血圧，50－59 歳高血圧の 5 つのデータ区分においては，ポアソンガンマモデルの方が DIC の値が小さく，モデルの当てはまりが良いという結果になった．つまり，期待件数が少ない傾向を示すデータ区分においては，ポアソンガンマモデルの方が良い結果にあり，反対に，比較的期待件数の多いデータ区分(高血圧，60－69 歳高血圧，70 歳以上高血圧)においては，ポアソン対数正規モデルの方は当てはまりが良いということがわかった．ただし，DIC の差は最も大きいものであっても，0-39 歳高血圧データでの約 4 程度となっており，著しい差ではない．

[24] 詳しくは，Lawson et al.(2003)あるいは，Spiegelhalter et al.(2002)を参照されたい．

表 6 階層ベイズ法によるモデルの DIC 値

	Poisson-gamma model	Log-normal model
高血圧	388.55	388.52
男性、高血圧	355.18	355.74
女性、高血圧	369.01	368.68
0～39歳、高血圧	144.01	147.75
40～49歳、高血圧	218.08	220.85
50～59歳、高血圧	296.34	297.41
60～69歳、高血圧	334.90	334.30
70歳以上、高血圧	372.28	372.04

最後に，上記の分析結果をデータ区分ごとにまとめると，以下のような結果となる．

① 高血圧

いずれの手法においても，単純指数との絶対値として示した補正差の最大値はおよそ 0.5 以下となっており，あまり大きな補正は行われなかった．

全体として単純指数との乖離度を示し，低い値であるほど良いとされる平均二乗誤差の値は，LWA が 0.012 と最も小さく，また，経験ベイズ法と階層ベイズ法は共におよそ 0.04 という値を示した．

手法間の差の程度を示す平均二乗誤差の値では，LWA と階層ベイズ法の間の値が 0.036 と比較的大きな値を示した．また，ベイジアンの手法間における比較では，3 つの手法間(経験ベイズ法と階層ベイズ法(ポアソンガンマモデルとポアソン対数正規モデル))で補正結果にほとんど差がなかった．さらに，階層ベイズ法による 2 つのモデルの比較では，DIC の値からポアソン対数正規モデルの方が良いという結果が示された．ただし，その差は 0.03 と非常に小さな値であった．

② 男性高血圧

絶対値として示した補正差の最大値は，すべての手法で 0.4 以下となっており，大きな補正は行われていないことがわかった．単純指数との平均二乗誤差では，LWA が最も大きく 0.017，経験ベイズ法で 0.004，階層ベイズ法では非常に小さな値となった．

また，手法間の平均二乗誤差の値は，LWA と階層ベイズ法間においてポアソンガンマモデル 0.013，ポアソン対数正規モデル 0.011 という値となり，最も単純な方法である LWA と最も精度が高いとされる階層ベイズ法の間に大きな差は生じていないことがわかった．また，経験ベイズ法と階層ベイズ法間においても，大きな差は生じていないことがわかった．DIC の値による階層ベイズ法の 2 つのモデルの比較では，ポアソンガンマモデルの方が良いという結果が得られたが，その差は非常にわずかな差であった．

③ 女性高血圧

絶対値として示した補正差の最大値は全体として非常に小さい傾向にあり，最も高い値で LWA の 0.29 であった．全体として，あまり補正が行われていないことがわかった．単純指数との平均二乗誤差では LWA の値が最も大きく，その他の手法においては非常に小さな値となった．

また，手法間の平均二乗誤差では，LWA と階層ベイズ法の間で 0.01 と非常に小さく，また，経験ベイズ法と階層ベイズ法，ポアソンガンマモデルとポアソン対数正規モデル間に差はほぼないことが

わかった．DIC による階層ベイズ法の 2 つのモデルの比較では，ポアソン対数正規モデルの方が良いという結果になった．

④　0－39 歳高血圧

絶対値として示した補正差の最大値は LWA で 2.65，経験ベイズ法で 1.82，階層ベイズ法(ポアソンガンマモデル)で 1.99，階層ベイズ法(ポアソン対数正規モデル)で 2.10 と，すべての手法において非常に大きな補正がなされていることがわかった．

また，単純指数との平均二乗誤差では，LWA で最も大きく 0.651，経験ベイズ法で 0.343，階層ベイズ法のポアソンガンマモデルで 0.383，階層ベイズ法のポアソン対数正規モデルで 0.46 と，経験ベイズ法が最も小さく，ついで階層ベイズ法のポアソンガンマモデルが小さい値となった．なお，経験ベイズ法による値とポアソンガンマモデルによる値にはわずかな差しか存在しなかった．

手法間の平均二乗誤差では，LWA と階層ベイズ法による 2 つのモデルが共におよそ 0.08 となり，最も大きな値を示した．両手法の補正結果は他の手法間の値よりも大きい結果となった．また，経験ベイズ法と階層ベイズ法による 2 つのモデル間では，ポアソンガンマモデルとの平均二乗誤差が 0.038，ポアソン対数正規モデルが 0.056 となり，階層ベイズ法の中でもポアソン対数正規モデルの方が経験ベイズ法との差が大きい傾向にあることがわかった．

また，DIC による比較では，ポアソンガンマモデルの方が良いという結果になった．

⑤　40－49 歳高血圧

絶対値として示した補正差の最大値は，LWA で 1.27，経験ベイズ法で 0.52，階層ベイズ法のポアソンガンマモデルで 0.48，階層ベイズ法のポアソン対数正規モデルで 0.61 となっており，0－39 歳高血圧のデータ区分についで，大きな値を示す結果となった．

単純指数との平均二乗誤差では，LWA の値が 0.171 と最も大きく，ついで経験ベイズ法による 0.067，階層ベイズ法のポアソンガンマモデルでは 0.051，階層ベイズ法のポアソン対数正規モデルでは 0.08 となり，階層ベイズ法のポアソン対数正規モデルが最も小さな値となった．

さらに，手法間の平均二乗誤差では，LWA と階層ベイズ法の 2 つのモデル間でおよそ 0.08 と最も大きな値となった．また，経験ベイズ法と階層ベイズ法の比較ではポアソン対数正規モデルとの平均二乗誤差の方がポアソンガンマモデルよりも大きい値を示していた．DIC による比較では，階層ベイズ法による手法のうち，ポアソンガンマモデルの方が良いという結果になった．

⑥　50－59 歳高血圧

絶対値とした補正差の最大値は，LWA で 1.04，経験ベイズ法で 0.29，階層ベイズ法のポアソンガンマモデルで 0.14，階層ベイズ法のポアソン対数正規モデルで 0.16 と，0－39 歳高血圧，40－49 歳高血圧のデータ区分についで比較的大きな補正がなされていることがわかった．

また，単純指数との平均二乗誤差では LWA で 0.09，経験ベイズ法で 0.01，階層ベイズ法のポアソンガンマモデルで 0.003，階層ベイズ法のポアソン対数正規モデルで 0.004 となっており，LWA が最も大きく，階層ベイズ法による補正が最も良いということが示唆された．

さらに，手法間の平均二乗誤差では LWA と階層ベイズ法による 2 つのモデル間の値は大きくなっており，およそ 0.08 であった．一方，階層ベイズ法と経験ベイズ法間の値は 0.006 前後と非常に小さな値となった．

DIC による階層ベイズ法の 2 つのモデルの比較では，ポアソンガンマモデルが良いという結果になった．

⑦　60－69 歳高血圧

絶対値で示した補正差の最大値は，LWA で 0.31，経験ベイズ法で 0.31，階層ベイズ法のポアソンガンマモデルで 0.08，ポアソン対数正規モデルで 0.12 といずれも著しく大きな補正はなされていないことがわかった．

単純指数との平均二乗誤差では，最大で LWA が 0.016，最小で階層ベイズ法の 2 つのモデルが 0.001 とすべての手法において小さな値を示している．また，手法間の平均二乗誤差でもすべての手法間で 0.015 以下となっており，手法間で差が生じていないことがわかった．

⑧　70 歳以上高血圧

絶対値として示した補正差の最大値は，LWA で 0.49，経験ベイズ法で 0.16，階層ベイズ法のポアソンガンマモデルで 0.04，階層ベイズ法のポアソン対数正規モデルで 0.07 となり，すべての手法において大きな補正はなされていないことがわかった．

また，単純指数との平均二乗誤差では，LWA で最大の 0.016，3 つのベイジアンの手法では共に 0.002 以下を示した．さらに，手法間の平均二乗誤差では，LWA と階層ベイズ法の値が最大となっており，0.014 前後であった．したがって，手法によって補正結果に大きな差は生じていないということがわかった．

最後に，表 7 から表 14 の指数一覧表を確認すると，0－39 歳高血圧のデータ区分において実績件数がゼロとなる市町村がいくつか存在していることがわかる．これらの指数は単純指数として示すと，当然ながらゼロと算出される．しかし，各種の補正手法により算出された値はゼロではなく，実数値が算出されている．したがって，McEwin and Elazar(2006)も指摘しているように，標本が得られないような場合にも SAE の手法は役立つものといえる．

また，経験ベイズ法の場合には，期待件数が少ないときに補正式中の ϕ 二乗が負値となり，ゼロで打ち切りとなる市町村がいくつかみられた．この場合には補正はなされず，結果として補正前の単純指数の値が返される．これは，モーメント法を用いた経験ベイズ法の欠点だと考えられる．

表 7 山形県 高血圧データの補正結果

	データ					Locally weighted average	Empirical beyes	Hierarchical bayes	
	市町村	被保険者数	実績件数	期待件数	単純指数		Poisson-gamma model (モーメント法)	Poisson-gamma model	Log-normal model
1	山形市	22,813,756	1,230	854	1.440	1.194	1.219	1.219	1.218
2	米沢市	8,276,604	2,223	1,510	1.472	1.424	1.408	1.407	1.406
3	鶴岡市	15,145,025	16,594	13,618	1.219	1.241	1.156	1.157	1.157
4	酒田市	12,351,192	7,013	4,981	1.408	1.405	0.978	0.979	0.980
5	新庄市	4,607,092	2,861	2,398	1.193	1.275	1.134	1.134	1.135
6	寒河江市	4,094,790	963	692	1.391	1.354	1.194	1.194	1.194
7	上山市	3,645,684	1,506	1,204	1.251	1.277	1.192	1.192	1.193
8	村山市	2,919,189	835	501	1.668	1.340	1.168	1.167	1.169
9	長井市	2,869,720	1,596	1,165	1.370	1.412	1.275	1.270	1.270
10	天童市	6,052,830	760	638	1.192	1.395	1.241	1.241	1.240
11	東根市	4,380,985	802	637	1.258	1.265	1.333	1.334	1.334
12	尾花沢市	2,604,504	711	567	1.254	1.221	1.245	1.245	1.245
13	南陽市	3,500,126	916	674	1.358	1.377	1.340	1.337	1.336
14	山辺町	1,389,794	480	426	1.127	1.148	1.427	1.431	1.427
15	中山町	1,149,701	900	752	1.196	1.170	1.379	1.385	1.380
16	河北町	1,968,918	513	433	1.186	1.292	1.251	1.251	1.251
17	西川町	708,624	821	681	1.206	1.178	1.623	1.640	1.627
18	朝日町	1,052,576	370	270	1.371	1.116	1.196	1.196	1.197
19	大江町	1,020,201	439	355	1.236	1.252	1.257	1.259	1.258
20	大石田町	1,002,071	523	382	1.369	1.214	1.253	1.255	1.254
21	金山町	840,455	2,668	2,240	1.191	1.246	1.137	1.137	1.143
22	最上町	1,406,370	2,075	1,781	1.165	1.229	1.195	1.199	1.200
23	舟形町	740,999	2,170	1,708	1.271	1.329	1.186	1.192	1.193
24	真室川町	1,235,430	4,120	3,321	1.240	1.164	1.204	1.210	1.210
25	大蔵村	526,288	3,290	2,464	1.335	1.259	1.326	1.358	1.350
26	鮭川村	675,472	1,847	1,484	1.245	1.292	1.226	1.239	1.239
27	戸沢村	709,919	2,633	1,966	1.339	1.318	1.335	1.360	1.353
28	高畠町	2,723,385	1,309	983	1.332	1.429	1.468	1.467	1.464
29	川西町	1,963,220	780	510	1.531	1.385	1.370	1.367	1.364
30	小国町	938,616	574	507	1.132	1.078	1.360	1.354	1.350
31	白鷹町	1,662,003	1,532	1,638	0.936	1.072	1.335	1.330	1.328
32	飯豊町	818,958	1,111	1,093	1.017	0.915	1.515	1.513	1.503
33	三川町	864,801	10,352	8,949	1.157	1.120	1.127	1.141	1.146
34	庄内町	2,804,193	7,280	7,446	0.978	0.990	0.939	0.943	0.948
35	遊佐町	1,816,885	2,754	2,433	1.132	1.155	1.018	1.025	1.030

データ出典:「国、都道府県の医療費適正化計画の重点対象の発見に関する研究」のデータより著者作成
※被保険者数とは、該当するデータ区分のみの被保険者数を指す。

表 8 山形県 男性 高血圧データ補正結果

	データ					Locally weighted average	Empirical beyes	Hierarchical bayes	
	市町村	被保険者数	実績件数	期待件数	単純指数		Poisson-gamma model (モーメント法)	Poisson-gamma model	Log-normal model
1	山形市	40,168	6,702	5,462	1.227	1.250	1.227	1.227	1.227
2	米沢市	14,858	2,728	1,984	1.375	1.391	1.367	1.372	1.371
3	鶴岡市	27,314	4,130	3,543	1.166	1.133	1.162	1.166	1.167
4	酒田市	21,856	2,848	2,938	0.970	0.984	0.973	0.973	0.975
5	新庄市	8,360	1,102	982	1.122	1.160	1.127	1.126	1.128
6	寒河江市	7,552	1,193	1,010	1.181	1.237	1.190	1.184	1.184
7	上山市	6,680	1,115	922	1.209	1.228	1.212	1.210	1.210
8	村山市	5,403	819	734	1.115	1.220	1.141	1.121	1.124
9	長井市	5,328	899	703	1.278	1.305	1.319	1.277	1.274
10	天童市	11,122	1,653	1,395	1.185	1.131	1.191	1.186	1.187
11	東根市	8,135	1,406	1,049	1.340	1.201	1.321	1.337	1.334
12	尾花沢市	4,893	718	623	1.153	1.250	1.172	1.158	1.160
13	南陽市	6,379	1,015	800	1.269	1.315	1.313	1.267	1.265
14	山辺町	2,532	474	353	1.342	1.163	1.299	1.334	1.326
15	中山町	2,143	402	296	1.359	1.282	1.304	1.347	1.338
16	河北町	3,646	619	504	1.228	1.253	1.228	1.229	1.229
17	西川町	1,295	363	212	1.712	1.314	1.466	1.649	1.618
18	朝日町	1,984	314	269	1.167	1.366	1.194	1.175	1.178
19	大江町	1,882	314	263	1.193	1.243	1.209	1.198	1.200
20	大石田町	1,899	275	232	1.183	1.165	1.205	1.191	1.192
21	金山町	1,602	183	171	1.068	1.126	1.110	1.096	1.113
22	最上町	2,647	346	309	1.120	1.132	1.134	1.131	1.138
23	舟形町	1,401	208	180	1.155	1.245	1.161	1.168	1.173
24	真室川町	2,299	313	279	1.122	1.131	1.136	1.134	1.140
25	大蔵村	971	165	108	1.528	1.131	1.344	1.460	1.428
26	鮭川村	1,277	189	147	1.284	1.222	1.234	1.276	1.269
27	戸沢村	1,355	232	161	1.442	1.191	1.329	1.408	1.387
28	高畠町	5,216	912	637	1.432	1.326	1.384	1.423	1.416
29	川西町	3,688	633	480	1.319	1.357	1.340	1.314	1.309
30	小国町	1,646	375	275	1.363	1.381	1.353	1.351	1.341
31	白鷹町	3,203	574	425	1.349	1.343	1.350	1.341	1.335
32	飯豊町	1,513	308	207	1.486	1.349	1.375	1.453	1.433
33	三川町	1,631	236	207	1.142	1.051	1.114	1.156	1.163
34	庄内町	5,168	625	670	0.933	1.041	0.951	0.947	0.959
35	遊佐町	3,267	409	424	0.965	0.915	0.985	0.985	1.000

データ出典:「国、都道府県の医療費適正化計画の重点対象の発見に関する研究」のデータより著者作成
※被保険者数とは、該当するデータ区分のみの被保険者数を指す。

表 9 山形県 女性 高血圧データ補正結果

	データ					Locally weighted average	Empirical beyes	Hierarchical bayes	
	市町村	被保険者数	実績件数	期待件数	単純指数		Poisson-gamma model (モーメント法)	Poisson-gamma model	Log-normal model
1	山形市	47,916	9,892	8,156	1.213	1.234	1.214	1.213	1.213
2	米沢市	17,098	4,285	2,996	1.430	1.414	1.427	1.428	1.427
3	鶴岡市	31,161	6,222	5,406	1.151	1.111	1.149	1.152	1.152
4	酒田市	25,832	4,432	4,508	0.983	0.994	0.986	0.986	0.988
5	新庄市	9,428	1,652	1,450	1.139	1.152	1.170	1.143	1.145
6	寒河江市	8,258	1,668	1,388	1.201	1.303	1.208	1.203	1.204
7	上山市	7,396	1,553	1,317	1.179	1.259	1.189	1.181	1.183
8	村山市	5,868	1,256	1,046	1.201	1.236	1.210	1.203	1.204
9	長井市	5,752	1,271	1,004	1.266	1.346	1.302	1.266	1.266
10	天童市	12,248	2,467	1,926	1.281	1.187	1.278	1.280	1.280
11	東根市	8,780	1,884	1,415	1.332	1.302	1.321	1.330	1.329
12	尾花沢市	5,163	1,129	861	1.311	1.322	1.299	1.310	1.308
13	南陽市	7,135	1,618	1,166	1.387	1.319	1.391	1.383	1.382
14	山辺町	2,834	756	501	1.508	1.215	1.431	1.492	1.483
15	中山町	2,296	561	397	1.414	1.409	1.357	1.403	1.395
16	河北町	3,956	887	700	1.267	1.295	1.264	1.268	1.267
17	西川町	1,441	472	289	1.635	1.359	1.471	1.596	1.575
18	朝日町	2,080	446	369	1.210	1.416	1.226	1.216	1.218
19	大江町	2,057	488	374	1.305	1.280	1.286	1.301	1.299
20	大石田町	1,970	436	334	1.304	1.261	1.284	1.301	1.298
21	金山町	1,643	297	255	1.166	1.164	1.193	1.180	1.187
22	最上町	2,783	554	443	1.249	1.196	1.208	1.251	1.252
23	舟形町	1,460	305	252	1.208	1.325	1.198	1.217	1.219
24	真室川町	2,471	508	402	1.265	1.210	1.210	1.266	1.265
25	大蔵村	1,061	205	162	1.266	1.107	1.203	1.268	1.266
26	鮭川村	1,331	250	208	1.203	1.274	1.198	1.213	1.218
27	戸沢村	1,386	291	221	1.315	1.230	1.211	1.310	1.305
28	高畠町	5,299	1,311	873	1.501	1.495	1.472	1.492	1.486
29	川西町	3,892	963	685	1.405	1.451	1.404	1.398	1.394
30	小国町	1,978	541	399	1.355	1.374	1.377	1.348	1.343
31	白鷹町	3,214	735	557	1.319	1.495	1.352	1.316	1.314
32	飯豊町	1,649	472	302	1.561	1.409	1.475	1.530	1.514
33	三川町	1,708	338	300	1.125	1.097	1.104	1.141	1.151
34	庄内町	5,659	907	967	0.938	1.094	0.955	0.950	0.960
35	遊佐町	3,748	702	669	1.049	0.914	1.052	1.060	1.070

データ出典:「国、都道府県の医療費適正化計画の重点対象の発見に関する研究」のデータより著者作成
※被保険者数とは、該当するデータ区分のみの被保険者数を指す。

表 10 山形県 0-39 歳 高血圧データ補正結果

	データ					Locally weighted average	Empirical beyes	Hierarchical bayes	
	市町村	被保険者数	実績件数	期待件数	単純指数		Poisson-gamma model (モーメント法)	Poisson-gamma model	Log-normal model
1	山形市	19,432	53	36	1.475	1.320	1.361	1.364	1.221
2	米沢市	6,548	19	12	1.564	1.339	1.564	1.306	1.170
3	鶴岡市	12,080	12	23	0.532	0.642	0.532	0.702	0.868
4	酒田市	9,420	17	18	0.971	1.002	0.971	0.991	1.011
5	新庄市	4,499	6	8	0.719	0.958	0.884	0.899	0.968
6	寒河江市	3,462	2	6	0.311	0.704	0.929	0.765	0.914
7	上山市	2,781	8	5	1.543	1.704	1.228	1.204	1.108
8	村山市	2,238	2	4	0.480	0.854	1.020	0.875	0.962
9	長井市	2,326	2	4	0.460	0.778	0.460	0.864	0.959
10	天童市	5,688	8	11	0.758	0.924	1.003	0.901	0.968
11	東根市	3,915	6	7	0.825	0.730	1.052	0.952	0.991
12	尾花沢市	2,156	3	4	0.749	0.906	1.072	0.954	0.996
13	南陽市	3,097	5	6	0.872	0.819	0.872	0.977	1.003
14	山辺町	1,004	3	2	1.602	1.612	1.183	1.118	1.068
15	中山町	933	2	2	1.150	1.380	1.142	1.049	1.038
16	河北町	1,466	3	3	1.100	0.855	1.135	1.045	1.032
17	西川町	377	0	1	0.000	0.923	1.100	0.973	1.009
18	朝日町	813	3	2	1.975	0.733	1.203	1.160	1.085
19	大江町	734	0	1	0.000	1.546	1.063	0.906	0.981
20	大石田町	845	4	2	2.521	0.848	1.248	1.233	1.122
21	金山町	832	1	2	0.636	0.845	1.015	0.988	1.009
22	最上町	1,317	5	2	2.031	0.944	1.382	1.220	1.120
23	舟形町	585	0	1	0.000	1.570	0.942	0.928	0.990
24	真室川町	1,126	4	2	1.903	1.534	1.320	1.184	1.097
25	大蔵村	503	3	1	3.218	0.570	1.398	1.224	1.119
26	鮭川村	616	2	1	1.733	1.019	1.215	1.107	1.062
27	戸沢村	635	0	1	0.000	1.060	0.931	0.925	0.988
28	高畠町	2,442	6	5	1.308	1.079	1.308	1.110	1.066
29	川西町	1,468	1	3	0.366	1.249	0.366	0.893	0.972
30	小国町	446	1	1	1.232	1.183	1.232	1.049	1.039
31	白鷹町	1,325	4	2	1.616	1.021	1.616	1.150	1.080
32	飯豊町	605	1	1	0.893	1.213	0.893	1.017	1.028
33	三川町	691	2	1	1.541	0.747	1.541	1.090	1.054
34	庄内町	2,167	4	4	0.988	0.769	0.988	1.019	1.024
35	遊佐町	1,324	2	2	0.809	0.915	0.809	0.990	1.013

データ出典:「国、都道府県の医療費適正化計画の重点対象の発見に関する研究」のデータより著者作成
※被保険者数とは、該当するデータ区分のみの被保険者数を指す。

表 11 山形県 40-49 歳 高血圧データ補正結果

	データ					Locally weighted average	Empirical beyes	Hierarchical bayes	
	市町村	被保険者数	実績件数	期待件数	単純指数		Poisson-gamma model (モーメント法)	Poisson-gamma model	Log-normal model
1	山形市	5,442	183	132	1.384	1.439	1.387	1.391	1.390
2	米沢市	2,093	83	51	1.620	1.859	1.681	1.585	1.542
3	鶴岡市	3,983	127	98	1.301	1.220	1.247	1.321	1.333
4	酒田市	2,976	83	73	1.142	1.101	1.154	1.193	1.232
5	新庄市	1,522	43	37	1.158	1.341	1.158	1.241	1.278
6	寒河江市	1,019	26	25	1.041	1.240	1.270	1.186	1.252
7	上山市	866	40	21	1.888	1.525	1.548	1.711	1.630
8	村山市	708	16	17	0.929	1.582	1.268	1.159	1.239
9	長井市	703	22	17	1.274	1.733	1.630	1.353	1.369
10	天童市	1,598	44	39	1.131	1.084	1.274	1.216	1.262
11	東根市	1,177	46	29	1.599	1.352	1.472	1.550	1.508
12	尾花沢市	799	41	20	2.092	1.395	1.601	1.823	1.717
13	南陽市	1,029	41	25	1.625	1.582	1.702	1.560	1.520
14	山辺町	344	11	8	1.315	1.326	1.381	1.396	1.401
15	中山町	250	7	6	1.141	1.629	1.364	1.351	1.370
16	河北町	488	14	12	1.170	1.354	1.347	1.317	1.347
17	西川町	145	7	4	1.957	1.890	1.433	1.557	1.506
18	朝日町	295	8	7	1.107	1.386	1.355	1.329	1.359
19	大江町	267	12	7	1.839	1.304	1.448	1.574	1.517
20	大石田町	292	10	7	1.400	1.402	1.394	1.431	1.422
21	金山町	303	9	7	1.208	1.069	1.208	1.361	1.374
22	最上町	471	17	12	1.469	1.259	1.469	1.458	1.440
23	舟形町	202	9	5	1.821	1.993	1.821	1.554	1.499
24	真室川町	356	14	9	1.606	1.585	1.606	1.514	1.473
25	大蔵村	216	11	5	2.078	0.813	2.078	1.632	1.551
26	鮭川村	246	7	6	1.166	1.208	1.166	1.361	1.382
27	戸沢村	229	10	6	1.768	1.425	1.768	1.544	1.492
28	高畠町	759	45	19	2.417	1.621	1.900	2.002	1.853
29	川西町	529	20	13	1.520	1.684	1.696	1.485	1.457
30	小国町	164	6	4	1.494	2.273	1.720	1.463	1.440
31	白鷹町	439	25	11	2.310	1.868	1.825	1.825	1.704
32	飯豊町	198	10	5	2.049	1.634	1.760	1.610	1.540
33	三川町	225	9	6	1.610	1.145	1.201	1.495	1.465
34	庄内町	776	12	19	0.630	1.079	1.048	0.973	1.121
35	遊佐町	468	10	12	0.866	0.959	1.123	1.180	1.262

データ出典:「国、都道府県の医療費適正化計画の重点対象の発見に関する研究」のデータより著者作成
※被保険者数とは、該当するデータ区分のみの被保険者数を指す。

表 12 山形県 50-59 歳 高血圧データ補正結果

	データ					Locally weighted average	Empirical beyes	Hierarchical bayes	
	市町村	被保険者数	実績件数	期待件数	単純指数		Poisson-gamma model (モーメント法)	Poisson-gamma model	Log-normal model
1	山形市	10,585	1,167	921	1.267	1.383	1.288	1.268	1.268
2	米沢市	4,112	526	355	1.483	1.673	1.483	1.475	1.473
3	鶴岡市	8,424	806	726	1.110	1.066	1.101	1.116	1.117
4	酒田市	6,701	516	581	0.889	0.903	0.896	0.903	0.908
5	新庄市	2,599	246	225	1.093	1.295	1.129	1.112	1.115
6	寒河江市	2,092	229	181	1.267	1.381	1.314	1.273	1.271
7	上山市	1,804	226	155	1.455	1.299	1.365	1.440	1.435
8	村山市	1,577	175	135	1.299	1.467	1.325	1.303	1.300
9	長井市	1,483	179	127	1.406	1.560	1.406	1.395	1.391
10	天童市	3,283	367	284	1.291	1.052	1.317	1.293	1.293
11	東根市	2,308	298	200	1.492	1.421	1.383	1.477	1.472
12	尾花沢市	1,638	196	140	1.396	1.281	1.348	1.387	1.383
13	南陽市	1,913	264	165	1.597	1.512	1.597	1.568	1.562
14	山辺町	747	103	65	1.589	1.122	1.366	1.529	1.518
15	中山町	593	70	51	1.366	1.372	1.336	1.356	1.349
16	河北町	1,014	138	87	1.580	1.534	1.374	1.533	1.524
17	西川町	272	24	23	1.031	2.067	1.318	1.162	1.178
18	朝日町	561	68	48	1.413	1.141	1.341	1.389	1.380
19	大江町	525	56	45	1.248	1.283	1.325	1.271	1.269
20	大石田町	680	74	58	1.273	1.144	1.326	1.284	1.284
21	金山町	572	54	49	1.098	0.952	1.186	1.162	1.167
22	最上町	889	100	76	1.309	1.089	1.282	1.312	1.308
23	舟形町	476	48	41	1.179	1.328	1.224	1.224	1.227
24	真室川町	761	88	65	1.347	1.403	1.298	1.341	1.338
25	大蔵村	333	45	29	1.574	0.987	1.348	1.475	1.458
26	鮭川村	465	57	40	1.430	1.134	1.318	1.396	1.387
27	戸沢村	481	70	41	1.706	1.157	1.424	1.587	1.571
28	高畠町	1,664	242	143	1.696	1.325	1.696	1.651	1.645
29	川西町	1,244	152	106	1.428	1.344	1.428	1.411	1.408
30	小国町	366	43	32	1.358	2.168	1.358	1.346	1.336
31	白鷹町	937	121	80	1.506	1.612	1.506	1.471	1.464
32	飯豊町	432	64	37	1.739	1.481	1.739	1.598	1.581
33	三川町	544	54	47	1.157	0.877	1.060	1.206	1.211
34	庄内町	1,714	102	148	0.691	0.891	0.763	0.763	0.795
35	遊佐町	1,129	75	97	0.771	0.724	0.840	0.864	0.891

データ出典:「国、都道府県の医療費適正化計画の重点対象の発見に関する研究」のデータより著者作成
※被保険者数とは、該当するデータ区分のみの被保険者数を指す。

表 13 山形県 60-69 歳 高血圧データ補正結果

	データ					Locally weighted average	Empirical beyes	Hierarchical bayes	
	市町村	被保険者数	実績件数	期待件数	単純指数		Poisson-gamma model (モーメント法)	Poisson-gamma model	Log-normal model
1	山形市	19,447	4,446	3,354	1.326	1.349	1.328	1.326	1.326
2	米沢市	7,262	1,836	1,253	1.466	1.428	1.471	1.462	1.460
3	鶴岡市	12,502	2,663	2,154	1.236	1.190	1.232	1.238	1.239
4	酒田市	10,784	1,829	1,849	0.989	0.990	0.994	0.996	1.001
5	新庄市	3,535	744	605	1.230	1.203	1.282	1.238	1.240
6	寒河江市	3,477	803	599	1.340	1.343	1.340	1.340	1.339
7	上山市	3,168	721	547	1.318	1.368	1.330	1.319	1.320
8	村山市	2,295	489	394	1.240	1.445	1.304	1.249	1.253
9	長井市	2,471	581	425	1.367	1.382	1.454	1.366	1.363
10	天童市	5,194	1,165	894	1.303	1.179	1.320	1.305	1.306
11	東根市	3,764	927	650	1.426	1.265	1.381	1.423	1.420
12	尾花沢市	1,835	414	318	1.303	1.517	1.329	1.308	1.310
13	南陽市	2,832	752	485	1.552	1.399	1.492	1.539	1.529
14	山辺町	1,179	297	202	1.469	1.333	1.369	1.451	1.440
15	中山町	974	249	168	1.486	1.505	1.368	1.463	1.449
16	河北町	1,718	407	296	1.377	1.424	1.351	1.375	1.370
17	西川町	600	177	105	1.693	1.629	1.385	1.610	1.573
18	朝日町	795	169	136	1.239	1.447	1.324	1.261	1.269
19	大江町	817	186	142	1.310	1.493	1.335	1.317	1.318
20	大石田町	670	161	114	1.407	1.389	1.349	1.394	1.384
21	金山町	573	127	98	1.298	1.327	1.303	1.312	1.313
22	最上町	968	233	167	1.394	1.381	1.313	1.387	1.380
23	舟形町	546	116	94	1.233	1.544	1.299	1.265	1.272
24	真室川町	913	205	158	1.301	1.313	1.303	1.310	1.310
25	大蔵村	358	86	61	1.399	1.227	1.307	1.382	1.370
26	鮭川村	443	117	76	1.538	1.535	1.315	1.481	1.454
27	戸沢村	500	125	86	1.448	1.410	1.311	1.421	1.405
28	高畠町	2,101	561	361	1.555	1.567	1.489	1.537	1.526
29	川西町	1,508	392	259	1.514	1.665	1.480	1.495	1.484
30	小国町	913	219	159	1.378	1.477	1.467	1.374	1.367
31	白鷹町	1,304	312	222	1.407	1.705	1.467	1.399	1.393
32	飯豊町	648	178	112	1.591	1.621	1.482	1.536	1.508
33	三川町	645	143	111	1.284	1.247	1.213	1.298	1.302
34	庄内町	2,288	381	395	0.965	1.130	0.987	0.995	1.018
35	遊佐町	1,474	254	254	1.002	0.979	1.026	1.042	1.070

データ出典:「国、都道府県の医療費適正化計画の重点対象の発見に関する研究」のデータより著者作成
※被保険者数とは、該当するデータ区分のみの被保険者数を指す。

表 14 山形県 70歳以上 高血圧データ補正結果

	データ					Locally weighted average	Empirical beyes	Hierarchical bayes	
	市町村	被保険者数	実績件数	期待件数	単純指数		Poisson-gamma model (モーメント法)	Poisson-gamma model	Log-normal model
1	山形市	33,178	10,745	9,175	1.171	1.184	1.171	1.171	1.171
2	米沢市	11,941	4,549	3,310	1.374	1.361	1.370	1.373	1.371
3	鶴岡市	21,486	6,744	5,948	1.134	1.101	1.131	1.134	1.134
4	酒田市	17,807	4,835	4,926	0.982	0.998	0.985	0.983	0.985
5	新庄市	5,633	1,715	1,558	1.101	1.113	1.112	1.104	1.106
6	寒河江市	5,760	1,801	1,587	1.135	1.240	1.141	1.137	1.138
7	上山市	5,457	1,673	1,511	1.107	1.190	1.116	1.110	1.112
8	村山市	4,453	1,393	1,230	1.133	1.130	1.140	1.135	1.137
9	長井市	4,097	1,386	1,134	1.223	1.279	1.245	1.222	1.222
10	天童市	7,607	2,536	2,094	1.211	1.175	1.210	1.212	1.211
11	東根市	5,751	2,013	1,579	1.275	1.237	1.268	1.274	1.272
12	尾花沢市	3,628	1,193	1,002	1.190	1.222	1.191	1.192	1.191
13	南陽市	4,643	1,571	1,286	1.222	1.259	1.242	1.222	1.221
14	山辺町	2,092	816	577	1.414	1.150	1.366	1.401	1.393
15	中山町	1,689	635	466	1.364	1.295	1.320	1.353	1.344
16	河北町	2,916	944	806	1.171	1.196	1.175	1.173	1.174
17	西川町	1,342	627	369	1.701	1.207	1.546	1.656	1.633
18	朝日町	1,600	512	444	1.152	1.409	1.164	1.156	1.159
19	大江町	1,596	548	443	1.238	1.188	1.227	1.238	1.234
20	大石田町	1,382	462	386	1.198	1.181	1.197	1.200	1.199
21	金山町	965	289	270	1.070	1.123	1.116	1.088	1.100
22	最上町	1,785	545	495	1.101	1.110	1.120	1.110	1.115
23	舟形町	1,052	340	292	1.165	1.192	1.139	1.172	1.173
24	真室川町	1,614	510	447	1.141	1.088	1.134	1.148	1.150
25	大蔵村	622	225	174	1.296	1.111	1.157	1.280	1.272
26	鮭川村	838	256	232	1.103	1.182	1.124	1.119	1.129
27	戸沢村	896	318	248	1.282	1.151	1.163	1.273	1.267
28	高畠町	3,549	1,369	983	1.392	1.383	1.376	1.385	1.380
29	川西町	2,831	1,031	784	1.315	1.334	1.318	1.310	1.306
30	小国町	1,735	647	479	1.351	1.284	1.341	1.342	1.334
31	白鷹町	2,412	847	667	1.269	1.310	1.287	1.267	1.264
32	飯豊町	1,279	527	355	1.485	1.297	1.411	1.461	1.443
33	三川町	1,234	366	342	1.070	1.051	1.065	1.084	1.094
34	庄内町	3,882	1,033	1,072	0.964	1.077	0.979	0.972	0.980
35	遊佐町	2,620	770	728	1.058	0.917	1.057	1.065	1.072

データ出典:「国、都道府県の医療費適正化計画の重点対象の発見に関する研究」のデータより著者作成
※被保険者数とは、該当するデータ区分のみの被保険者数を指す。

5.1.2. 考察

以上の分析結果より，全体として補正が大きくなされたデータ区分は，0－39 歳高血圧，40－49 歳高血圧，50－59 歳高血圧であり，これらのデータ区分は他のデータ区分に比べ期待件数が少ない傾向にあることがわかった．これら 3 つの区分では，すべての手法において大きな補正がなされていることから，補正の必要性が高いデータ区分であるものと考えられる．例として，0－39 歳高血圧の補正結果を地図化したものが，図 25 である．精度の高い手法(右に位置する手法ほど精度が高い)による補正結果ほど指数の変化のグラデーションが滑らかになっていることが視覚的に確認することができる．手法により補正結果の大きく異なるデータ区分の場合には，マッピングした結果も大きく異なっていることから，政策判断材料として用いる際の補正の重要性が示唆される．

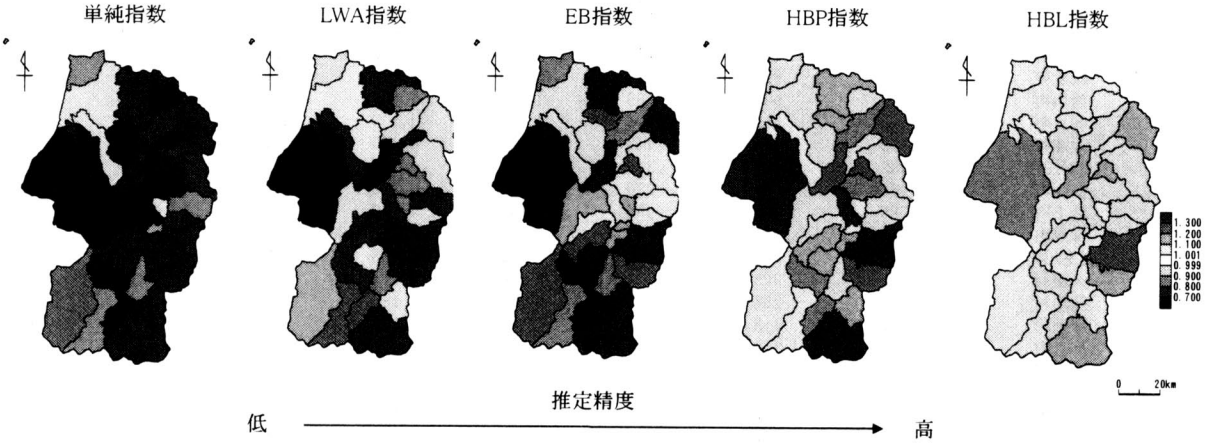


図 25 0－39 歳高血圧，医療費指数の補正結果

一方，上記 3 つのデータ区分以外(高血圧，男性高血圧，女性高血圧，60－69 歳高血圧，70 歳以上高血圧の 5 つ)においては，各手法による補正結果の差に著しい差異は観察されず，上記 3 つのデータ区分に比べて，補正の必要性が低いものと考えられる．例として，高血圧と男性高血圧の補正結果を地図化したものが，図 26，図 27 である．上記に示した 0－39 歳高血圧の医療費指数ほどは補正がなされていないことが視覚的に確認できる．また，高血圧のデータ区分では，経験ベイズ法と階層ベイズ法による補正結果は酷似していることが確認できる．さらに，男性高血圧のデータ区分では，殆ど補正がなされておらず，地図の様子もほとんど変化していないことが視覚的に確認できる．この点は，上記の 0－39 歳の補正結果と大きな異なっており，特徴的である．

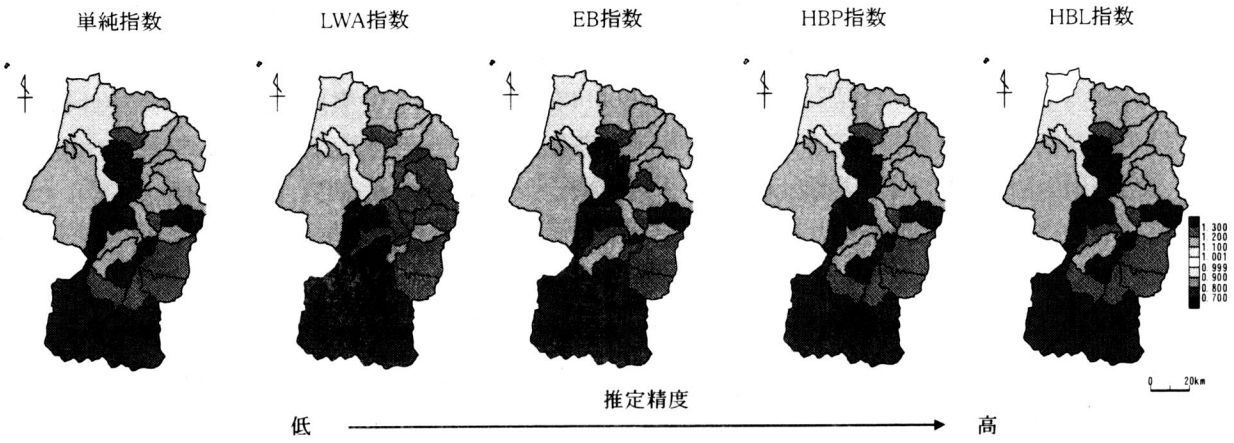


図 26 高血圧，医療費指数の補正結果

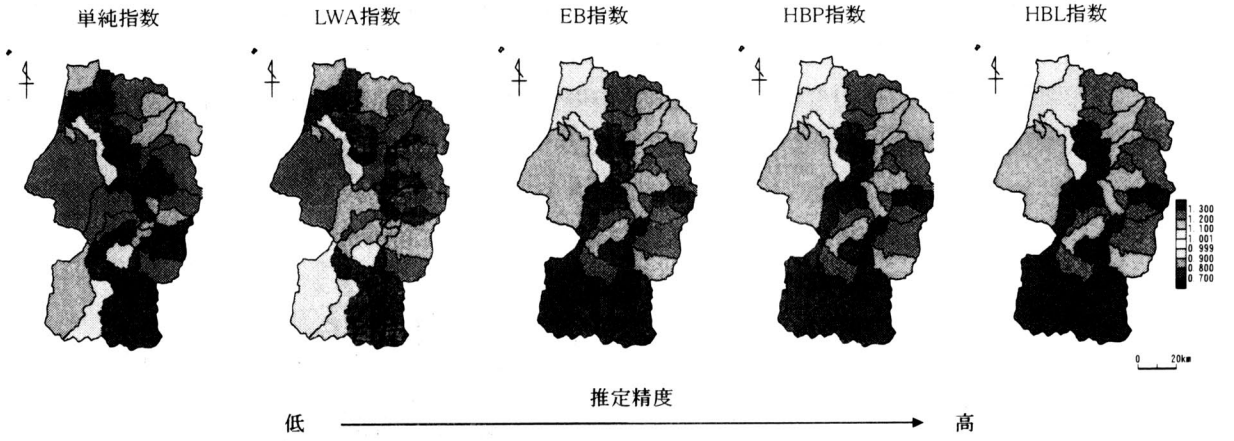


図 27 男性高血圧，医療費指数の補正結果

このように，期待件数が低い指数ほど大きく補正がなされることは，佐伯ほか(2008)による研究結果と同様の結果であった．佐伯ほか(2008)では，死亡率の補正について分析を行っており，結論として疾患によらず経験ベイズ法による補正を行うべきと結論付けている．ただし，この研究ではSMRと経験ベイズ法の2つの方法(モーメント法と最尤法)による比較を行っているため，本研究とは比較しているものが異なる．本研究では，佐伯ほか(2008)の研究結果を踏まえ，補正手法はどのようなデータであっても市町村別の地域データであれば用いるべきとの前提に立ち，複数あるSAEの手法のうち，4つの手法[25]を選択し，そのうちいずれを用いるのが適切かを検討することに主眼を置いた．

その結果，平均二乗誤差とDICの値から，期待件数の多い傾向にあった3つのデータ区分(0－39歳高血圧，40－49歳高血圧，50－59歳高血圧)においては，推定精度を重視し，階層ベイズ法を選択することが適切だと考えられる．ただし，これらの区分において，経験ベイズ法を階層ベイズ法の代替手法として用いて補正を行うことも有用だと考えられる．なぜならば，平均二乗誤差の値から，階層ベイズ法による結果から著しく乖離してはいないためである．ただし，期待件数が100－500の場合に

[25] Locally Weighted Average Method，経験ベイズ法，階層ベイズ法によるポアソンガンマモデルとポアソン対数正規モデルの4つである．

経験ベイズ法と階層ベイズ法による補正結果の差が大きくなる傾向がみられたため，そのようなデータを含む場合には，より慎重に手法を選択する必要がある．

また，高血圧，男性高血圧，女性高血圧，60－69 歳高血圧，70 歳以上高血圧の期待件数の大きい 5 つのデータ区分では，補正の必要性があまり高くはないため，LWA あるいは経験ベイズ法による補正手法を代替的に用いることは妥当だと考えられる．つまり，精度が最も良く，複雑なモデルや統計ソフトウェアを用いる階層ベイズ法を用いなくとも良いという結論に至った．上記のデータ区分においては LWA あるいは経験ベイズ法による推定結果の精度もある程度保たれているものと考えられる．

つまり，期待件数が大きなデータ区分においては，手法の簡便性を重視し，反対に期待件数の少ないデータ区分においては精度を重視した，手法選択を行うべきということが確認された．

以上の結果を簡略化し，手法の精度と複雑さによる位置づけを概念的に示した図にデータ区分ごとの補正推奨度を加えると以下のような結果となる．データ区分については縦軸の補正の推奨度によって位置づけを行い，各種手法については縦軸が推定精度，横軸が手法の複雑さにより位置づけを行った．期待件数が少なく，補正の必要性が高い 3 つのデータ区分については階層ベイズ法による 2 つのモデルで補正を行うことが推奨される．また，高血圧については LWA による補正でも良いと考えられるが，期待件数のばらつきが大きいということを考慮に入れ，経験ベイズ法による補正がより適切と考えた．残りの 4 つのデータ区分では，男性高血圧，60－69 歳高血圧，70 歳以上高血圧，女性高血圧の順に補正の必要性が高いものと考えられる．

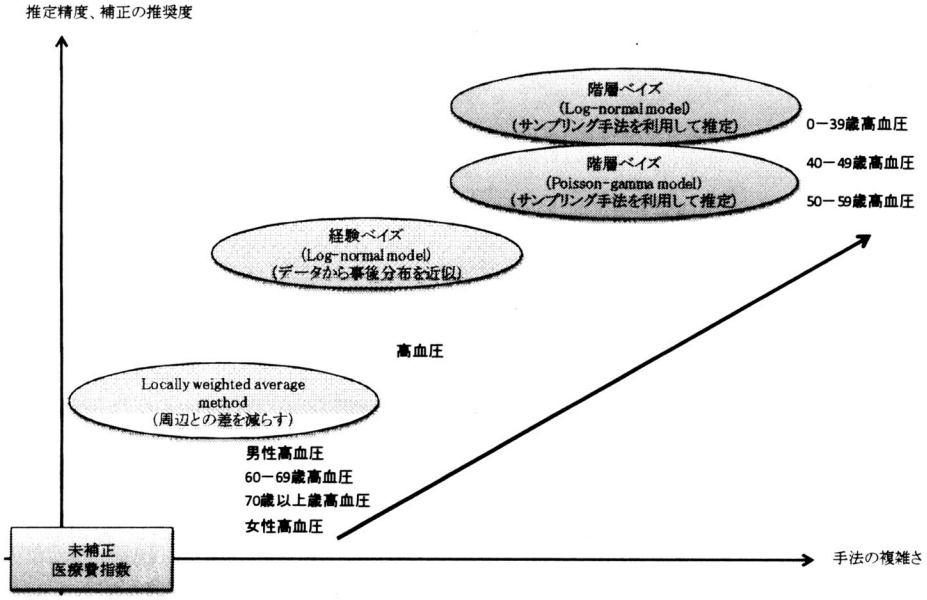


図 28 手法とデータ区分の位置づけ

ただし，Aylin et al. (1999)が述べているように，正確さという点では経験ベイズ法は階層ベイズ法に劣ることが指摘されている．また，本研究の分析結果においても，データ区分によっては，階層ベイズ法と経験ベイズ法による補正結果に差が生じていることも事実であり，期待件数が低いデータセットの場合には注意が必要であることもわかった．特に，期待件数が 100－500 の場合に，階層ベイズ法と経験ベイズ法の補正結果の差が大きくなる傾向が確認されたため，このようなデータ区分において常

に経験ベイズ法を用いることは，精度を重視する場合に問題となる可能性がある．Devine et. al(1994)が指摘しているように，経験ベイズ法は過度に収束しすぎる(Overshrinkage)ことが欠点とされており，推定値の分布が真のリスクの分布よりも狭い傾向にあるとされている．このような欠点から，精度という観点から経験ベイズ法よりも階層ベイズ法を含むフルベイズの手法の方が優れた手法であるとの議論がなされている(C.MacNab et al(2004))．

したがって，医療費適正化計画の策定にあたって医療費指数を改良する場合には，100－500程度の期待件数を含むデータ区分の場合には，可能であれば階層ベイズ法を用いるべきであるが，代替手段として経験ベイズ法を用いることも可能だといえる．

なお，期待件数が低い場合に，補正の必要性が高いものと考えられるが，具体的に件数がいくつ以下の場合に補正が必要といえるのかについては明らかでない．本研究においては，補正差が絶対値で0．5以上の指数は期待件数が500以下となる傾向がみられたが，この点については別途詳細な研究が必要であろう．

この点について，McEwin and Elazar(2006)は，オーストラリアにおける公的な統計指標が必要となる人口規模は10,000－200,000人程度だと述べ，これらの人口規模においてSAEの問題がしばしば生じると述べている．しかし，罹患率のあまり高くない疾患データを分析するような場合には，データ区分の人口規模が10,000－200,000人以上であったとしても補正の必要が生じる可能性もある．つまり，分析地域は小さくなくともデータの出現率が低い区分を分析する場合には，SAE の問題を生じる可能性があり，注意が必要である．ここでいう，データの出現率が低いデータとは，罹患率の非常に低い疾患の医療費データや，他の年齢層では罹患率が高いが分析対象となる年齢階層においては罹患率が低いような区分の医療費データを指している．

また，期待件数が低いとき，被保険者数も少ない傾向にあるといえるが，医療費データを扱う場合には，これはすべてに当てはまるとはいえない．例えば，データ区分を年齢階層別に区切った場合に，高血圧に対する罹患率があまり高くないと考えられる 0－39 歳，40－49 歳等の年齢階層では，被保険者数の人口規模が多い市町村であっても，罹患している患者は少ないために受診件数は低いという状況が生じる．反対に，疾患別・性別・年齢階層別に医療費データを分けて考えた場合，高齢になるにつれて罹患しやすい疾患であれば，人口規模がある程度小さな市町村であっても罹患している患者が多ければSAEの問題は生じない可能性がある．

したがって，医療費データの補正を行う場合には，人口規模のみで SAE の誤差の生じ方は決定されない可能性がある．データあるいは疾病の発生頻度によって，SAE の問題の起こりやすさは異なると考えられ，杓子定規に人口規模のみで補正の必要性を決めることはできない．分析を行う際には，対象データの期待件数がどの程度であるのか確認する必要があるといえよう．また，分析結果より期待件数以外の補正の必要度を決定する要因として，単純指数(未補正指数)の標準偏差とはずれ値の有無があげられよう．これら 2 つの要因は 0-39 歳，40－49 歳，50－59 歳，高血圧のデータ区分より観察された．結果として，補正の必要度を決定する要因は①期待件数の小ささ，②単純指数(未補正指数)の標準偏差と③単純指数(未補正指数)のはずれ値の有無の 3 つである．

そして，①人口規模の小さな地域を分析単位とする場合，②発生頻度の少ない疾患データを分析する場合，③任意の疾患に対して発生頻度の少ない年齢階層のデータを分析する場合の 3 つについて，SAE の問題を考慮に入れて推定を行うべきと考えられる．

これら 3 つの場合は，すべて結果的に期待件数が少ないという特徴を有するが，この特徴を有する

区分は，具体的に期待件数がいくつ以下の場合に生じるものなのかを明らかにする必要がある．疾患別に人口規模あるいは期待件数がどの程度であれば SAE の問題を生じるのかを明らかにすることが期待される．この点については更なる研究が必要であり，今後の課題だといえる．

また，補正手法を用いることにより，実績件数がゼロである指数を算出することができた．このように実績件数がゼロとなるのは，分析データが平成19年度5月診療分と単月データであるため，該当する月に0－39歳に該当する高血圧疾患患者の受診がなかったということを示している．高血圧疾患は高齢であるほど罹患率が高いと考えられるため，0－39 歳の年齢階層でそのような市町村が生じることは当然の結果だともいえる．

補正を行っていない単純指数の場合，実績件数がゼロとなるとき指数自体もゼロとなってしまい，該当市町村の医療費の高低を評価することが出来ないという問題が生じていた．しかし，補正手法を用いることで指数の算出が可能であることがわかった．また，これは 4 つ手法すべてにおいて可能であることが確認された．現状では，電子化されているレセプトデータは 5 月診療分のみであるため，データ区分を細かくする，あるいは分析地域を市区町村等に設定する場合に，このような問題はたびたび生じ得るものと考えられる．したがって，実績件数がゼロとなる場合に，指数の算出を行うために補正手法を用いることは有用だと考えられる．ただし，手法によって算出される値は異なる結果となっているため，いずれの手法を用いるべきかについて別途検討が必要であろう．

以上より，期待件数が小さいときは精度を重視した階層ベイズ法によるポアソンガンマモデルが適切であるが，経験ベイズ法による代替的な補正も意味を持つことが示唆された．さらに，反対に期待件数が大きいときは，LWA あるいは経験ベイズ法による代替が可能であることがわかった．また，すべてのデータ区分において，階層ベイズ法のポアソンガンマモデルとポアソン対数正規モデルでは，ほぼ同じ値に補正される．DIC の値からは，期待件数の少ない傾向を示すデータ区分では，特にポアソンガンマモデルの方は当てはまりが良いという結果が示唆された．

手法の利用可能性という観点から，政策的にはより単純な手法であるほど良いと考えられる．したがって，常に精度を重視した手法を用いるべきとはいえない．期待件数の大きなデータ区分の場合には，補正結果があまり変わらないのであれば，精度のみを重視した手法を使わなくてもよい．ただし，今後ベイジアンの方法がもっと分かりやすく，ベイズ統計に慣れ親しんだ人でなくとも分析できるくらい簡単なソフトウェアが開発されたならば，最も推定精度の高いとされる階層ベイズを選択されるべきだといえる．

最後に，SAE の問題はノイズが入っている以上，それを取り除く策が何かしら必要だと考えられる．しかし，真の値がわからない以上，推定精度を評価することが非常に難しい．先行研究においても，この点は指摘されている点であり，今後改善が期待される．

5.2. 適正化重点対象地域の選定

5.2.1. 結果

手法選択のための分析から，補正の必要性が非常に高いことが示唆されたデータ区分 0－39 歳高血圧を用いて医療費適正上重要な地域を選定する．医療費適正上重要な地域とは，医療費が著しく高いあるいは低い地域を指し，標準偏差分類という手法を用いて地域の選定を行う．標準偏差分類は，指数の平均値から著しく乖離する地域を選定する手法である．算出された指数の平均値から標準偏差×±2 以上の値を異常値と仮定し，検出することとする．

標準偏差分類により，重要地域の検出を行った結果が表 15 である．表中では，高医療費圏となる＋2σ値を赤，低医療費圏となる－2σ値を緑で表示している．さらに，その結果を地図上に表現したものが図 29 である．

表 15 より，補正に用いた手法により，異常値として検出される地域が全く異なっていることがわかる．標準偏差という同一の尺度にもとづいて重要地域の選定を行ったとしても，補正手法により選定される地域が全く異なっている．政策的に重要な地域を選定する際には，はじめにどの手法を用いて補正を行うかが非常に重要な選択であることを示唆する結果となった．

具体的に，単純指数による重要地域として，平均値から標準偏差2つ分の値以上を示したのは大蔵村の 3.22 となり，高医療費圏として選定された．大蔵村は＋2σ(標準偏差)以上の値として検出されているが，一方の－2σの値は負値となったことから検出がなされなかった．

LWA による補正結果からは，上山市の 1.70 が高医療費圏として検出された．－2σの値は正値であったが，それ以下となる指数が存在しなかったため，低医療費圏の検出はされなかった．

経験ベイズ法による補正結果からは，低医療費圏として長井市 0.46，川西町 0.37 という値が検出された．高医療費圏は検出されなかった．

階層ベイズ法のポアソンガンマモデルとポアソン対数正規モデルでは，高医療費圏と低医療費圏共に同様の地域が検出された．高医療費圏としては山形市の 1.36(ポアソンガンマモデル)と 1.22(ポアソン対数正規モデル)が検出され，低医療費圏としては鶴岡市の 0.70(ポアソンガンマモデル)と 0.87(ポアソン対数正規モデル)が検出された．

図 29 の結果より，手法の精度が向上するにつれて指数の高低のグラデーションがより鮮明となっている．また，視覚的に判断される重要そうだと考えられる地域と指数の値から選定される地域とが一致するようになっていることがわかる．

ただし，今回の重要地域の選定では，全市町村の指数の平均値から著しく乖離する地域を選定するという標準偏差分類という手法を用いたが，手法によっては±2σの値が負値となるものが生じたため，高医療費圏あるいは低医療費圏のみの選定しか行われなかった．したがって，地域の選定にあたっては，さらに優れた手法を用いるなど適切な地域の選定方法を用いる必要があるといえる．

表 15 標準偏差分類による重要地域の選定(0－39 歳高血圧)

	単純指数	Locally weighted average	Empirical beyes	Hierarchical bayes	
			Poisson-gamma model(モーメント法)	Poisson-gamma model	Log-normal model
山形市	1.48	1.32	1.36	1.36	1.22
米沢市	1.56	1.34	1.56	1.31	1.17
鶴岡市	0.53	0.64	0.53	0.70	0.87
酒田市	0.97	1.00	0.97	0.99	1.01
新庄市	0.72	0.96	0.88	0.90	0.97
寒河江市	0.31	0.70	0.93	0.77	0.91
上山市	1.54	1.70	1.23	1.20	1.11
村山市	0.48	0.85	1.02	0.87	0.96
長井市	0.46	0.78	0.46	0.86	0.96
天童市	0.76	0.92	1.00	0.90	0.97
東根市	0.82	0.73	1.05	0.95	0.99
尾花沢市	0.75	0.91	1.07	0.95	1.00
南陽市	0.87	0.82	0.87	0.98	1.00
山辺町	1.60	1.61	1.18	1.12	1.07
中山町	1.15	1.38	1.14	1.05	1.04
河北町	1.10	0.85	1.14	1.05	1.03
西川町	0.00	0.92	1.10	0.97	1.01
朝日町	1.98	0.73	1.20	1.16	1.09
大江町	0.00	1.55	1.06	0.91	0.98
大石田町	2.52	0.85	1.25	1.23	1.12
金山町	0.64	0.84	1.02	0.99	1.01
最上町	2.03	0.94	1.38	1.22	1.12
舟形町	0.00	1.57	0.94	0.93	0.99
真室川町	1.90	1.53	1.32	1.18	1.10
大蔵村	3.22	0.57	1.40	1.22	1.12
鮭川村	1.73	1.02	1.21	1.11	1.06
戸沢村	0.00	1.06	0.93	0.92	0.99
高畠町	1.31	1.08	1.31	1.11	1.07
川西町	0.37	1.25	0.37	0.89	0.97
小国町	1.23	1.18	1.23	1.05	1.04
白鷹町	1.62	1.02	1.62	1.15	1.08
飯豊町	0.89	1.21	0.89	1.02	1.03
三川町	1.54	0.75	1.54	1.09	1.05
庄内町	0.99	0.77	0.99	1.02	1.02
遊佐町	0.81	0.92	0.81	0.99	1.01
標準偏差	0.74	0.30	0.28	0.15	0.07
平均	1.08	1.04	1.09	1.03	1.03
＋2σ値	2.56	1.64	1.65	1.33	1.17
－2σ値	-0.39	0.43	0.52	0.73	0.89

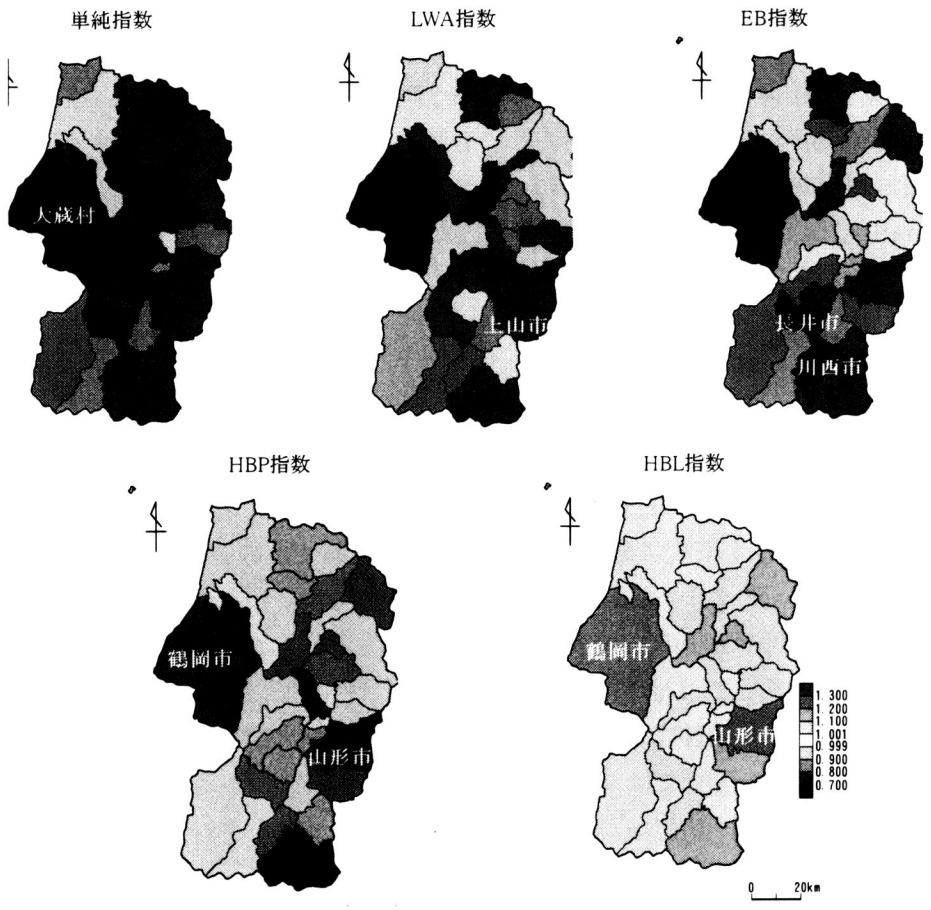


図 29 各手法による重要地域の選定結果(0－39 歳, 高血圧)

5.2.2. 考察

期待件数が大きく, 補正の度合が大きかった 0－39 歳高血圧のデータ区分を用いて医療費適正化からみた重要地域の選定を行った結果, 補正手法により選定される地域が大きく異なることが確認された. この結果から, 政策的に重要な地域を選定する際に, 特に SAE の問題を生じやすいと考えられるデータ区分に対しては精度の高い手法を用いて補正を行うべきということがあらためて示唆された.

今回, 重要な地域を選定する際には標準偏差分類による方法を用いたが, 単純指数の－2σ 値では負値となるなど, すべてのデータに対して有効ではないことが確認されたため, 重要地域をどのように選定するのかについては, 別途詳細な研究が必要であろう.

6. 研究のまとめと課題

本研究では，医療費適正化計画の策定にともない，策定主体者である国や都道府県は，地域の医療費の利用水準を相対的に把握するための詳細な医療費分析が求められている．地域の医療費の利用水準を相対的かつ詳細に把握する一つの有力な方法として，医療費指数を性・年齢・疾病別に算出し，それを地域単位に分析することがあげられる．しかし，医療費指数を細分化した医療費で算出すると，Small Area Estiomation(SAE)の問題が生じてしまう．SAE とは，分析地域の人口が少ない，あるいは分析単位が小さいために誤差変動が大きくなり，精度の高い推定結果を得られない問題である．

もちろん，これまでに SAE の問題を扱った研究はなされている．その代表例が市区町村別の死亡率等を経験ベイズ法により補正を行う研究であるが，医療費の分析では未だ実施されていない．また，補正手法には，経験ベイズ法を用いた方法以外にもいくつかあるが，適切な手法選択については十分な議論がなされていない．一方，国外の研究では，精度を高めることを念頭に複数の補正手法が開発された結果，補正手法が高度に発達する一方で，精度と簡便性の間にトレードオフの関係が生じることが知られている．

そこで本研究では，次の 2 つのことを目的として研究を実施した．第 1 はベイズ統計を用いて医療費指数の改良を行うこと，第 2 は小地域推定の方法選択に関する知見を獲得することである．研究目的に対応して，1）補正手法の選択に関する分析と，2）適正化重点対象地域の選定に関する分析の 2 つを実施した．なお，実際の分析では，山形県の高血圧疾患のデータを用いて分析した．下記に，その主要結果を記す．

1）補正手法の選択に関する分析のまとめ

既存研究のレビューから，医療費指数の改良には次の4つの手法が有用であることがわかった．① Locally Weighted Average(LWA)，②経験ベイズ法(EB)，階層ベイズ法(HB)の③ポアソンガンマモデルと④ポアソン対数正規モデルである．そこで，この 4 つの方法を山形県の高血圧疾患の受診件数データに適用し，補正の精度を実証的に分析した．実際の分析では，高血圧全体，性別の 2 種類，年齢階層の 5 種類（0－39 歳，40－49 歳，50－59 歳，60—60 歳，70 歳以上）の合計 8 種類のデータ区分に，4つの手法を適用した．分析の目的は，データ区分(性・年齢区分)ごとに適切な補正手法と重視すべき点(精度あるいは簡便性のどちらか)を明らかにし，さらに補正前と補正後で選択される適正化重点対象地域に違いがないかを比較することである．

分析の結果，高血圧の医療費データでは，データ区分によっては，学術的に支持されている精度の高い手法の代わりに簡便な補正手法を代用したとしても，一定の推定精度を保つことが可能であることが示唆された．具体的には，発生頻度の大きい年齢階層のデータ(高血圧，男性高血圧，女性高血圧，60—60 歳高血圧，70 歳以上高血圧)では，簡便な手法である LWA，経験ベイズ法で問題がないことがわかった．一方，①人口規模の小さな地域のデータ(期待件数がおおむね 500 件以下)，②発生頻度の少ない年齢階層のデータ(0－39 歳，40－49 歳，50－59 歳)を補正する場合には，より精度の高い方法(階層ベイズ法，経験ベイズ法)が必要である．また，補正の必要度を決定する要因は，①期待件数の小ささ，②単純指数(未補正指数)の標準偏差，③単純指数(未補正指数)のはずれ値の有無の 3 つである．最後に，補正前と補正後では，適正化重点対象地域が大きく異なることを確かめ，指数の補正を行う必要があることを実証的に示した．

また，0－39 歳高血圧のデータ区分のように期待件数が低い傾向にある市町村では，実績件数がゼロとなる市町村がいくつか存在し，このような市町村データの指数を算出するに際して SAE の手法が有用であることが確かめられた．補正をしなければ，分子がゼロになってしまうため，指数自体もゼロとなってしまう．しかし，本研究で用いた補正を施せば，実績件数がゼロにはならないため，この問題を解決することができる．なお，少し統計的に込み入った話になるが，経験ベイズ法の場合には，期待件数が少ないときに補正式中の ϕ 二乗が負値となり，ゼロで打ち切りとなる市町村がいくつかみられた．この場合には指数の補正は一切なされないため，この点がモーメント法を用いた経験ベイズ法の欠点だと考えられる．

SAE の問題に対する方法は複数あり，実際に政策の現場で用いる場合にはトレードオフの関係にある推定精度と手法の簡便性の双方のバランスをとる必要がある．残念ながらわが国の既存研究では，それらの議論があまりなされぬまま，特定の手法のみが用いられている現状にあった．本研究の最大の成果の一つは，図 30 に示すように，データの区分ごとに適切な補正方法を明らかにした点にある．データ区分や地域の人口規模が小さい場合には精度を重視して SAE の手法を選択すべきであり，またデータ区分や人口規模があまり小さくない場合には，簡便性を重視して手法を選択しても補正結果の推定精度にはある程度保たれることを確認した．

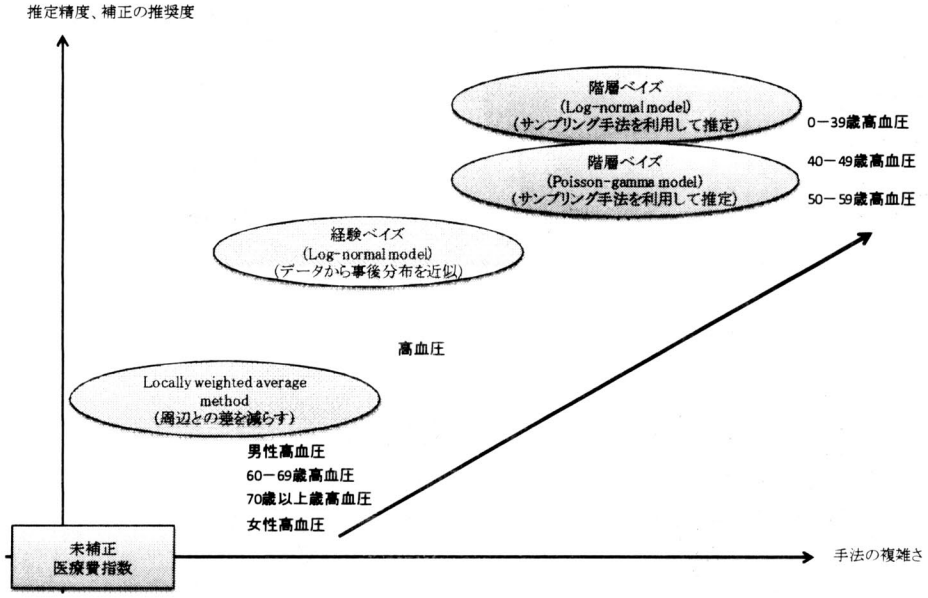


図 30　手法とデータ区分の位置づけ(再掲)

2)適正化重点対象地域の選定に関する分析のまとめ

1)補正手法の選択に関する分析から，0－39 歳高血圧は，階層ベイズ法による精度の高い補正が必要であることがわかった．そこで実際に補正前と補正後で，医療費適正化上重要とされる地域に違いが生じるかを確認した．地域の選定にあたっては，標準偏差分類という手法を用いて全市町村の指数の平均値より突出して高いあるいは低い地域を選定した．また，補正前と各手法による補正後の指数をマッピングし，結果の違いを考察した(図 31)．0－39 歳という高血圧疾患の発生頻度が低いと考えられるデータ区分では，各手法によって指数の値が大きく異なるのと同時に，手法の精度が向上す

るほど指数の高低に関する地理的な分布がより鮮明なグラデーションとなることが確認された．また，結果として選定された地域は，補正手法によって大きくことなっていることが確認されたため，補正の必要性が高いデータ区分では精度を重視した手法選択が求められることをあらためて確認する結果となった．

今後の課題としては，SAE の問題は具体的に分析データの人口に対して期待される受診件数がいくつ以下の場合に生じる可能性があるのか，疾患別に明らかにすることがあげられる．また，今回は受診件数を用いた医療費指数の補正を行ったため，他のの医療費の要素である点数と日数についても SAE の補正を行うことと，他県データでも同様の分析を行うことがあげられる．さらに，補正手法については時系列データを用いた SAE の補正方法も提案されている．医療費についての時系列データを有する場合には，それを補助データとして用いることで補正した結果と他の手法による補正結果とを比較することにより，どちらが望ましいと考えられるのかを検討することが期待される．

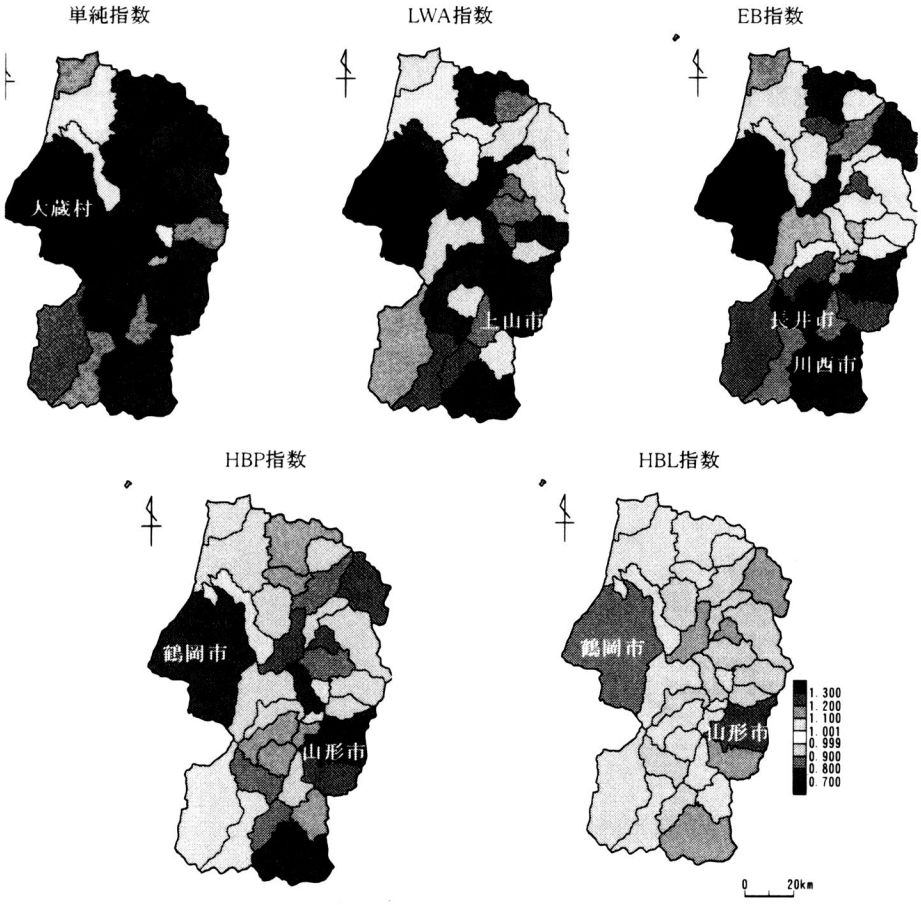


図 31 各手法による重要地域の選定結果(0－39 歳，高血圧)(再掲)

参考文献

相田潤・安藤雄一・青山旬・丹後哲郎・森田学 (2004)「経験的ベイズ推定値を用いた市町村別3歳児う蝕有病率の地域比較および歯科保健水準との関連」『口腔衛生会誌』54, 566-576

上畑鉄之丞 (2003)『根拠に基づく健康政策のすすめ方 政策疫学の理論と実際』医学書院

久保川達也 (2006)「線形混合モデルと小地域の推定」『応用統計学』, 35, 3, 1-24

慶應義塾大学小暮研究会II (2007)『WinBUGS 入門 ～導入・使用法・実例～』

厚生労働省 (2006)『平成18年度国民医療費の概況』

厚生労働省 (2007)『平成19年度版 厚生労働白書』

厚生労働省 (2008)『平成20年度版 厚生労働白書』

厚生労働省 (2007)『医療費適正化に関する施策についての基本的な方針(案)』

古城隆雄・印南一路 (2009)「医療費の地域差要因と適正化重点対象グループの発見」『国,都道府県の医療費適正化計画の重点対象の発見に関する研究』, 11-64

小西貞則・超智義道・大森裕浩 (2008)『計算統計学の方法』朝倉書店

佐伯圭吾・平子哲夫・中田正 (1999)「人口動態市区町村別統計へのベイズ統計の応用について(2)合計特殊出生率への応用」『厚生の指標』第46巻第11号, 3-10

佐伯圭吾・岡本希・森田徳子・車谷典男 (2005)「SMRの経験的ベイズ推定量についての検討－奈良県市町村別死因統計を用いて－」『厚生の指標』第52巻第11号, 7-13

高橋邦彦・横山徹爾・丹後俊郎 (2008)「疾病地図から疾病集積性へ」『Public Health』57

丹後俊郎 (2000)『統計モデル入門』朝倉書店

丹後俊郎・横山徹爾・高橋邦彦 (2007)『空間疫学への招待 疾病地図と疾病集積性を中心として』朝倉書店

中田正・齊藤重正・六車史 (2008)「最近のベイズ推定研究の小地域の人口動態指標推定への応用の研究」『厚生の指標』第55巻第5号, 1-11

中谷友樹・谷村晋・二瓶直子・堀越洋一 (2004)『保健医療のためのGIS』古今書院

平子哲夫・佐伯則英・中田正 (1999)「人口動態市区町村別統計へのベイズ統計への応用について(1)標準化死亡比への応用」『厚生の指標』,第46巻第10号, 3-11

古谷知之 (2008)『ベイズ統計データ分析-R&WinBUGS-』朝倉書店

水嶋春朔 (2000)『地域診断のすすめ方 根拠に基づく生活習慣病対策と評価』医学書院

村山祐司・柴崎亮介 (2008)『GISの理論』朝倉書店

GIS by ESRI (2000)『ArcMapユーザーズ・ガイド』

Bernardinelli, L., D.Clayton and C.Montomoli (1995) "Bayesian Estimates of Disease Maps: How important are Priors?"

Bernardinelli, L., and C.Montomoli (1992) "Empirical Bayes versus Fully Bayesian

ANALYSIS Of Geographical Variation In Disease Risk", *Statistics In Medicine,* VOL. 11, 983－1007

Best, N., S.Richardson, P.Clarke, and V.Gomez-Rubio (2008) "A Comparison of model-based methods for Small Area Estimation"

Best, N., S.Richardson, and A.Thomson (2005) "A comparison of Bayesian spatial models for diseasemapping", *StatisticalMethods in Medical Research*, 14, 35-59

Brackstone, J.G. (2002) "Strategies and Approaches for Small Area Statistics", *Statistics Canada*, Vol. 28, No. 2, 117-123

Carlin, P., and T. A.Louis (2000) *Bayes and Empirical bayes methods for data analysis*, Champan and Hall

Carlin, P.B., and T. A.Louis (2009) *Bayesian Methods for Data Analysis*, Champan and Hall

Casella, G. (1985) "An Introduction to Empirical Bayes Data Analysis", *The American Statistician*, Vol. 39, No. 2, 83-87

Clayton, D., and J.Kaldor (1987) "Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping", *International Biometric Society*, Vol. 43, No. 3, 671-681

Devine, O.J., and Louis TA (1994) "A constrained empirical Bayes estimator for incidence rates in areas with small populations", *Statistics in Medicine*, Vol.13, 1119-1133

Ghosh, M., and J.N.K.Rao (1994) "Small Area Estimation: An Appraisal", *Statistical Science,* Vol. 9, No.1, 55-76

G´omez-Rubio, V., and A. L´opez-Qu´ılez (2006) "Empirical and Full Bayes estimators for disease mapping"

G´omez-Rubio, V., N.Best, and S.Richardson (2009) "Bayesian Statistics for Small Area Estimation"

Gomez-Rubio, V. (2009) "Approximate Bayesian Inference for Small Area Estimation"

Gilks, R.W., S.Richardson and D.J.Spiegelhalter (1996) *Markov Chain Monte Carlo in Practice,* Champan and Hall

Heisterkamp, S.H., G. Doornbos, and M.Gankema (1993) "Disease Mapping Using Empirical Bayes and Bayes Methods on Mortality Statistics in the Netherlands", *Statistics in Medicine*, Vol.12, 1895-1913

Kafader, K. (1994) "Choosing among two-dimensional smoothers in practice", *Computational Statistics & Data Analysis*, Vol18, 4, 419 – 439

Lawson, A.B (2009) *Bayesian Disease Mapping*, Champan and Hall

Lawson, A.B (2009) *Statistical methods in spatial epidemiology*, Wiley

Lawson, A.B, Browne, William J and Carmen L.Vidal Rodeiro (2003) *Disase Mapping with WinBUGS and MLwiN*, Wiley

Lawson, A.B, A.Biggeri, D.Bohning, E.Lesaffre, J-F.Viel, and R.Bertollini (1999), *Disease Mapping and Risk Assessment,* Wiley

MacNab, C.Y. (2003) "Hierarchical Bayesian Modeling of Spatially Correlated Health Service Outcome and Utilization Rates", *Biometrics*, 59, 305–316

MacNab, C.Y., P.J.Farrell,P.Gustafson,and S.Wen (2004)," Estimation in Bayesian Disease Mapping", *Biometrics,* 60, 865–873

Maiti, T. (1998) "Hierarchical Bayes estimation of mortality rates for disease mapping", *Journal of Statistical Planning and Inference*, 69, 339-348

Martuzzi, M., and P. Elliott (1996) "Empirical Bayes Estimation of Small Area Prevalence of Non-rare Conditions", *Statistics in Medicine*, Vol.15, 1867-1873

McEwin, M., and D.Elazar (2006) "Regional Statistics : Small Area Estimation in Offcial Statistics"

Meza, L.J. (2003) "Empirical Bayes estimation smoothing of relative risks in disease mapping", *Journal of Statistical Planning and Inference* 112, 43 – 62

Mollie, A., and S.Richardson (1991) "Empirical Bayes Estimates of Cancer Mortality Rates Using Spatial Models", *Statistical in Medicine*, VOL.10, 95-112

SAE package developers (2007) "Introduction to Small Area Estimation"

Marshall, J.R. (1991) "Mapping Disease and Mortality Rates Using Empirical Bayes Estimators", *Journal of the Royal Statistical Society*, Vol. 40, No. 2, 283-294

Morris, C.N. (1983) "Parametric Empirical Bayes Inference: Theory and Applications", *Journal of the American Statistical Association*, Vol. 78, No. 381, 47-55

Owen J. Devine, Thomas A. Louis, and M. Elizabeth Halloran (1994) "Empirical Bayes Methods for Stabilizing Incidence Rates before Mapping", *Epidemiology*, Vol. 5, No. 6, 622-630

Rao, J.N.K. (2003) *Small area estimation*, Wiley

Salvati, C.N.,and C.Ray (2009) "Small Area Estimation for Spatially Correlated Populations-A Comparison of Direct Indirect Model-Based Methods-"

Saei, A., and R.chambers (2003) "Small area estimation:A review of Methods Based on the Aplication of Mixed Models", *S3RI Methodology Working Paper M03/16*

Satorra Albert and Ventura Eva (2006) "Small- Area Estimation at Idescat: Current and

Related Reserch"

Saunderson, R.T., and I.H. Langford (1996) "A Study of the Geographical Distribution of Suicide Rates in England And Wales 1989-92 Using Empirical Bayes Estimates", *Social Science & Medicine*, Vol.43, No.4,498-502

Shi, X., E.Duell, E.Demidenko, T.Onega, B.Wilson, and D.Hoftiezer (2007) "A polygon-based locally-weighted-average method for smoothing disease rates of small units" , *Epidemiology*,Vol.18, No.5, 523-528

Singh,A.C., Stukel, D.M., and Pfeffermann,D. (1998) "Bayesian versus frequentist measures of error in small area estimation", *Royal Statistical Society*, 60, 2, 377-396

Trevisani, M., and N.Torelli (2002) "Comparing hierarchical Baysian models for Small Area Estiomation"

Trevisani, M., and N.Torelli (2004) "Small area estimation by hierarchical bayesian models: Some practical and theoretical issues", *Atti della XLII Riunione Scientifica della Societ`a Italiana di Statistica*, 273–276

Ugarte, M. D., T.Goicoa, B.Ib´anez and A. F. Militino (2009) "Evaluating the Performance of Spatio-Temporal Bayesian Models in Disease Mapping", *Environmetrics*

Waller, A.L., and C.A.Gotway (2004) *Applied Spatial Statistics for Public Health Data*, Wiley

Yasui, Y., H.Liu, J.Benach, and M.Winget (2000) "An empirical evaluation of various priors in the empirical Bayes estimation of small area disease risks", *Statistics in Medicine*, 19, 2409-2420

Zhou, M.Q., and Y.You (2008) "Hierarchical Bayes Small Area Estimation for the Canadian Community Health Survey"

謝辞

はじめに，本研究は平成 21 年度厚生労働科学研究「国，都道府県の医療費適正化計画の重点対象の発見に関する研究」の協力自治体から提供されたデータを利用させて頂きました．貴重なデータをご提供頂いた協力自治体に心より御礼を申し上げます．

ここからは，本研究を行うにあたり，ご指導ならびにご協力いただいた方々に対し，この場を借りて御礼を申し上げます．全ての方々のお名前をあげることはできませんが，ご容赦頂ければ幸いです．

誰よりもまず，SFC に入学後，学部 2 年生の頃より計 5 年間のご指導を頂いた慶應義塾大学総合政策学部の印南一路教授に心からの御礼を申し上げます．印南先生なくして，今の私はあり得なかったと思うほど，印南先生の存在や発言は私にとって大きなものでした．先生には，学問や研究に関してのご指導を頂きましたが，それ以上に物事に対する考え方や生き方を教わったように思います．「努力するプロセス自体を楽しみなさい」という先生の言葉や，「一番を目指すのではなく自分がやりたいことで進路を選択しなさい」という先生のアドバイスは，進路について悩み，浪人していたことで自信をなくしていた私が徐々に自分を取り戻すきっかけとなった言葉でした．先生から頂いた様々な言葉が，私が大学で得た最大の教訓だと思っています．これまで印南先生とめぐり合い，先生の下で勉強できたことに心から感謝しています．これから，また新たな場所で精一杯努力するつもりですが，先生の言葉を節々に思い返して，胸を張って近況をご報告できるよう，そのプロセス自体を楽しみながら精一杯目標に向かって走りたいと思います．改めて，この場を借りて印南先生に心からの感謝をお伝えしたいと思います．本当にありがとうございました．

次に，感謝の気持ちをお伝えしなければならないのは，同大学総合政策学部の小暮厚之教授です．小暮先生には，大学院に入学後様々な場面でご指導を頂きました．幾度となく研究や統計についてのご指導を頂き，未熟な私にベイズ統計に関する書籍の翻訳という機会も与えて下さいました．小暮先生の存在がなければ，統計の面白さを知ることもなく，修士論文を完成させることも出来なかったと思います．また，就職に際しては，「黒島さんは真面目だから日本銀行がいいと思いますよ」とアドバイスを頂き，私はそれを鵜呑みにして出願しました．小暮先生から金融の世界の面白さを教えて頂き，また励ましのお言葉を頂かなければ，このような結果には決して至らなかったと思います．大学院の面接で初めてお会いした時から， 先生の温和なお人柄に惹かれる一方で，難しい統計の概念を目から鱗が出るほど分かりやすく説明なさってくれる先生に尊敬と憧れの念を抱いていました．先生のようになりたいと何度も思い，必死に机にかじりついたのを覚えています．SFC の大学院に入学し，小暮先生とお会い出来たこと，そして統計学の一端を小暮先生から学べたことに心から感謝の気持ちでいっぱいです．この場を借りて，改めて感謝の気持ちをお伝えしたいと思います．本当にありがとうございました．

つづいて，もう一人の副査である同大学総合政策学部の古谷知之准教授にもお礼を申し上げます．先生には，ベイズ統計や空間モデリング，またデータマイニングといった，統計の知識の土台を形成すべく素晴らしい授業をして頂きました．先生の授業を受けることがなければ，本研究の内容が現在のように定まることもなかったと思います．まだまだ現段階でも未熟であることは否めませんが，今のレベルまで到達することもなかったと思います．SFC の中でも，先生の授業は他の授業に比べ自分で補わなければならない知識や範囲が非常に多く，必死で勉強したのを覚えています．しかし，その半面授業を終えること

で得られるものが非常に多く，先生の授業はすべて受けたいと思うほどでした．これからも，先生の素晴らしい授業が続くことを個人的に願っています．改めて，古谷先生にも感謝の意をお伝えしたいと思います．本当にありがとうございました．

東海大学の堀真奈美准教授にも，印南研究室の偉大な先輩として，また，医療政策分野の先生としても様々な場面でご指導頂きました．堀先生には海外の医療制度について学ぶ機会を様々な場面で与えて頂き，実際に海外の社会保障政策の現場を見るという貴重な機会も頂きました．堀先生を交えた印南先生方とのミーティングでは，様々な角度からご指摘される堀先生にいつも感動し，自分も多角的な視点から物事を考えられるようになりたいといつも感じていたのを覚えています．先生の有言実行な姿勢と男勝りな働き方は私の憧れです．Warm heart, and Cool head とはまさに堀先生を体現した言葉だとずっと思っていました．これから先，少しでも先生に近づけるよう，ひたむきに努力していきたいと思います．この場を借りて，改めて堀先生に感謝の意をお伝えしたいと思います．本当にありがとうございました．

そして，印南研究室の偉大な先輩としてもう一人，私の SFC での生活で決して忘れてはならない先輩，それは政策・メディア研究科博士課程の古城隆雄先輩です．はじめて印南研に入った学部2年生の頃から，呆れられるほど稚拙な質問と度重なる質問に，何度古城さんの研究を妨げ，怒られたかわかりません．時に，生意気な態度を取ってしまったことも反省しています．しかし，どんなときも未熟な私を見捨てることなく，研究の相談から進路の相談，それから個人的な悩みに至るまで親身に接して下さり，5 年間で私が古城さんから学んだことは図り知れません．SFC で出会った先輩の中で，私が最も影響を受けた先輩，それは古城さん以外にいません．SFC で古城さんという先輩に出会えたこと，そして，一緒に研究することができたこと，これ以上に幸せなことはありませんでした．これからも，大きな夢に向かって努力する心を忘れず，変わらない古城さんの後輩黒島でいたいと思います．この場を借りて改めて感謝の意をお伝えしたいと思います．本当にありがとうございました．

大学院での生活では，家にほとんどいることはなく，大学院棟で朝から晩まで，ほとんどの時間を過ごしていました．その中で出会った先輩や同期の仲間は，かけがえのない存在です．それぞれに異なる専門分野と強い個性に，いつも刺激をもらい，友達や先輩っていいなと改めて他者と関わることの大切さを認識していました．

特に，政策・メディア研究科博士過程の渡辺大輔先輩には，「勉強するな遊べ」という私に対する助言とは裏腹に，毎日趣味で難解な本を読むその姿と何を聞いても答えてくれるその博学さに，いつも尊敬の念を抱いていました．就職に際しても，様々な場面でいつも親身に相談に乗って下さり，力を貸して下さいました．渡辺さんには，心からの感謝の気持ちで一杯です．これからも趣味でお勉強するというその姿勢を見習っていきたいと思います．この場を借りて，改めて感謝の気持ちをお伝えしたいと思います．本当にありがとうございました．

そして，私の学部・大学院での生活を明るく，楽しいものにしてくれたのは，以下にあげる先輩・同期・後輩を含む友達の存在があってこそでした．印南研で仲良くなった両角彩さんや増井貴子さん，高橋怜子さん，門倉由梨さん，研究室は違いましたが仲良くなった吉道さゆりさんや横山夏来さん，彼女たちの存在は学生生活の中で本当に大きかったです．みんなと過ごした時間を思うと，人の幸せは人とのつながりから得られるものなくしては，あり得ないなと心から思います．みんなの存在があったから，研究生活で一人寂しく感じることもありませんでした．本当にありがとうと伝えたいです．

また，大学院に入学後，一年生の頃から仲良くなった修士 2 年の原悠樹くんには，たびたびプログラミングの相談に乗ってもらいました．プログラミングがさっぱりなのに、R を使わなければならなかった私にとって，原くんのプログラミングの指導は，本当に心強かったです．忙しいのにいつも助けてくれて，本当にありがとう．原くんの美味しいご飯と楽しい会話にはいつも癒されました．この場を借りて御礼をお伝えしたいと思います．

つづいて，大学院で小暮研や授業など様々な場面でお世話になったのが，修士 1 年の兎原義弘さんです．兎原さんには，同じベイズ統計を学ぶ学生であったことや同じ授業を取っていることが多かったこともあり，本当にさまざまな場面で色々なことを相談させて頂きました．兎原さんの存在が本当に心強かったです．この場を借りて御礼をお伝えしたいと思います．

そして，修士2年の大場章弘くんには，GISについて貴重なアドバイスを頂きました．また，大学院の博士過程の先輩であった中島多恵子先輩や今村晴彦先輩，荒井悠介先輩，丸山洋平先輩，裵潤先輩にもさまざまな場面でお世話になりました．印南研の後輩である安達くん，切貫くん，福山くん，平村さん，張さん，板垣くん，みんなのおかげで私は楽しい研究会生活を過ごすことができました．安達くんには，集計作業を手伝ってもらったこともありました．

この場を借りて，ここにあげた皆様に改めて御礼を伝えたいです．ありがとうございました．

皆と出会えたおかげで，学生生活が本当に楽しく素晴らしいものになりました．他にも，お世話になった先輩や同期がいますが，ここにすべての方のお名前を書くことはできないことをお許しください．SFC で出会えた素晴らしい学問と友人や先輩に巡り合えたこと，本当に嬉しく思っています．このような場を提供してくれた SFC という素晴らしい学びの場が今後も存在し続けてくれることを心から願っています．

最後に，小さなころから努力することの大切さを教えてくれ，いつも温かく見守ってくれていた母，言葉を交わすことは少ないけれども心ゆくまで勉強できる環境を提供してくれた父，2 人の両親に心からの感謝の意を伝えたいと思います．また，アメリカにいるため，あまり話す機会の少ない父には，この場を借りて，この世に生を授けてくれたことに改めて感謝したいと思います．加えて，努力家な姉と大切な弟にもこの場を借りて，感謝の意を伝えたいと思います．

このように大学生活を振り返ってみれば，人間が一人でできることや学べることはごく僅かであって，多くの人の存在があってはじめて多くのことを学び，得られるのだと思います．今の自分は， 先生や先輩方，両親や祖父母，家族，そして友人の存在があってはじめて成り立ったものだと思います．これから先の長い人生で，私も周囲の人間に良い影響を与えられる人間となれるよう，感謝の気持ちを忘れず，自分なりに精一杯の生きていきたいと思います．改めて，これまで支えて下さったすべての方に御礼を伝えたいと思います．

2010 年 1 月 11 日

黒島テレサ

ベイズ統計を用いた医療費指数の改良
－エビデンスに基づく医療政策に向けて－

2010年 3 月 25 日　　初版発行

著者　黒島テレサ

監修　印南一路

発行　慶應義塾大学 湘南藤沢学会
〒252-0816　神奈川県藤沢市遠藤5322
TEL:0466-49-3437

Printed in Japan　　印刷・製本　　ワキプリントピア

ISBN 978－4－87762－235－0
SFC-MT 2009-004

■本論文は修士論文において優秀と認められ、出版されたものです。