

**特集：地域保健活動における評価の現状と課題****保健活動における評価の現状と課題**

尾島俊之

浜松医科大学健康社会医学講座

**Assessment and Evaluation of Public Health Programs**

Toshiyuki OJIMA

Department of Community Health and Preventive Medicine,  
Hamamatsu University School of Medicine**抄録**

評価には、事前評価と事後評価（地域全体の評価や事業参加者の評価）がある。全国の市町村での実施割合は、何らかの保健事業の評価8割、地域の状況を把握するための既存資料の分析7割等である。住民の健康指標等の把握割合は、死因別死亡者数・各種保健事業参加率等は9割、標準化死亡率（SMR）5割等である。事業評価のデザインとして、ケーススタディーデザイン、前後比較デザイン、準実験デザイン、実験デザイン（無作為比較対照試験）等がある。実験デザインは科学的に最も正しい方法であると言われることが多いが、倫理的な問題や外部妥当性などの欠点がある。現場での評価には前後比較デザインが最もお薦めであるが、事業以外の影響による変化等の問題に注意を要する。事業の改善を目的とした評価においては、質的評価、特に観察が重要である。

**キーワード：** 事前評価、事後評価、前後比較デザイン、実験デザイン、質的評価

**Abstract**

This paper outlines the assessment and evaluation of public health programs. Among municipalities in Japan, 80% have evaluated public health programs and 70% have analyzed existing data on their jurisdictions. The proportions of municipalities that have monitored statistics were 90% for cause-specific mortality and participation rates of various programs, and 50% for standardized mortality ratios. Programs have been evaluated by various types of design, including case study design, before-after design, quasi-experimental design, and experimental design (randomized controlled trials). Although experimental design are usually regarded as the best design, this approach has limitations including ethical difficulties and poor generalizability. Here, the adoption of the before-after design is recommended for practical evaluation of programs, paying attention to outside effects other than the programs themselves. In conducting evaluation to improve programs, qualitative methods, especially on-site observation, are useful.

**keywords:** assessment, evaluation, before-after design, experimental design, qualitative methods

### I. はじめに

高齢化社会を迎え、財政が逼迫する中で、科学的根拠に基づく保健活動の重要性が叫ばれている。すなわち、地域特性などをしっかりと把握し、真に必要なや効果の高い事業を見分けて実施していくことが求められている。そこで、健康増進計画の策定やその中間評価、生活習慣病予防プログラムの評価の実際などを踏まえながら、保健活動における評価の現状と課題を考察したい。なお、「評価」という概念にはどのようなことまでが含まれるか、いろいろな考え方がありうるだろうが、本稿では表1に示すようになるべく広めにとらえたい。

表1. 評価の種類

事前評価 assessment
事後評価 evaluation
地域全体の評価
事業参加者の評価
数量的手法
質的手法
事前評価に有用な既存資料の種類
国勢調査、住民基本台帳
人口動態統計
健康診査データ
国民健康保険レセプト
介護保険データ
事業の評価のデザイン
ケーススタディーデザイン
前後比較デザイン
準実験デザイン
実験デザイン（無作為比較対照試験）

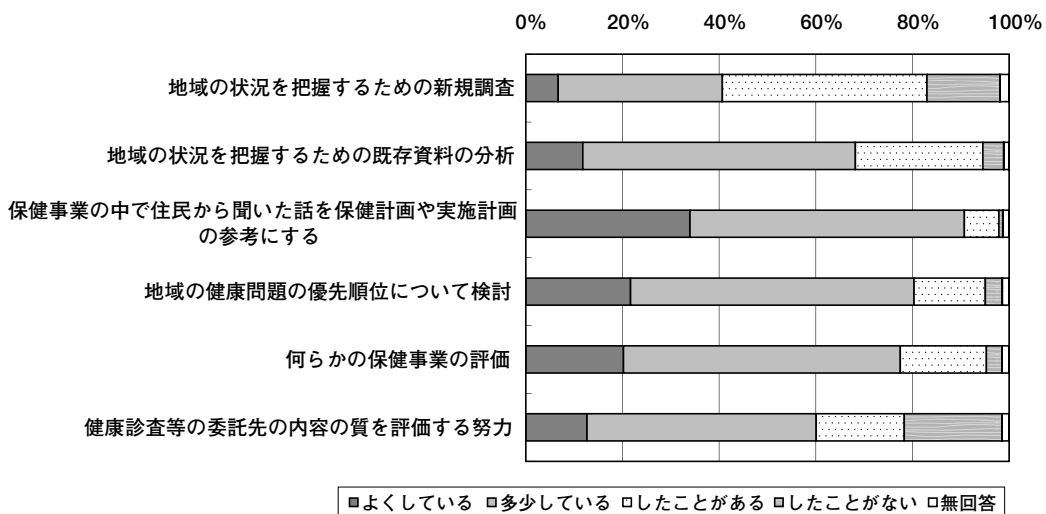
### II. 事前評価と事後評価

評価には、事前評価assessmentと事後評価evaluationがある。例えば、健康増進計画の策定の際には、ほとんどの場合に、事前評価、即ち地域診断が行われている。その際に、既存の統計資料を用いたり、一定の年齢の住民を無作為抽出してアンケート調査を行ったりすることが多い。また、住民へのインタビューや座談会（フォーカスグループディスカッション）によって質的な情報の収集が行われることも多い。既存統計資料としては、国勢調査・住民基本台帳による属性別人口、人口動態統計による死亡数・死因別死亡割合・標準化死亡比（SMR）、健康診査のデータ、国民健康保険レセプト、介護保険データなどが一般的に有用である。事前評価の分析方法としては、記述疫学として行うのが王道である。具体的には、ヒト：性・年齢階級別の分析、場所：自分の地域と県や国などの広域との比較や地域間の比較、時間：年次推移を見るのが基本である。

事後評価としては、健康増進計画の中間評価などでは、事前評価と同様の項目の評価を再度行い、事前評価時との変化を見る形が一般的である。一方で、ある事業の評価を行う場合には、地域住民全体ではなく、その事業の参加者などを対象にして評価を行う場合が多い。その際の評価デザインなどについては、後に詳述する。

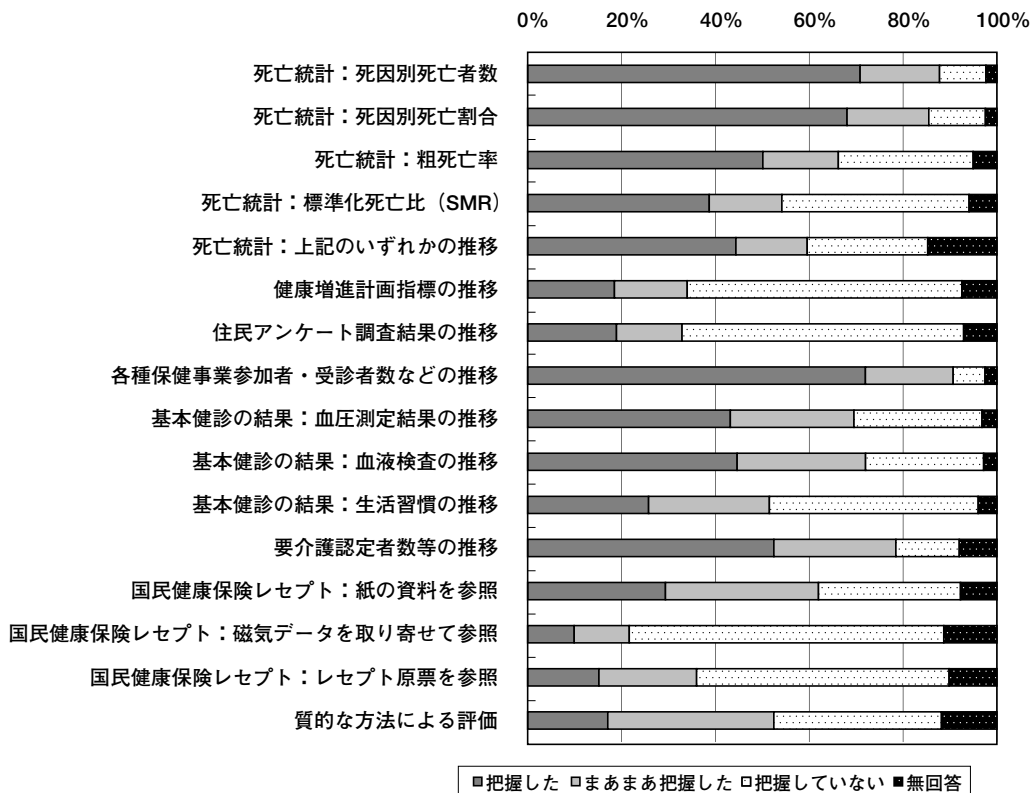
### III. 全国の市町村における評価の実施状況

全国保健センター連合会の調査による市町村における評価の実施状況を図1に示す。最近では、特定健診・保健指導に焦点を絞った評価の状況に関する調査などは行われていると考えられるが、幅広く評価の状況を調べたものとしては若干古いデータになってしまい恐縮である。「よくしている」または「多少している」と回答した市町村の割合は、



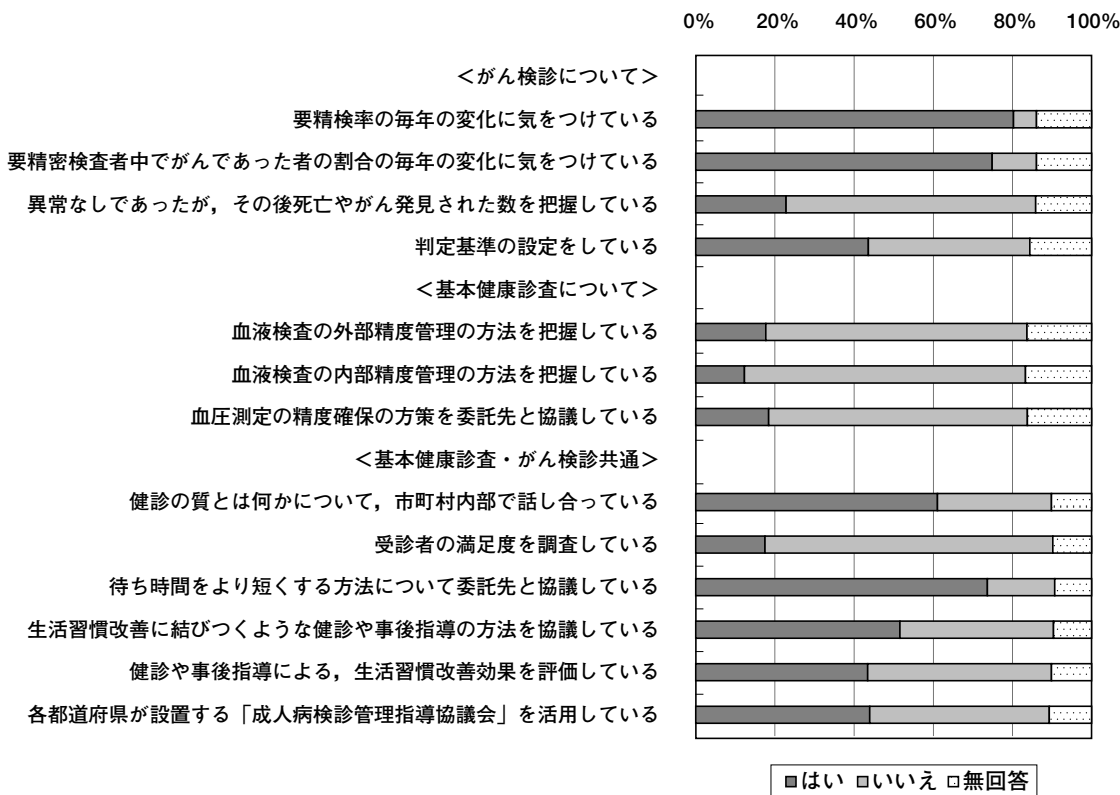
平成16年全国保健センター連合会調べ N=2,573

図1. 市町村における評価の実施状況 (過去3年間の実施状況)



平成17年全国保健センター連合会調べ N=1,844

図2. 住民の健康に関する指標等の把握（過去3年間の実施状況）



平成17年全国保健センター連合会調べ N=1,844

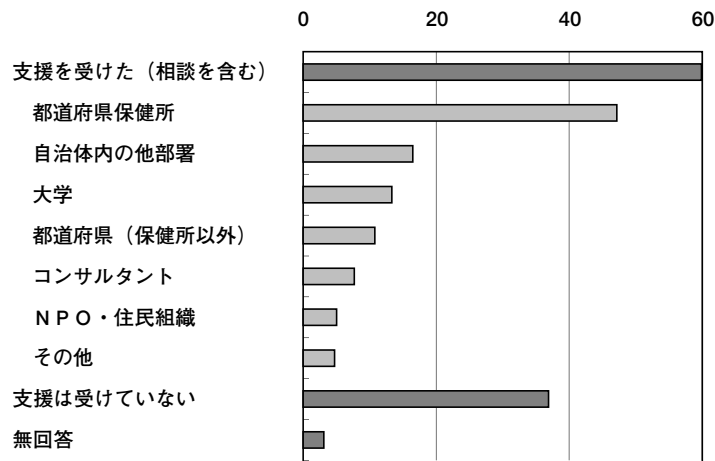
図3. 健診の評価及び精度管理の状況

何らかの保健事業の評価において8割近くに上る。また、地域の状況を把握するための既存資料の分析は7割程度に上るが、地域の状況を把握するための新規調査は4割程度にとどまる。保健事業の中で住民から聞いた話などの質的な情報の活用は9割近くに及ぶ。

住民の健康に関する指標等の把握の状況を図2に示す。「把握した」または「まあまあ把握した」との回答が多い指標としては、死因別死亡者数、死因別死亡割合、各種保健事業参加者数・受診者数などの推移について9割近くに達した。一方で、標準化死亡比（SMR）5割強、基本健

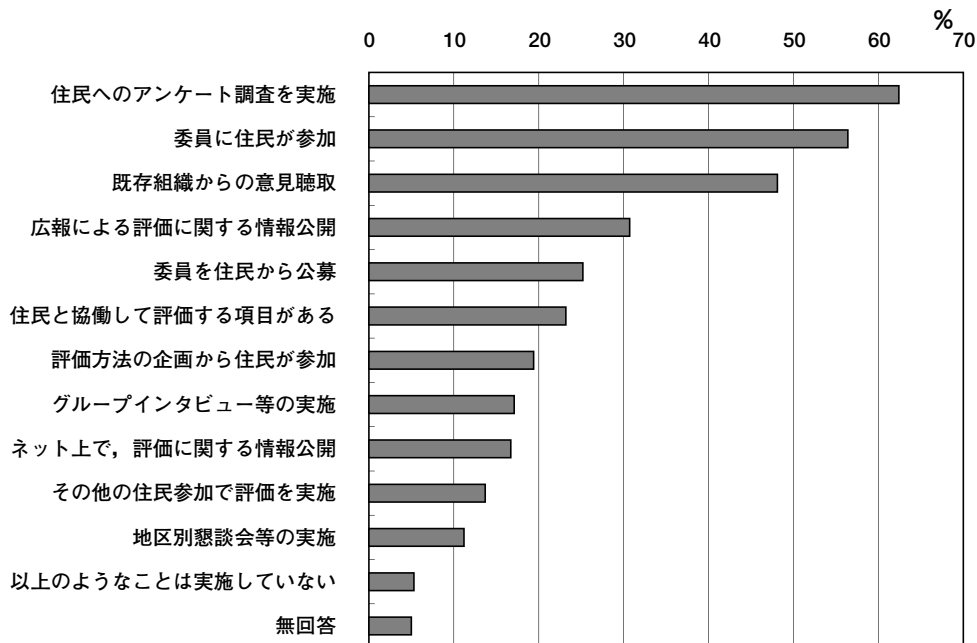
診による血圧測定結果の推移7割、基本健診による生活習慣の推移5割とまだまだ少ない。国民健康保険レセプトの活用については最も多い紙の資料の参照でも6割強に留まる。質的な方法による評価は5割程度であったが、ちょっとした質的な評価を含めるともっと多いのではないかと考えられる。

健診の評価及び精度管理の状況を図3に示す。要精検率や、要精密検査者中でのがんであった者の割合（陽性反応適中度）の毎年の変化に気をつけていると回答した市町村は8割近くに上った。しかし、異常なしであったがその後死



平成17年全国保健センター連合会調べ N=1,844

図4. 情報収集や活用において支援を受けたか（過去3年以内の状況）



平成17年全国保健センター連合会調べ N=659  
健康増進計画中間評価に取り組んでいる市町村のみが回答

図5. 健康増進計画中間評価における住民参加の状況

亡やがん発見された数（敏感度）を把握している市町村は2割強に限られていた。健診や事後指導による生活習慣改善効果を評価していると回答した市町村は約4割であった。

情報収集や活用において支援を受けている状況を図4に示す。全体で6割の市町村が支援を受けており、最も多かったのは都道府県保健所からの支援であり5割弱の市町村が受けていた。次いで、自治体内の他部署の支援2割弱、大学の支援1割強であった。

健康増進計画中間評価における住民参加の状況を図5に示す。最も多いものは、住民へのアンケート調査を実施で6割強である。評価方法の企画から住民が参加、グループインタビュー等の実施は2割弱に留まった。

#### IV. 事業評価のデザイン

事業の評価を行う際には、どのようなデザイン、つまり大枠としての方法を採用するかが重要である。評価のデザインには表1に示したような種類がある。

ケーススタディーデザインcase study designは、事業の後に1回だけ調査を行う方法である。また、ある一人一人のケースについて、物語的narrativeに変化を記述する方法も含まれると考えられる。このデザインは、事業の改善すべき点などを質的に評価しようとする場合には有用である。しかしながら、効果の度合を数量的に検証することはできない。ただし、思い出しバイアスの影響が大きい可能性があるものの、事業後の1回の調査で、事業前の状態と、事業後の状態の両方を調査して、数量的に検討を行おうとする場合もある。

前後比較デザインbefore-after design, time series designは、事業の前と事業の後の2回、同じような調査を行って、その変化を分析することによって効果を検証しようとする方法である。保健現場で事業の評価を行おうとする際に、十分に実施可能であり、最もお薦めの方法である。特定保健指導の評価においても、事業前後の体重や腹囲、また生活習慣の変化によって効果を評価しようとする試みは多い。しかしながら、前後比較デザインの最大の欠点は、事業以外の理由によって改善したものと、事業の効果とを区別できない点である。例えば、保健指導によって体重が減ったのではなく、猛暑による夏やせで体重が減ったなどの可能性があるかもしれない。そのようなことがないかを含めて、結果を解釈する必要がある。前後比較デザインのもう一つの重要な問題としては「平均への回帰」がある。詳細は成書をご覧頂きたいが、個別健康教育においては、平均への回帰の問題を軽減するために、健診の後の健康教育の第1回目の時に再度採血を行うようになっている。特定保健指導においては、保健指導の際に採血は行わないのが標準的な方法となっているため、そのような定めはない。

準実験デザインquasi-experimental designは、本人の希望等による対照群を設定する方法である。例えば、特定保健指導該当者にその案内を行い、参加を希望した人を介入群、希望しなかった人を対照群とする。そして、1年後の

特定健康診査結果で1年間の変化を比較するなどの方法である。対照群を設けているため、前述の前後比較デザインでの猛暑の影響などを取り除くことができる。しかし、準実験デザインにも致命的な問題がある。それは、事業への参加を希望するような人は、そもそも健康への関心が高いため、仮に事業の効果がなかったとしても自分の力で健康状態を改善している可能性がある点である。

実験デザインexperimental design（無作為比較対照試験randomized controlled trials）は、無作為にくじ引きなどによって介入群と対照群を割り付ける方法である。健康への関心度を含めて、あらゆる交絡因子が介入群と対照群で均等になることが期待され、科学的に最も正しい方法であると言われている。しかし、実験デザインにも致命的な欠点がある。ひとつは、倫理的な問題である。ある意味、人体実験であるため、きちんとしたインフォームドコンセントを受けて実施する必要がある。また健康教育の介入を行う場合、対照群に選ばれてしまった人は健康教育を受けることができないという不利益が生じる。この点については解決策があり、対照群には、研究期間が終了した後に、時間差で健康教育を行うなどの方法がとられることが多い。もうひとつの問題は、外部妥当性である。実験デザインの被験者になっても良いという人は、非常に特殊な、奇抜な人であると考えられることができる。そこで、ある実験デザインで出た結果が、一般の人にも適用できるかという大きな疑問である。ある健康教育について、実験デザインによる研究としては効果が出たが、それを一般集団に適用した場合には、対象集団のモチベーションの違いなどによって、ほとんど効果がでなかったということもありえるのである。しかしながら、もし実験デザインの対象者でさえ効果が出なかったら、一般集団ではもっと効果がでないだろうと言うこともできる。

以上をまとめると、どの方法も一長一短あり、どのデザインが絶対によい、または絶対に悪いということはなさそうである。私見としては、現場での事業効果の評価には実施上簡便である前後比較デザインが最も適すと考えている。また、ケーススタディーデザインで物語的に評価する方法も魅力的である。一方で、国を挙げて、新しい事業を始めようとする場合には、いくつかのモデル地区で、現場と研究者の連携によって実験デザインによる検証を行うことが重要であると考えられる。

#### V. 質的評価の重要性

事業の評価を行おうとする際に、その目的は2つあると考えられる。ひとつは、事業の効果があるかないかの検証、もうひとつは事業の改善である。現場においては、特に後者の目的で評価を行うことも多いと考えられる。前項の種々のデザインで数量的に評価を行った際に、効果の有無についての結果は出すことができるが、事業の実施方法のどこをどのように改善すべきかということについては、答えがでないことが多い。事業の改善のためには、数量的評

価よりも質的評価の方が役に立つのではないだろうか。質的評価のための情報収集としては、インタビュー、フォーカスグループディスカッション、アンケート調査の記述回答質問、観察などがある。いずれも有用であるが、特に観察が有用であると個人的には思っている。事業を実施しながら、健康教育の話の内容やそれに対する参加者の反応などを観察することにより、いろいろな改善点を挙げるができるであろう。製造現場における改善や品質管理においても、数量的なデータ収集を行うこともあるが、まずは現場の観察が行われる。なお、質的評価によって、事業の改善案ができたが、現行の方法と、改善案の方法と、どちらの方が効果が高いか確信が持てないという場合には、数量的な評価で効果の大小を検証することも有用であろう。ただし、本来的には、1つの実施方法の違い毎に、それぞれきちんとした数量的評価を行う必要が生じる。改善可能性のある方法の変更箇所が100箇所あった場合には、100個の評価研究を行わなくてはならないことになる。そこで、実務的には、ほとんどの改善箇所について、直感的に考えて効果が高そうな方法を選ぶのが良いのと思う。

## VI. 生活習慣病予防事業評価の例

事業評価の例として、日本看護協会で、特定保健指導が始まる前の平成19年度にモデル的に実施した生活習慣病予防事業について、その評価分析の結果をお示ししたい。評価デザインは、実施の容易さを考えて、前後比較デザインを採用した。4回シリーズの保健指導（グループ支援等）により、生活習慣改善を目指すプログラムである。通常の特定期間とは異なり、1回目と4回目の保健指導の際に採血も行っているが、それ以外の項目については、通常の特定期間でも同様の評価を行うことができる。表2に検査値の変化の結果を示した。基本的に、個人々人について、1回目と4回目の差を計算し、その平均値を求め、また対応のあるt検定を行った。その結果、体重、腹囲、HbA1c等で統計学的有意に効果があることを検証すること

ができた。次に、各項目について、改善、維持、悪化がどのような割合であったかについて図6にまとめた。ここでは数値に少しでも変化があった場合には改善または悪化と分類し、全く変化が無かった場合に維持とした。これも、体重、腹囲、HDLコレステロール等で改善者が多いという結果となった。

検査値の有効数字が細かく変化に対して鋭敏な項目については、ほとんどが改善または悪化に分類され、逆にHbA1cなど検査値の有効数字が粗い項目については維持の割合が大きい。この評価において、体重は0.1kgだけ減った場合でも、改善したと大きな声で言うて良いのか、例えば3kg以上減った場合に改善とした方が良いのではないかという意見が出た。そこで、どの程度改善した人が何%くらいいるのかを分析した結果を図7に示す。やや左に裾を引いたベル型曲線となった。図6のように0.1kg以上の改善で良ければ改善者割合は71.1%であるが、図7で3kg以上改善した人は18.1%という結果であった。合格ラインを甘くすれば合格者は多く、厳しくすれば少ないということになる。その評価の検討において、この合格ラインを決めたいという話も出たが結局断念した。権威のある専門家が集まって、コンセンサスが得られれば、エイヤアと合格ラインを決めることもできるだろうが、科学的根拠を持って決めることは不可能であろうと考えている。図7において、このベル型曲線の面積を左右にちょうど半々に分けるような線を引くことができる。それは表2で示した体重の差の平均値である-1.35kgである。この値で評価するのが最も良いのではないかと考えている。

日本看護協会のこの事業は、その後も毎年実施方法及び評価方法を改善しながら継続している。平成20年度の保健指導については、平成21年度の健診結果も使いながら現在評価を行っているところで、今年度中に報告書をまとめる予定である。また、平成21年度については事業前後の質問票等を改善して、現在、保健指導を行っているところである。

表2. プログラムの1回目と4回目の検査値の差

	n	1回目 平均値	4回目 平均値	差の 平均値	差の95%信頼区間		有意確率
					下限	上限	
体重	381	63.80	62.46	-1.35	-1.574	-1.117	<0.001
BMI	320	24.61	24.04	-0.58	-0.685	-0.468	<0.001
腹囲	378	87.54	85.74	-1.81	-2.205	-1.409	<0.001
HbA1c	407	5.61	5.56	-0.05	-0.083	-0.019	0.002
血糖	295	106.03	105.32	-0.71	-3.154	1.730	0.567
中性脂肪	267	153.08	149.67	-3.41	-17.509	10.692	0.635
LDL	269	134.07	136.24	2.17	-0.605	4.947	0.125
HDL	272	55.91	58.38	2.47	1.476	3.464	<0.001
SBP	321	135.82	134.89	-0.93	-2.538	0.681	0.257
DBP	321	82.53	81.60	-0.93	-2.101	0.232	0.116

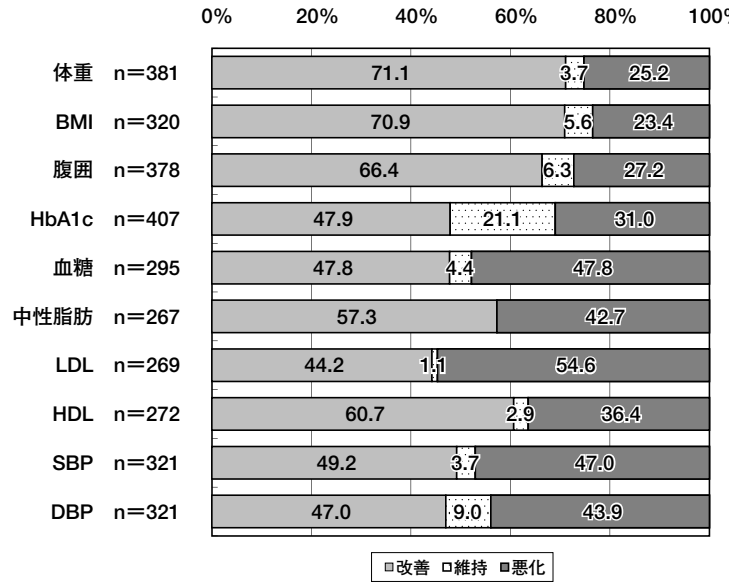
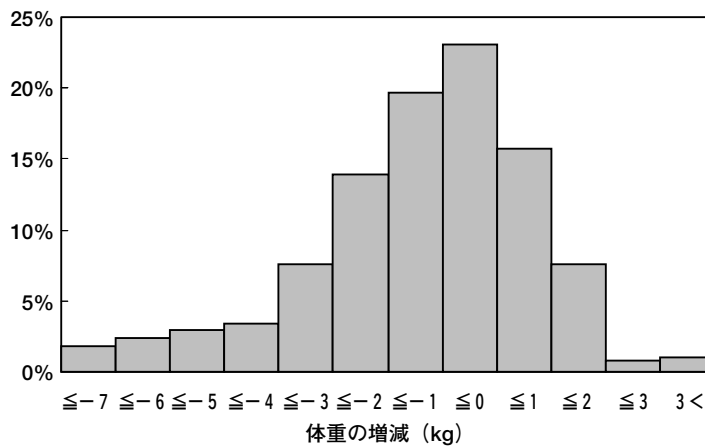


図6. プログラム1回目と4回目の検査値の改善



n=381, ちょうど境界の場合は下の階級に含まれている。

図7. 体重の1回目と4回目の差の分布

## VII. まとめ

保健事業のPDCA (Plan-Do-Check-Action) サイクルの中で、評価Checkをしっかりと行っていく必要がある。しかしながら、PDCAの中で最も重要なことは実施Doであると思っている。時々、計画や評価に熱心になる余り、実施がおろそかになっている例を見かけることがあるが、それでは本末転倒であろう。大部分の力を実施に注ぎつつ、適度な労力の割合で評価を行っていく必要が有ろう。

また、評価デザインで述べたように、科学的に完璧な評価は不可能である。一方で、完璧な評価でなくても、有る程度の事業効果の検証や、また事業の改善は行うことができる。平成21年に成立した民主党政権は、官僚主導から政治主導へということのひとつの旗印にしている。このことは言い換えると、専門家の視点から住民の視点重視へとい

うことであろうと考えている。そこで、評価を行う場合には、アウトカム評価や科学的により正しい評価を追求することに加えて、住民の心に響く評価も重要なのであろうと思う。住民に広報する場合など、一人の事例のリアルな質的評価の方が、数字の羅列よりも力強い説得力を持つ場合も有ろう。

評価においては、情報の利用が必須となるが、最近、いくつかの法的環境の変化が起きている。そのひとつとして、平成21年度から施行された統計法の改正がある。新統計法では、公的統計は「国民にとって合理的な意思決定を行うための基盤となる重要な情報である」とし、「公的統計は、広く国民が容易に入手し、効果的に利用できるものとして提供されなければならない」という基本理念が掲げられている。地方の現場にある業務上の情報と、この統計法による公的統計とは完全に重なるわけではないが、この理念は

非常に重要なものであると考えられる。評価を行う際に、その評価だけを目的に新たに情報収集することは多大な労力と経費を要する。個人情報の保護を万全にしつつ、既存の情報を有効に活用して評価を行っていくことが必要であろう。

### 参考文献

- 1) 全国保健センター連合会. 市町村保健センターのあり方研究会報告書. 2005.
- 2) 全国保健センター連合会. 平成17年度市町村保健センター及び類似施設調査市町村保健活動調査 調査研究結果報告書. 2006.
- 3) 日本看護協会. 平成19年度先駆的保健活動交流推進事業 生活習慣病予防活動支援モデル事業報告書. 2008.
- 4) 武藤孝司, 福渡靖. 健康教育・ヘルスプロモーションの評価. 東京: 篠原出版; 1994.
- 5) 尾島俊之. 調査デザインの基礎. 柳川洋, 他編集. 地域保健活動のための疫学 第2版. 東京: 日本公衆衛生協会; 2006. p.266-73.