

ウツタイン様式に基づく院外心肺機能停止 症例の搬送記録における欠損データの分析 及び対処方法について

(財) 消防科学総合センター
研究員 胡 哲 新

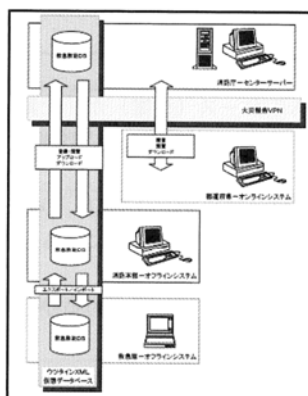
1 はじめに

院外心肺機能停止傷病者の搬送記録について、日本では 1994 年より独自の「救急蘇生指標」を用いたデータ収集を開始し、2005 年 1 月からは、総務省消防庁(以下、消防庁と略す)が世界的に統一された「ウツタイン様式」を用いて、全国の消防本部を通じて国レベルのデータ収集を行ってきた(図 1)。これらのデータを用いて、救急救命処置等による救命効果や、救急救命率の向上に係る施策の確立等に関する検討¹⁾²⁾も行われてきているところである。

一般的に、広範囲にわたる調査結果の有効活用を図るには、調査記録に関わる者の知識レベルをはじめとする種々の要件を「均質に」することが重要な前提の一つとなる。しかし、現在収集しているウツタイン統計データには、例えば、救急活動「時刻」に関する記録の基準となる時計が合わせられていない¹⁾ことや、「バイスタンダーCPRの有無」についての記録者の判断の違いなど¹⁾により、データの均質性が確保されない問題が考えられる。データ分析の視点からすれば、これらのデータが欠損しているという問題を意味している。また、現状のウツタイン統計データには、空白・矛盾等の異常値もあり、不完全なデータといえる。不完全なデータは英語では「missingdata」、日本語では「欠損データ」、「欠測データ」あるいは「欠落データ」と呼ばれているが、本稿では「欠損データ」という表現を使うことにする。

欠損データの問題は、統計学をはじめ医学、政治学など様々な学術分野において、看過できない問題として注目されている。欠損データを無視したり、無根拠に除外したりすると、偏りのある、役に立たない結果が得られる恐れが出てくる。一方、欠損データへの対処法は必ずしも確立してはいない³⁾⁴⁾。分析者各自がその場しのぎ的に対応しているというのが現状であろう。

本稿は、ウツタイン統計データの有効活用にあたり、消防庁が収集しているウツタイン統計データ(以下、「消防庁ウツタイン統計データ」という)における欠損データの発生状況を考察し、欠損データがもたらす問題とその対処法を検討したい。



- ※消防庁は全国の消防本部のデータを閲覧可能
- ※都道府県は管轄内の消防本部のデータを閲覧可能
- ※消防本部は自本部のみのデータを閲覧可能
- ※消防本部によっては、独自のシステムを有しているところがある。

このような本部は、消防庁システムの登録画面からデータを登録していない。独自システムに登録されたデータに対し、消防庁システムのデータベースの構造に整合するように変換ルール設定を行い、データコンバートを行ったものを、消防庁システムに登録している。

図1 消防庁ウツタイン統計データ収集システムの概要

(出典：総務省消防庁、ウツタイン統計活用検討会 報告書、2008年3月)

2 消防庁ウツタイン統計データにおける欠損データの発生状況

(1) 消防庁ウツタイン統計データの概要

消防庁ウツタイン統計データの内容を図2(左)に示す。全国の消防本部から収集したデータをExcel等のソフトにおいて、一番上の行が「変数名」、その下の各行が個々の「ケース」となるように、データベース化することができる(図2右)。このようなデータベースには、主に表1に示す7つの欠損パターンが考えられる。

「ウツタイン様式」に基づく心肺機能停止傷病者記録票

事例No.

発生年月日

年

月

日

性別

□男 □女

年齢

救急隊士乗乗

□あり □なし

医師の乗乗

□あり □なし

医師の2次乗乗

□あり □なし

1. 心停止の目撃

□ 目撃、または警を聞いた

時

分

□ 目撃

□ その他のバスターンダー(□友人 □同僚 □通行人 □その他)

□ 同僚

□ 既に心肺機能停止(発見時)

2. バイスタンダー-CPR

□あり □なし

□心臓マッサージ

□人工呼吸

□両者による除動

□なし

バイスタンダー-CPRまたは両者による除動開始時刻

時

分

□確定 □推定 □不明

□心電図診断

3. 初期心電図診断

□V1心電図診断

□Defib V1(無断性心電図診断)

□PEA(無断性心電図診断)

4. 救急隊員処置等の内容

□除動機(□二相性 □単相性)

初期除動機実施時刻

時

分

秒

進行回数

回

□気道確保

□特定行為隊員使用(□ILM □気道開放式マスク □気管内チューブ)

□静脈経管挿入

□薬剤投与

初期投与時刻

時

分

秒

投与回数

回

5. 病院搬送

未知

時

分

秒

搬送

時

分

秒

CPR開始

時

分

秒

病院収容

時

分

6. 心停止の確定診断

□心原性 □確定 □院外診断による心原性

□非心原性 □脳血管障害 □呼吸器系疾患 □悪性腫瘍 □外因性 □その他()

7. 転送及び予後

病院収容時の心拍再開

□あり □なし

初期心拍再開時刻

時

分

□1ヶ月未満(回復) □あり □なし

□1ヶ月未満(回復) □あり □なし

□脳機能回復(□GCS)

□OPC1 機能良好 □OPC2 中等度障害 □OPC3 高度障害

□OPC4 昏睡 □OPC5 死亡、脳死

□多発臓器障害(□OPC6)

□OPC1 機能良好 □OPC2 中等度障害 □OPC3 高度障害

□OPC4 昏睡 □OPC5 死亡、脳死

変数(データ項目)名

ID	X 1 発生年月日	X 2 目撃の有無	X 3 覚知時刻	X 4 心拍再開の有無	...
1	2006/7/4	1	10:03:00	2	

1つのケース(事案)

1つのデータ

図2 消防庁ウツタイン統計データの内容(左)及びデータベースの構造(右)

表1 消防庁ウツタイン統計データにおける主な欠損パターン

	区 分	欠損パターン	説 明
特定困難 ^{注1}	ケース（事案）ごとの欠損	① ケース数の不均質性	例えば、あるケース（事案）が「記録対象」とされるべきかどうかの理解・判断の違いによるもの
	変数（データ項目）ごとの欠損	② データ値の不均質性	例えば、ある変数（データ項目）の定義に対する理解・判断の違いによるもの
特定可能 ^{注2}	ケース（事案）ごとの欠損	③ 空白ケース	すべてのデータ項目（変数）の値が記録されていないケース
		④ 重複ケース	同一事案を、別事案として重複記録されているケース
	変数（データ項目）ごとの欠損	⑤ 空白データ	ある変数（データ項目）の値が記録されていないデータ
		⑥ エラーデータ	ある変数（データ項目）の値がエラーとなっているデータ
		⑦ 矛盾データ	変数（データ項目）間の値が矛盾しているデータ

注1：記録に関わる者の判断の違い等により、記録データのみでは欠損かどうかの判明が困難なもの。

注2：記録データを分析することにより、欠損の特定が可能なもの。

(2) 欠損データが持つ問題性

欠損データが持つ問題性の本質は、欠けているデータの分量にあるのではなく、欠損の発生メカニズムと密接に関わっている。即ち、欠損が生じている変数とそれ以外の変数との間の因果関係の有無が問題になる。例えば、「重複ケース」という欠損が「消防本部」においてランダムに発生しているか、ランダムでない形で発生しているかによって、その対処法は大きく異なってくるのである。

ランダムに発生する欠損データを除外して分析を行う場合に、ケース数の減少によって分析の精度がやや低下する以上の問題性を持たない。言い換えれば、「無視可能」な欠損である。しかし、ランダムでない形で発生する欠損データを除外して分析を行う場合に、分析結果をゆがめる恐れがあるため、「無視できない」欠損となる。一つの例を見てみると、社会調査におけるサンプリングを行うことが問題にならないのは、データの欠損をランダムに発生させているからである。

(3) 消防庁ウツタイン統計データにおける欠損の発生状況

上記(2)の考え方にに基づき、消防庁ウツタイン統計データ(2007年)を例として、以下の三点を分析、集計し、欠損の発生状況を把握してみた。結果を表2にまとめる。

ア 欠損データの項目

イ 欠損データの件数

ウ 欠損データが母集団に占める割合(欠損率)

表2において、「国単位」、「都道府県単位」の欠損率を考察しているのは、現状の消防庁ウツタイン統計データがテンプレートに従って国及び都道府県単位で集計・公表されているからである。また、「消防本部単位」の欠損率にも着目しているのは、消防庁ウツタイン統計データは、全国の消防本部から収集されており、本部毎の欠損率を把握することにより、欠損発生メカニズム及びその対処法を検討できるからである。

表2に示すように、様々な欠損データの中に、ランダムに発生するものもあれば、ランダムでない形で発生するものもあることが明らかとなった。〈「消防本部単位」でランダムでない形で発生している欠損項目及びそれぞれの欠損率を示す箇所は灰色で表示されている〉

表2 消防庁ウツタイン統計データ（2007年）における欠損状況の一覧

欠損データの種類		該当件数	該当する 都道府県 (消防本部)数	欠損率 (国単位)	最大欠損 率(都道 府県単 位)	最大欠損 率(消防 本部単 位)
＜年間のデータ(ケース)数＞		110,518	47(807)	－	－	－
① ケース数の不均質性 (例:「記録対象」とすべき事案の選取基準の違いによる)		悉皆調査がないため、定量的な分析は困難であるが、「医療機関で既に二次救命処理が行われている」ケースや、「ドクターカー、ドクターヘリに搬送した」ケースなどを「記録対象」とすべきか否かの判断基準は、入力要領(Ver.2.01)には明示されていないため、欠損の可能性があると考えられる。				
② データ値の不均質性 (例:「時刻の入力基準となる時計＜指令室時計、救急車時計、除細動時計など＞が合わせられていない」 ^{注3)})		19,323	42(173)	17.48%	55.57%	100%
③ 空白ケース		0	0	0	0	0
④ 重複 ケース	「消防本部コード」、「性別」、「年齢」、「覚知時刻」を基準としてチェック	1,951	42(122)	1.77%	17.54%	50%
	「年齢」	939	17(31)	0.85%	38.53%	100%
⑤ 空白 データ	「特定行為器具種別」 ^{注4)}	249	2(2)	0.23%	3.63%	23.46%

欠損データの種類		該当件数	該当する 都道府県 (消防本部) 数	欠損率 (国単位)	最大欠損 率(都道 府県単位)	最大欠損 率(消防 本部単位)
⑥ エラーデータ	「発生年月日」	43	23 (32)	0.04%	0.24%	4.55%
	「年齢」	1	1	0.00%	0.04%	0.33%
	「救急救命者乗車」	8	7 (8)	0.01%	0.16%	5.56%
	「目撃時刻」	115	30 (53)	0.10%	0.66%	9.09%
	「CPR開始時刻」	79	23 (40)	0.07%	0.39%	3.13%
	「初回除細動実施時刻」	8	8 (8)	0.01%	0.11%	1.75%
	「初回心拍再開時刻」	5	4 (5)	0.00%	0.11%	1.31%
	「1ヶ月生存」	6	6 (6)	0.01%	0.11%	3.57%
	「医師の乗車」	15	8 (11)	0.01%	0.23%	5.56%
	「医師の2次救命処置」	11	3 (4)	0.01%	0.15%	25.93%
⑦ 矛盾データ	「目撃なし」で「目撃時刻あり」	609	25 (35)	0.55%	6.89%	59.38%
	「目撃なし」で「バイスタンダー種別あり」	69	10 (11)	0.06%	0.87%	20.37%
	「バイスタンダー CPRあり」で「バイスタンダー CPR種別なし」	93	30 (67)	0.08%	0.60%	5.56%
	「バイスタンダー CPRなし」で「バイスタンダー CPRの実施時間の<確定/推定/不明>の選択あり」	1,111	10 (12)	1.01%	14.44%	55.00%
	「除細動なし」で「除細動種別の選択あり」	67	7 (8)	0.06%	0.71%	14.69%
	「除細動なし」で「除細動の施行回数が1回以上」	1,060	7 (8)	0.96%	22.41%	100%
	「気道確保の特定行為器具使用あり」で「器具種別の記載なし」 ^{注5}	249	2 (2)	0.23%	3.63%	23.46%
	「気道確保の特定行為器具使用なし」で「器具種別の記載あり」 ^{注5}	169	20 (24)	0.15%	4.53%	30.34%
	「薬剤投与なし」で「薬剤投与時刻の記載あり」	282	13 (23)	0.26%	12.04%	80.30%
	「薬剤投与なし」で「薬剤投与回数が1回以上」	833	17 (23)	0.75%	15.27%	100%
	「非心原性」で「心原性種別の記載あり」	104	2 (2)	0.09%	1.70%	36.01%
	「心原性」で「非心原性種別の記載あり」	282	3 (3)	0.26%	12.94%	35.53%
	「心拍再開なし」で「心拍再開時刻の記載あり」	13	5 (5)	0.01%	0.32%	2.08%
	「1ヶ月生存あり」で「脳機能カテゴリー CPC5」	720	45 (241)	0.65%	3.68%	95.92%

欠損データの種類		該当件数	該当する 都道府県 (消防本 部)数	欠損率 (国単位)	最大欠損 率(都道 府県単 位)	最大欠損 率(消防 本部単 位)
⑦ 矛盾 デー タ	「1ヶ月生存なし」で「脳機能カテゴリー CPC1-4」	705	45 (219)	0.64%	3.41%	21.77%
	「1ヶ月生存あり」で「全身機能カテゴリー OPC5」	719	45 (240)	0.65%	3.68%	95.90%
	「1ヶ月生存なし」で「全身機能カテゴリー CPC1-4」	776	45 (248)	0.70%	3.60%	21.77%

注3:「消防庁ウツタイン様式データの活用に関するアンケート調査結果(2008年)」^{参考文献1}に基づいて集計したもの

注4:「気道確保」かつ「特定行為器具使用」が選択されている場合

注5:「気道確保あり」が選択されている場合

3 欠損データがもたらす問題の分析例

前節の2(2)では、欠損発生メカニズムによってその問題性が変容するということを述べたが、一般的に、欠損データを含んだ調査結果を手にしてその欠損メカニズムを特定する作業は、必ずしも容易ではない。また、その手法を一般化することも極めて困難なのが現状である^{3) 4)}。

本稿では、何らかの理由で欠損データが存在した場合に、その欠損部分を分析に含めて分析しても、含めずに分析しても、分析結果に大きな違いがなければ、その欠損を「無視可能な欠損」とし、他方、分析結果に大きな違いをもたらす場合、その欠損を「無視できない欠損」とする。それぞれについて、一例を挙げ図3、図4に示す。

図3は「無視可能な欠損」の分析例である。具体的に、表2にある「⑦矛盾データ」の中で、「1ヶ月生存あり」で「脳機能または全身機能カテゴリー5」と「1ヶ月生存なし」で「脳機能または全身機能カテゴリー5」という欠損データを含めた場合と、それらの欠損データを含むケースを取り除いた場合の「都道府県別の1ヶ月生存率」の結果比較を示すものである。図3でわかるように、これらの欠損は分析結果に大きく影響することがなく、「無視可能な欠損」と推測される。

図4は「無視できない欠損」の分析例である。具体的に、表2にある「④重複ケース」について、「重複ケース」を含めた場合と、それらのケースを除外した場合の「都道府県別の発症率」の比較を示すものである。図4に示すように、「重複ケース」による欠損は分析結果(都道府県間の序列)を大きく左右することがわかる。したがって、この種の欠損は「無視できない欠損」となる。

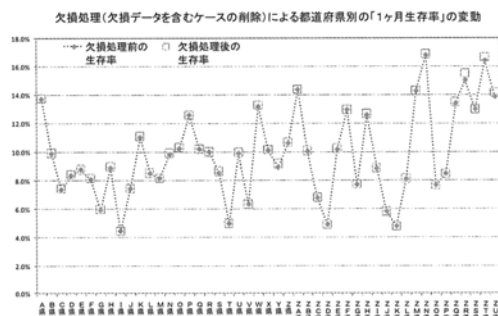


図3 (無視可能な) 欠損データの分析例

※ ここで「1ヶ月生存率」とは、一般市民により心肺機能停止の時点が目撃された心原性の症例数に対して、「1ヶ月生存あり」の症例が占める割合のことである。

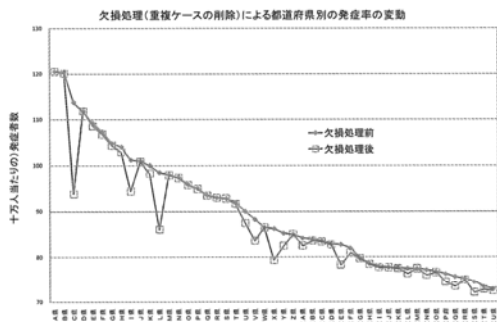


図4 (無視できない) 欠損データの分析例

※ ここで「発症率」とは、都道府県ごとの年間の心肺機能停止症例数が総人口数に占める割合のことである。

※ 図3、図4にある都道府県名の記号は関連性のないものである。)

4 消防庁ウツタイン統計データにおける欠損への対処法

一般的に、欠損データへの対処法としては、主に「欠損データを分析から除外する」、「何らの値で代入する」及び「再調査により欠損を補完する」という3つの方法が挙げられる。

それぞれの長所と短所を表2にまとめる。

消防庁ウツタイン統計データにおける欠損について、分析の対象によって、表2に示す方法の組み合わせを利用すると効果的であろう。例えば、国単位の分析を行う場合に、年間十万件以上の膨大な母集団数に対して、欠損データが微小であれば、分析結果への影響も微小と考えられる

表2 欠損データへの対処法

	長 所	短 所
①除外法： 欠損データを分析から除外する方法	●最も簡単で、複雑な統計分析は必要としない。	●統計分析に使えるデータ数が減ってしまう。 ●解析結果に偏りが生じる可能性がある。
②代入法： 欠損データに何らかの値を代入する方法	●データを無駄にしない。 ●適切に埋め込めれば、解析結果の偏りを防ぐことが可能である。	●適切に代入するためには、かなり専門的な統計知識を必要とする。 ●埋め込み方法によって、偏る分析結果が得られる可能性もある。
③再調査法： 欠損データを補完するために再度調査する方法	●欠損データを埋め込むためのもっとも精度の高い方法である。	●再調査にかかる労力は大きい。 ●すべての欠損を埋め込むことが現実的に困難である。

ため、「除外」しても差し支えないと考えられる(ただし、どの程度までを「微小」と考えるべきかという課題が残る)。また、「再調査」という方法により、一部の欠損データが復元できるとともに、欠損の原因をある程度限定できた場合に、「代入法」を組み合わせると、欠損データを最大限に埋め込むことができる。

5 まとめ

本稿は、2007 年消防庁ウツタイン統計データにおける欠損データの発生状況を考察し、欠損データがもたらす分析上の問題及びその対処法について検討を試みた。欠損データが持つ最大の問題は、欠損が無視できるか否かについて配慮せずに、欠損データを除外してしまうと、分析結果に偏りが生じる可能性がでてくることである。一方、「代入法」または「再調査法」を通じて欠損データを補完するような対処法には、多くの手間と時間がかかる場合があり、「無視可能な欠損」については、「除外法」の簡便性を考慮し、十分に活用することも重要である。即ち、消防庁ウツタイン統計データにおける欠損について、分析事項及び対象母集団などに応じて、表 2 に示す 3 つの対処法の組み合わせを図ることが有効であると考えられる。

どのような対処法にも、それぞれ長所と短所があり、最善の方法は欠損データを生み出さないことである。そのためには、ウツタイン統計に係るデータ記録、システムへの入力等の過程で、あらかじめ欠損が出ないような(ソフト的・ハード的な)工夫をすることが最も重要であろう。

参考文献

- 1) 総務省消防庁、ウツタイン統計活用検討会報告書、2008 年 3 月
- 2) 総務省消防庁、救急統計活用検討会報告書、2009 年 3 月
- 3) 数理社会学会と謝野有紀ほか、社会の見方、測り方計量社会学への招待、2006 年 7 月
- 4) 浅野正彦、欠損データの分析法、選挙学会紀要第 1 号(101-123)、2003 年