

## 乳幼児健康診査及び学校保健統計の活用を想定した乳幼児身体発育曲線の試作

研究協力者 加藤則子 (十文字学園女子大学教育人文学部)  
研究分担者 磯島 豪 (国家公務員共済組合連合会虎の門病院小児科)

### 研究要旨

新型コロナウイルス感染症 (COVID-19) 拡大の影響下で実施が延期されてきた乳幼児身体発育調査の、調査の方向性が継続的に検討されてきた。従来のように集団健診方式で年月齢に関して連続的なデータを取っていく方向性の中で、必要にして十分なサンプルサイズを検討する必要が生じた。また公的統計調査実施の承認を得ていく上でのサンプルサイズに関する根拠として、誤差率の算出が必要であるが、本調査に特化した計算方法を考案する必要があった。

分母を真の“3パーセンタイル値と50パーセンタイル値”の間隔、真の“50パーセンタイル値と97パーセンタイル値”の間隔とし、3パーセンタイル値と97パーセンタイル値の推計値の標準誤差をそれぞれ対応する分母で割ったものを誤差率と定義した。サンプリングに関し、2010年乳幼児身体発育調査公表値の算出に用いたサンプル数に対して、病院調査は0.3倍、一般調査0歳児は0.7倍、1歳児は0.42倍、2歳以降は1.5倍とすることで、3~50パーセンタイル値の間隔及び50~97パーセンタイル値の間隔が概ね5%以内の誤差率で算出されることが明らかになった。

### A. 研究目的

当初令和2(2020)年に予定されていた厚生労働省乳幼児身体発育調査は、新型コロナウイルス感染症 (COVID-19) 拡大の影響で延期されている。乳幼児身体発育調査は過去半世紀以上にわたって、10年に一度の間隔で行われ、その一般調査は調査対象者を集めた集合健診の形式を取ってきた。COVID-19の影響が続く中で、従来のような方式での調査は感染予防の観点から好ましくないという自治体等からの意見もあり、令和3年度から令和4年度にかけて、自治体でルーティンに行っている乳幼児健康診査(以下健診)活動の中収集された身体計測データを活用していくという代替案が浮かび上がった。

これを受けて本研究では令和3年度に、平成22(2010)年乳幼児身体発育調査データの中から、一般調査データにおいては乳幼

児健診で得られる年月齢のデータを抽出し、病院調査データにおいては系統抽出法を用いてよりデータ数をコンパクトにし、それらに加えて5歳及び6歳の学校保健統計調査データを用いて、身体発育曲線の平滑化を試み、2010年公表値との比較を行うなどして評価を行った。

このような経緯から、令和4年度当初においては、自治体においてルーティンで行われる健診の年月齢のみから発育曲線を作成する場合の追加検討として、一般調査の地区数を増やすことを仮定して、乳幼児健診に対応する各年月齢のサンプル数を増やして検討してみるという課題が持ち上がった(研究1)。

一方で、調査方法に関する検討が進む中で、従来通りの健診年月齢に絞るのではなく、対象地区に住む子どもをすべて対象とし、年月齢に関する連続的な分布において

サンプリングする方法が浮上してきた。そのため、データセットの年月齢分布を至適なものとするのが課題となった(研究2)。

公的統計調査実施に関する承認手続きでは、一般的に、標準誤差が標本平均(や割合)の5%以内となるようにサンプルサイズを定めるとされている。乳幼児の身体計測値に関して、標本平均に対する誤差率5%を考えると、例えば体重10kgに対し、5%すなわち500gという幅は、出来上がりの、3, 10, 25, 50, 75, 90, 97の7つのパーセンタイル値の間隔にほぼ匹敵することとなり、これを誤差率と設定するようなサンプリングは、非現実的である。2010年同調査のサンプルに関して、3, 97パーセンタイル値の誤差率の考え方については令和3年度本研究班横山分担研究報告でまとめられており、これを活用して、適切な分母を見出し、この調査に特化した誤差率の計算方法を開発することが課題(研究3)となる。

そして、以上のような考案、検討を行うことによって得られた、適切な年月齢分布のサンプリング方法に対して、学術的な裏付けがあったほうが望ましい。誤差率の検討によって望ましいとされた年月齢列のサンプルサイズに関して、Cole(2021)(文献1)が推奨している、正規変換データをもとに算出された各パーセンタイル値の標準誤差と矛盾しないことが示されれば、サンプリングが妥当であることの傍証とすることができる。この検証も今年度の課題の一つ(研究4)とする。

## B. 研究方法

### 研究1

令和3年度の本研究では、2010年乳幼児身体発育調査におけるサンプルに関して、自治体におけるルーティンの乳幼児健診の

年月齢に相当するサンプルのみを抽出し、これにデータのバランスを勘案して出生から1か月までのデータとして2010年病院調査の10分の1サンプル、5~6歳以降のデータとして2010年学校保健統計調査の50分の1サンプルを追加したデータセットを作成し、出来上がりとしては、一応安定した曲線が得られた。

この追加解析として、一般調査の調査地区が増やされることを想定し、自治体におけるルーティンの乳幼児健診の年月齢に相当するサンプルを2度用いる方法をとって倍加させ、これに病院調査データ、学校保健調査データを同様に追加したものをデータセットとして、GAMLSS on RのBCToモデルを用いて、パーセンタイル曲線の平滑化を行った。

### 研究2

至適なサンプルの年月齢分布を探索するために、まずすべての年月齢区分に関して系統抽出法により2分の1にした。その結果を見ながら、必要に応じて年月齢区分ごとの標本数を増減させた。平滑化結果の評価としては、Cole TJ(2021)(文献1)の推奨に従って、正規化データにおける中央値の標準誤差が0.041に収まっていることを確認した。

データの正規変換は、以下のコードによる。

```
m0<-lms(wt, age,
families=c("BCCGo", "BCPEo", "BCTo"),
data=mws, k=9, calibration=F,
trans. x=T, cent=cent)
wtz<-resid(m0)+10 #0値を避けるために、
10を足してある
mws$Tage<-(mws$age)^(m0$power)
```

中央値の標準誤差の算出と出力は以下のコードによる。

```

predml<-
predict(m1, what="mu", type="response", s
e. fit=TRUE)
plot(mws$age, predml$se. fit)
predml$fita<-predml$fit-10
plot(mws$age, predml$fita)

```

### 研究 3

誤差率の定義を以下のように設定した。分母を真の“3 パーセンタイル値と 50 パーセンタイル値”の間隔、真の“50 パーセンタイル値と 97 パーセンタイル値”の間隔とする。“3 パーセンタイル値と 50 パーセンタイル値”の推計値の間隔、“50 パーセンタイル値と 97 パーセンタイル値”の推計値の間隔の標準誤差を、それぞれ対応する分母で割って誤差率を求める。これらの算出は横山徹爾作成プログラム（令和 3 年度分担研究報告書）による。

この定義による誤差率の算出を、2010 年公表値算出のデータセットと、本調査設計を検討する際に提案された地区数から、回収率 0.7 倍として見込まれる標本数によるデータセット（表 1）で行った。病院調査データは十分にあることから 2010 年公表値作成データセットの 0.3 倍、一般調査の 0 歳は、WHO の実績（文献 2）に従って 2010 年公表値作成データセットの 0.7 倍、2 歳以降は 2010 年の場合のデータから減らすことができない（前後するが、研究 2 の結果に従って）ため余裕をもって 2010 年公表値作成データセットの 1.5 倍、1 歳は 2010 年の場合ほど多い必要はないため、2 歳以降に年齢別の数を合わせ、2010 年公表値作成データセットの 0.42 倍としたものである。サンプルデータは系統抽出法を用いて年月齢区分ごとに増減して、データセットを作成した。

### 研究 4

表 1 のデータセットから GAMLSS on R BCCGo モデルにて求められる平滑化パーセンタイル値に関し、Cole TJ (2021) に示されているように、GAMLSS on R による平滑化結果に対し、R 上の Bootstrap を行い、正規変換データパーセンタイル値の推計誤差をもとめ、推奨されている 3, 97 パーセンタイル値に関して 0.061 以内、中央値に関して 0.041 以内という目安との比較を行う。

GAMLSS on R による平滑化結果に対する、R 上の Bootstrap は以下のコードによる

```

library(gamlss)
library(gamlss.foreach)
m0<-lms(wt, age, families=c("BCCGo"),
data=mw, k=9, calibration=F,
trans. x=T, cent=cent)
nage<-seq(0, 78, 1)
boC221021a<-
centiles.boot(m0, xname="age",
xvalues=nage, cent = c(3, 50, 97), B=10)

```

データの正規化は、以下のコードによる。計算の途中で正規化値がゼロとなると、エラーが起こるため、これに 10 を加算したものを、平滑化の計算に用いている。

```

m0<-lms(wt, age, families=c("BCCGo"),
data=mw,
k=9, calibration=F,
trans. x=T, cent=cent)
nx=mw$age^(m0$power)
mw$wtz<-resid(m0)+10 #正規変換値を求
める
newd<-
data.frame(mw, nx=mw$age^(m0$power))

```

正規化データに対する Bootstrap の実施は以下のコードによる。

```

m3<-
gamlss(wtz~pb(nx), sigma. fo=~pb(nx), nu.
fo=~pb(nx), tau. fo=~pb(nx),

```

```
family=BCTo, data=newd)
boC221021c <- centiles.boot(m3,
xname="nx", xvalues=nage, data=newd,
power=m0$power, cent = c(3, 50, 97),
B=100)
```

中央値並びに 3, 97 パーセンタイル値に関する標準誤差の出力は以下のコードによる

```
summary(boC230208b, sd)
```

### C. 研究結果

#### 研究 1

表 2 のデータセットから図 1 の平滑化結果が示されることは、昨年度の本研究において報告した。ここで、3, 6, 10, 18, 36 か月のデータを、同じデータを 2 度使うことにより倍加させ(表 3)、平滑化を試みた(図 2)。図 2 に示すように、このようなデータセットでは、平滑化曲線が乱れ、発育の目安として用いることができるような曲線が得られないことが分かった。

#### 研究 2

2010 年公表値の男子体重データセットから求めた平滑化曲線は図 3-1 のとおりであり、その際の正規化データ (z スコア) を平滑化した場合の中央値の標準誤差は図 3-2 であり、0.041 を下回った。これについては昨年度前に報告済みである。ここで、すべての年月齢区分におけるサンプルサイズを半分になるようなデータセットで平滑化したところ、平滑化曲線は月齢 60 以上で安定を欠き(図 4-1)、正規化データ (z スコア) を平滑化した場合の中央値の標準誤差は、60 か月以上で 0.041 を超え、高い年齢において、サンプルを半減させたことによる悪影響が大きく出ることが分かった。このため、サンプルを 2 歳以降はそのままとし、サンプルサイズに余裕のある病院調査

を 10 分の 1、一般調査の 0 歳代と 1 歳代を 0.3 倍としたデータセットで平滑化を行ったところ、平滑化曲線は安定し(図 5-1)、正規変換データ (z スコア) を平滑化した場合の中央値の標準誤差は 0.041 に収まった(図 5-2)。

#### 研究 3

2010 年公表値の男子体重データセットで、研究方法に記載した算出方法による誤差率、97 パーセンタイル値と 3 パーセンタイル値について、年月齢別に求めた結果を図 6 に示す。97 パーセンタイル値において、3 歳以降、誤差率が 5%を上回って増加していった。表 1 に示した、調査設計の検討段階における地区数に相当したサンプリングをもとに算出した誤差率を男子体重、女子体重に関して、図 7 に示す。男女ともに、ほぼすべての年月齢に関して、5%をほぼ下回る誤差率が算出された。

#### 研究 4

2010 年男子体重の計算に用いたデータセットで、GAMLSS on R BCCGo モデルで平滑化を行った結果に対し、R 上の Bootstrap を 100 回かけた結果を、標準誤差をグレーの帯として、図 8-1 に示す。また、このデータを正規化変換したもの(に 10 を加算したもの)に対して、R 上の Bootstrap を 100 回かけた結果を図 8-2 に示す。いずれの図でも、標準誤差範囲を示すグレーの帯が、40 か月を超えるあたりから広がっていることが分かった。正規化変換したデータの 100 回の Bootstrap による 3, 50, 97 パーセンタイルの標準誤差の値を、年月齢に対して示したのが図 8-3 である。50 パーセンタイルの標準誤差はあらゆる年齢で 0.041 を下回ったが、3, 97 パーセンタイルの推計誤差は、40 か月を超えるころから 0.061 を上回って増加していった。

図 5-1 における男子体重の計算に用いたデータセットで、GAMLSS on R BCCGo モデルで平滑化を行った結果に対し、R 上の Bootstrap を 100 回かけた結果を、標準誤差をグレーの帯として、図 9-1 に示す。また、このデータを正規化変換したもの（に 10 を加算したもの）に対して、R 上の Bootstrap を 100 回かけた結果を図 9-2 に示す。いずれの図でも、標準誤差範囲を示すグレーの帯の 40 か月以上の部分の幅が、図 8-1、図 8-2 よりも狭まっていた。同データセットに関して、正規化変換したデータの 100 回の Bootstrap による 3, 50, 97 パーセンタイルの推計誤差の値を、年月齢に対して示したのが図 9-3 である。50 パーセンタイルの標準誤差はあらゆる年齢で 0.041 を下回ったが、3, 97 パーセンタイルの標準誤差は、40 か月と 48 か月前後を除けば、0.061 に収まっていることが分かった。

表 1 に示すデータセットの正規変換したデータにおいて、GAMLSS on R BCCGo モデルで平滑化を行った結果に対し、R 上の Bootstrap を 1000 回かけ、3, 50, 97 パーセンタイルの推計誤差の値を、年月齢に対して示したのが図 10 である。50 パーセンタイルの標準誤差はあらゆる年齢で 0.041 を下回った。3, 97 パーセンタイルの推計誤差は、0 歳、48 か月前後、60 か月以降を除けば、ほぼ 0.061 に収まっていることが分かった。

## D. 考察

### 研究 1

自治体のルーティンの乳幼児健診年月齢のみに限ったデータのように、年月齢分布が間隔をあけて散在するようなデータセットでは、データセットのバランスが崩れやすく、局所的にデータ数を増やすことで却

って思わしくない平滑化結果が得られることが分かった。既得の乳幼児健診データを用いて発育曲線を作成する方針が、その後別の方向に変わったため、この検討は、ここまでとする。

### 研究 2

サンプルの年月齢別分布をいくつかのタイプに分けて設定してみたところ、2 歳以上の大きい年月齢においてサンプルを減らすと良好な平滑化が得られず、2 歳未満でサンプルを減らした場合は、比較的良好な平滑化結果が得られた。次回調査のサンプリングを考える上では、費用対効果の観点から、2 歳未満では、2010 年のものより少ないサンプルサイズ、2 歳以上では 2010 年のものより少し増やしたサンプルサイズとすることが妥当であることが分かった。

### 研究 3

公的統計調査においてサンプルサイズを決める上で、誤差率の概念は重要である。本調査の性格に特化した誤差率の計算法が開発され、そのことにより、設定された年月齢別のサンプルの分布が適切であることが導き出された。この設定に基づいて、公的統計調査実施の承認を得ていく上での調査地区数の算出につながったことは、大きな意義があると言える。

### 研究 4

研究 3 で設定された年月齢ごとのサンプルサイズによるデータを用いて、Cole TJ (2021) (文献 1) に示されている検証を試みたところ、中央値に関しては、正規変換値の標準誤差が推奨値より十分小さいことが分かった。3, 97 パーセンタイル値に関しては、多くの年月齢で推奨値に近い値を示したが、0 歳と 60 か月以降、そして 48 か月前後では、これを上回った。GAMLSS では、データ

の両端の推計誤差が大きくなる傾向があるため、これを反映しているものと考えられる。48 か月前後で誤差が大きくなる理由は不明である。

GAMLSS on Rによる平滑化及びBootstrapによる誤差率の計算方法に関しては習熟の途上であるため、本検討の結果を直ちにサンプルサイズの設定に反映させるものではないと考える。

## E. 結論

公的統計調査のサンプルサイズ算定においては、標準誤差を平均(や割合)で除した誤差率が問題になる。分母を真の“3 パーセンタイル値と 50 パーセンタイル値”の間隔、真の“50 パーセンタイル値と 97 パーセンタイル値”の間隔とし、“3 パーセンタイル値と 50 パーセンタイル値”の推計値の間隔、“50 パーセンタイル値と 97 パーセンタイル値”の推計値の間隔の標準誤差をそれぞれ対応する分母で割ったものを誤差率と定義した。サンプリングに関し、2010 年公表値の算出に用いたサンプル数に対して、病院調査は 0.3 倍、一般調査 0 歳児は 0.7 倍、1 歳児は 0.42 倍、2 歳以降は 1.5 倍とすることで、3, 97 パーセンタイル値が概ね 5% 以内の誤差率で算出されることが明らかになった。

### 【参考文献】

- 1) Sample size and sample composition for constructing growth reference centiles. Cole TJ. Stat Methods Med Res. 2021 Feb;30(2):488-507.

- 2) Group WHOMGRS. WHO Child Growth Standards based on length/height, weight and age. Acta Paediatr Suppl 2006; 450:76-85.

## F 健康危険情報

なし

## G. 研究発表

### 1. 論文発表

- 1) 伊藤 善也, 水流 聡子, 安井 清一, 西岡 桃子, 村田 光範, 田中 敏章, 神崎 晋, 加藤 則子, 横谷 進, 大藪 恵一, 長谷川 奉延, 学校健康診断における年齢、身体発育値や体格に関する指標の表記法. 日本成長学会雑誌 2022;28(1):7-10
- 2) 田中敏章, 伊藤 善也, 加藤 則子, 村田 光範, 磯島 豪, 他 小域成長研究データに基づく日本人女子の成長(第 3 編) 思春期開始の時期が暦年齢、骨年齢に対する成長率に与える影響. 日本成長学雑誌 2022;28(1):19-25

### 2. 学会発表

なし。

## H. 知的財産権の出願・登録状況

### 1. 特許取得

なし。

### 2. 実用新案登録

なし。

### 3. その他

なし。

表1 2010年公表値作成データセットと研究3データセットの人数

	2010年調査		次回調査			変更理由
	集計人数※1		集計人数			
	男女計	男女計	1歳当たり	男女各※2		
病院調査	4,698	1,409	1,409	705		十分な精度があるので、2010年調査の3割に減らす。
一般調査計	7,520	6,931		3,466		
0～1歳未満	2,275	1,593	1,593	796		WHO2006を参考に、2010年調査の7割に減らす。
1～2歳未満	2,333	971	971	2,669		2010年調査は、1～2歳未満の誤差率が小さく、2歳以上特に5歳以上での誤差率が大きいことから、2～6歳半は1歳当たりの人数を1.5倍とし、1～2歳未満はそれに合わせて減らす。
2～6歳半	2,912	4,368	971			

※1：集計人数は体重について検討した。

※2：半数ずつと仮定しシミュレーションで用いた。

表2 乳幼児健診からデータが得られる月齢のみを抽出し、病院調査と学校保健統計調査データを加えたデータセット（昨年度報告）

病院調査	1/10抽出	218
3カ月		91
6か月		116
10か月		101
18か月		104
36か月		31
学校保健66か月	1/50抽出	397
学校保健78か月	1/50抽出	432

表3 表1の乳幼児健診月齢分を2倍にしたデータセット

一般調査相当のデータを倍にしてみた

病院調査	1/10抽出	218
3カ月		182
6か月		232
10か月		202
18か月		208
36か月		62
学校保健66か月	1/50抽出	397
学校保健78か月	1/50抽出	432

(単純に同じデータを2度計算に組み込む)

図1 表2の平滑化結果

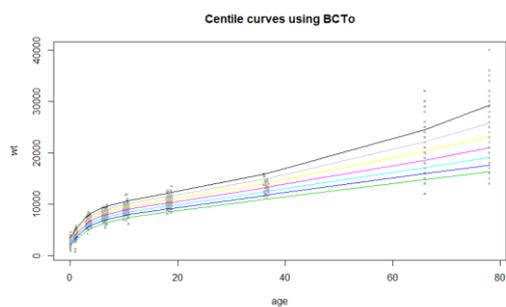


図2 表3の平滑化結果

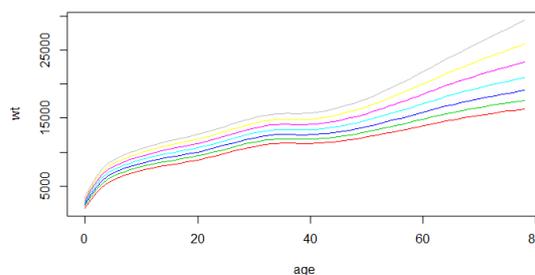


図 3-1

2010 年公表値男子体重データセットから求めた平滑化曲線

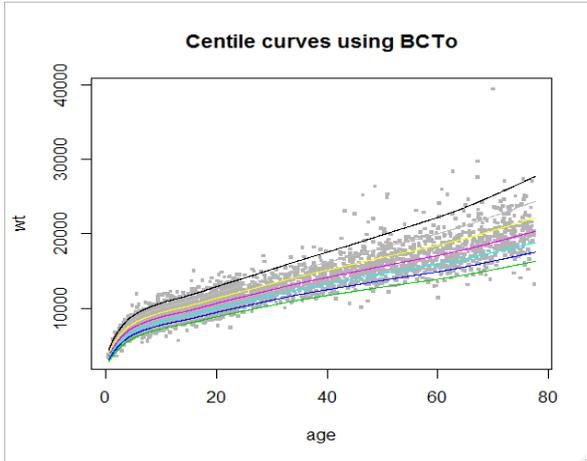


図 4-1

男子体重データを  
全年齢区分で 2 分の 1 にしたデータ  
セットから求めた平滑化曲線

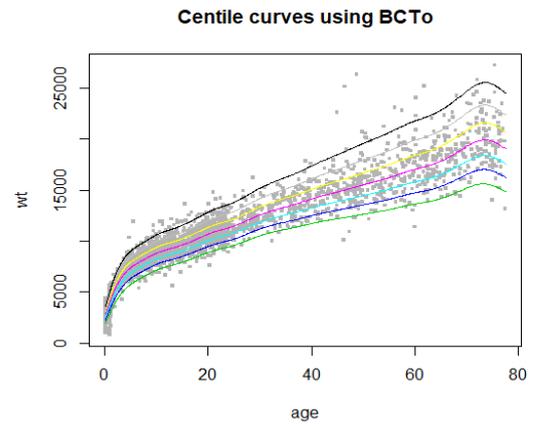


図 3-2 図 3-1 の

z スコアの中央値を推計した場合の SE

Cole2020 の推奨 0.041 を超えない  
全体的に 0.041 より小さい

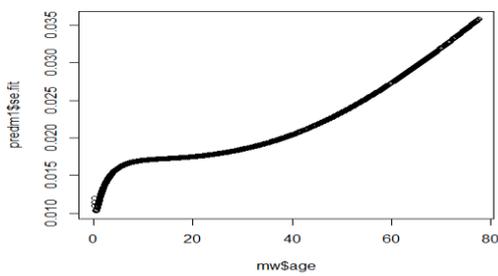


図 4-2 図 4-1 の

z スコアの中央値を推計した場合の SE

5 歳以降は 0.041 を超えて不適

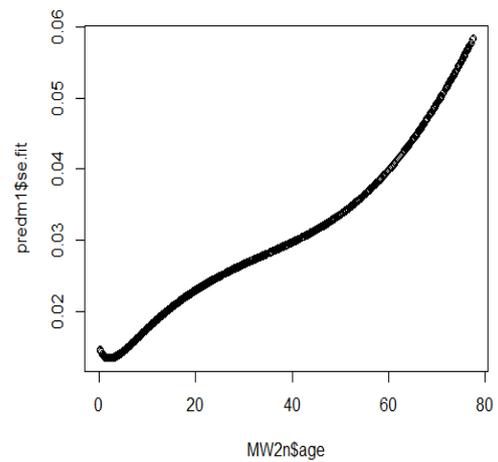


図 5-1

男子体重 病院調査 1/10、0 歳及び 1 歳 0.3 倍、2 歳以降そのままのデータセットから求めた平滑化曲線

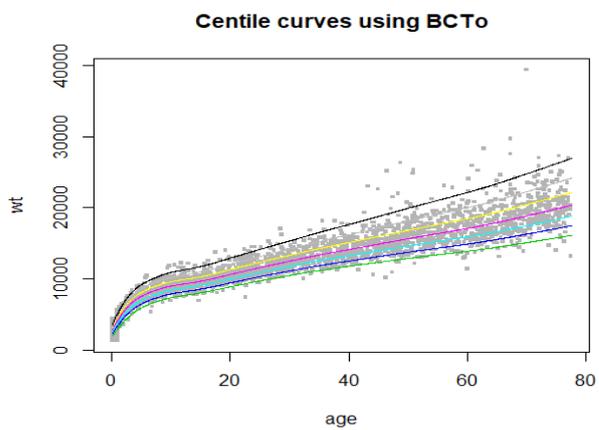


図 5-2 図 5-1 の z スコアの中央値を推計した場合の SE

全体的に 0.041 より小さい

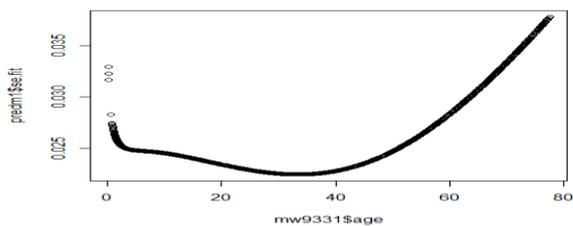
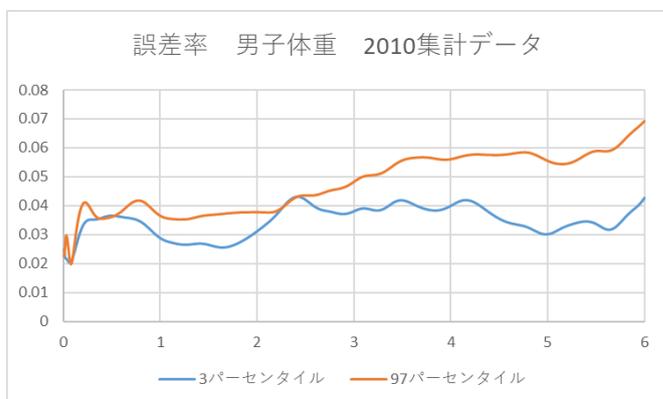
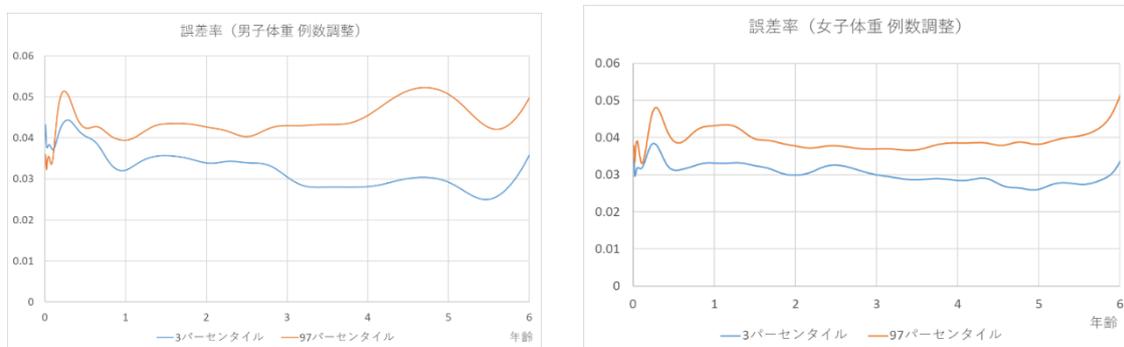


図6 2010年公表値に用いたデータセットで誤差率<sup>\*</sup>を計算



<sup>\*</sup>分母を真の“3パーセンタイル値と50パーセンタイル値”の間隔、真の“50パーセンタイル値と97パーセンタイル値”の間隔とし、“3パーセンタイル値と50パーセンタイル値”の推計値の間隔、真の“50パーセンタイル値と97パーセンタイル値”の推計値の間隔の標準誤差を、それぞれ対応する分母で割ったもの

図7 調査設計の検討時における地区数に相当したサンプリング<sup>\*</sup>をもとに算出した誤差率



※年月齢別のデータ数 2010年公表値に用いたデータセットに対し  
 病院調査 0.3倍 一般調査0歳 0.7倍 一般調査1歳 0.42倍 一般調査2歳以降 1.5倍で  
 誤差率を計算

図 8-1 2010 年集計データ  
平滑化結果に対し  
R 上の Bootstrap を施行(100 回)

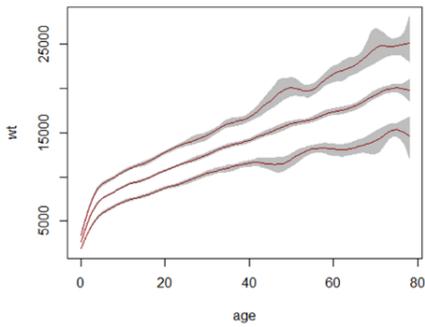


図 9-1 2010 年集計データに対し  
病院調査 0.3 倍 一般調査 0 歳 0.7 倍  
一般調査 1 歳 0.42 倍 一般調査 2 歳以降 1.5 倍  
平滑化結果に対し  
R 上の Bootstrap を施行(100 回)

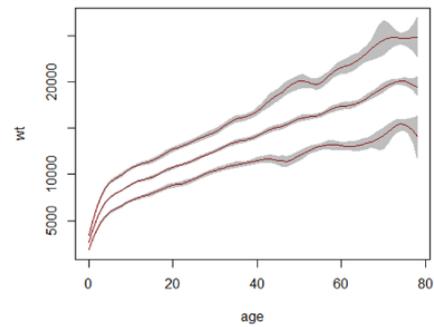
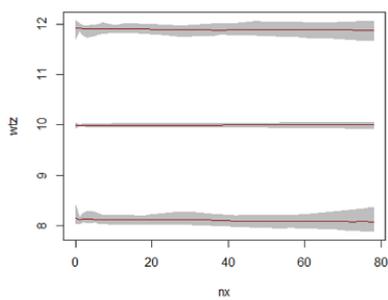


図 8-2 図 8-1 に使用したデータを  
正規変換したものの平滑化結果に対し  
R 上の Bootstrap を施行

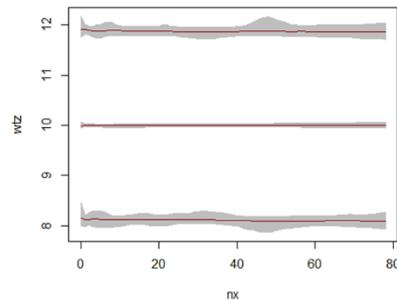
(100 回)



(縦軸は正規変換値に 10 を加算したもの)

図 9-2 図 9-1 に使用したデータを  
正規変換したものの平滑化結果に対し  
R 上の Bootstrap を施行

(100 回)



(縦軸は正規変換値に 10 を加算したもの)

図 8-3 図 8-1 に使用したデータを正規変換したものの Bootstrap による平滑化結果 (100 回)

3,50,97 パーセンタイル値推計 SE

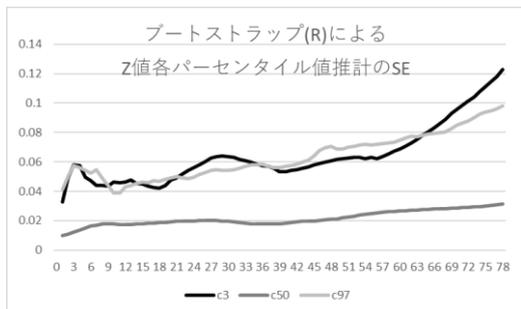


図 9-3 図 9-1 に使用したデータを正規変換したものの Bootstrap による平滑化結果 (100 回)

3,50,97 パーセンタイル値推計 SE

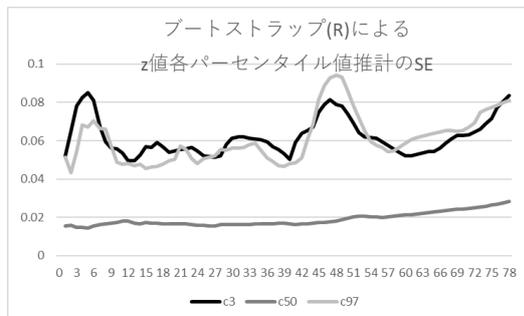


図 10 表 1 に基づくサンプリングによる平滑化結果 (データを正規変換したもの) 平滑化結果の Bootstrap (1000 回) による 3,50,97 パーセンタイル値推計 SE

