

# 実用的な抽出調査を自ら企画してみよう ～途上国での方法を日本に活用するには～

座長：磯 博康（国立健康危機管理研究機構 グローバルヘルス政策研究センター）

蜂矢 正彦（国立健康危機管理研究機構 国際医療協力局）

演者：駒田 謙一（国立健康危機管理研究機構 国際医療協力局）

細澤 麻里子（国立健康危機管理研究機構 グローバルヘルス政策研究センター）

ファシリテーター：宮野 真輔（国立健康危機管理研究機構 国際医療協力局）

本田 真梨（国立健康危機管理研究機構 国際医療協力局）

# 本日の流れ

◆本セッションの趣旨

◆演題1

「開発途上国における無作為サンプリング調査の実例と日本への応用の可能性」

◆グループワーク1：サンプリングを含む調査デザイン

◆演題2：

「現実的に考えられる調査方法と回答率を上げるための取り組み～日本における実例～」

◆グループワーク2：調査実施上の工夫

◆まとめ

# 開発途上国における無作為サンプリング 調査の実例と日本への応用の可能性

駒田 謙一(国立健康危機管理研究機構 国際医療協力局)

# 日本公衆衛生学会 COI 開示

発表者名：駒田 謙一

演題発表に関連し、発表者らに開示すべき  
COI 関係にある企業などはありません。

# 内容

- ・代表的な無作為抽出方法
- ・ラオスにおける実例
- ・日本における活用のポイント
- ・住民台帳が使えない場合のオプション

# 背景

- ・エビデンスに基づいた保健政策が重要であるが、肝心のエビデンスが不足していることも珍しくない  
→ 時には自ら調査する必要あり
- ・途上国で政府主導の疫学調査を行う場合、回答率が高く、比較的厳密な研究デザインを組めることがある。  
→ 日本に活用できるか

# 単純無作為抽出(Simple Random Sampling)

- ・さまざまなサンプリングの基本。
- ・全ての標本が均等に選ばれる可能性がある。  
(100人から10人を選ぶなら、各個人が選ばれる確率は $1/10$ ずつ)  
→ selection biasがかかりにくい
- ・コストと時間の問題で大規模調査で用いるのは難しい。  
(全国から600人を選んで訪問するより、10都道府県で各60人の方が現実的)

# 収束抽出(クラスターサンプリング)

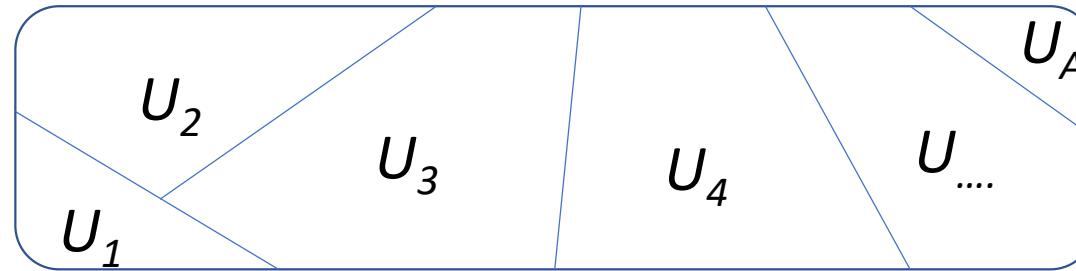
- まず居住区や組織などグループ単位(クラスター)で抽出し、抽出されたクラスターの中から最終的な個人等を抽出する。

例) ある県の有病率を調査する場合に、まず市町村を10個抽出し、その後抽出された市町村でそれぞれ60名を抽出。

- 県を母集団とする2段階サンプリング
- 調査対象が広い地域に散らばっている時に、労力・時間・費用(交通費)を抑えた調査が可能になる。

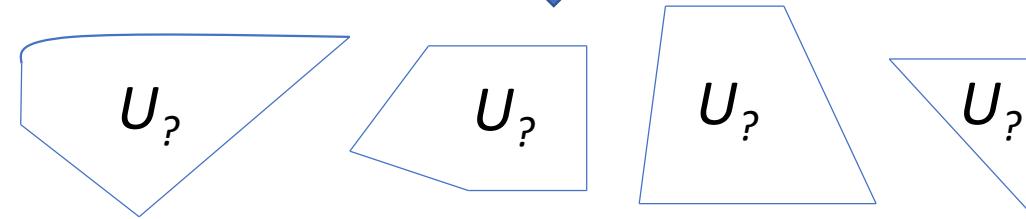
ただし、最終段階のクラスターでは最終単位(個人等)のサンプリングフレーム(台帳)が必要。

母集団  
Aクラスター



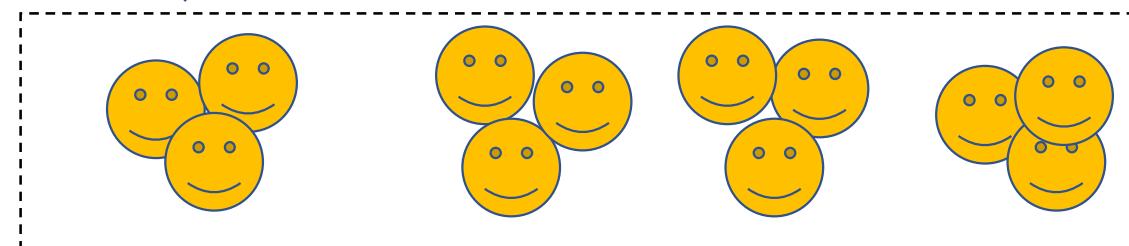
第1段サンプリング

第1次抽出単位  
a単位



第2段サンプリング

第2次抽出単位  
各b人



調査する標本  $a * b$  人

## 参考：サンプルサイズ

$$N \geq \frac{1.96^2 \times p(1-p)}{d^2} \times DEFF \times \frac{1}{R}$$

想定される有病率p: 70%

95%信頼区間の幅d: ±10%

デザイン効果DEFF: 2

回答率R: 80%

とすると、必要サンプルサイズNは、**202**

(母集団の大きさも関係ない)

# ラオスでの実例(2014年)

- ・ ラオス全国における一般人口の麻疹・風疹に対する抗体保有率を調査
- ・ 必要サンプルサイズを2184と推定  
多段階クラスターサンプリングを利用  
第1段階: 全国144郡から26郡を抽出  
第2段階: 各26郡で2村ずつを抽出 ( $26 \times 2 = 52$ 村)  
第3段階: 各52村で42名を抽出  
( $52 \times 42 = 2184$ 名、各村で住民リストをもとに42名を抽出)
- ・ 質問票調査と血液採取、日本で抗麻疹IgG抗体、抗風疹IgG抗体を測定
- ・ 過去の予防接種プログラムの実績と比較し、予防接種の効果を評価

# ラオスでの実例(2014年)

## 確率比例抽出

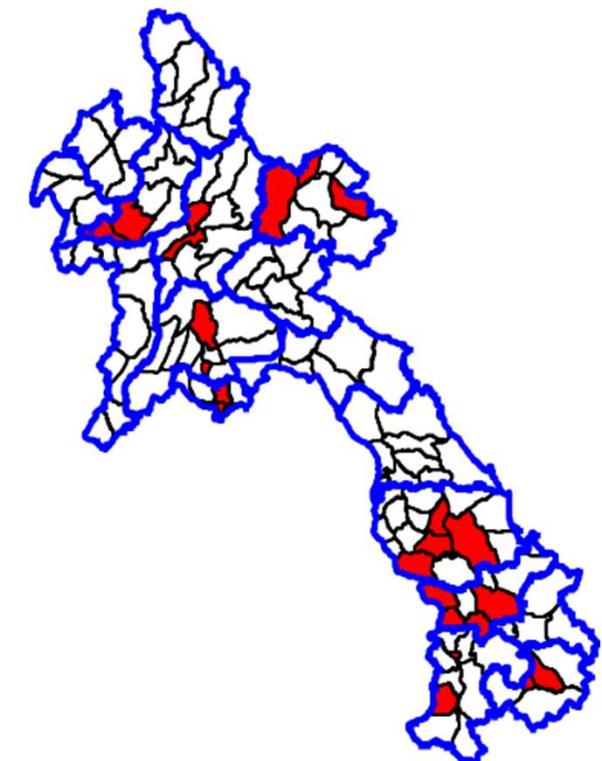
Probability Proportionate to Size (PPS)  
Sampling

<サンプリングインターバルの計算>  
 $60,000,000 \div 26 = 230,769$

<インターバル間で乱数を発生>  
Between 1 and 230,769 → 1,250

<乱数とインターバルを元に26郡を選択>  
District 1: 1,250  
District 2: 1,250+230,769=232,019  
District 3: 232,019+230,769=462,788  
:  
District26: 5,539,706+230,769= 5,770,475

Districts	Population	Cumulative Population
A	50,000	50,000 ← District1
B	150,000	200,000
C	90,000	290,000 ← District2
D	60,000	350,000
E	100,000	450,000
F	400,000	850,000 ← District3 District4
G	20,000	870,000
H	600,000	1,470,000 ← District5 District6
I	700,000	2,170,000
:	:	:
LAST	800	60,000,000 ← District26
合計	6,000,000	



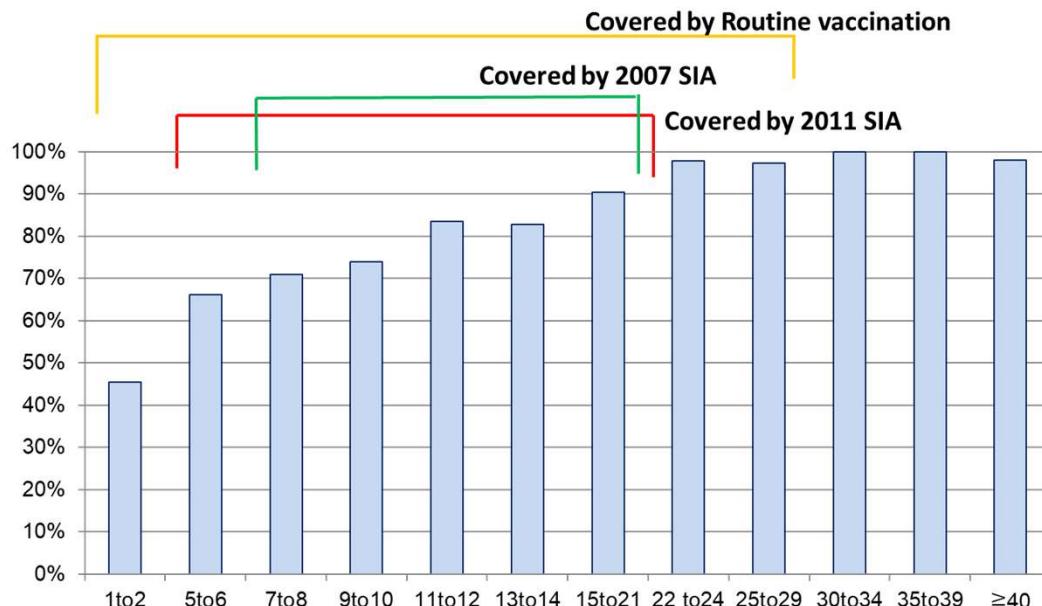
人口の大きなクラスターは選ばれやすく、  
小さなクラスターは選ばれにくい。

大きなクラスターでは各人が選ばれる確率は低く、  
小さなクラスターではその逆。

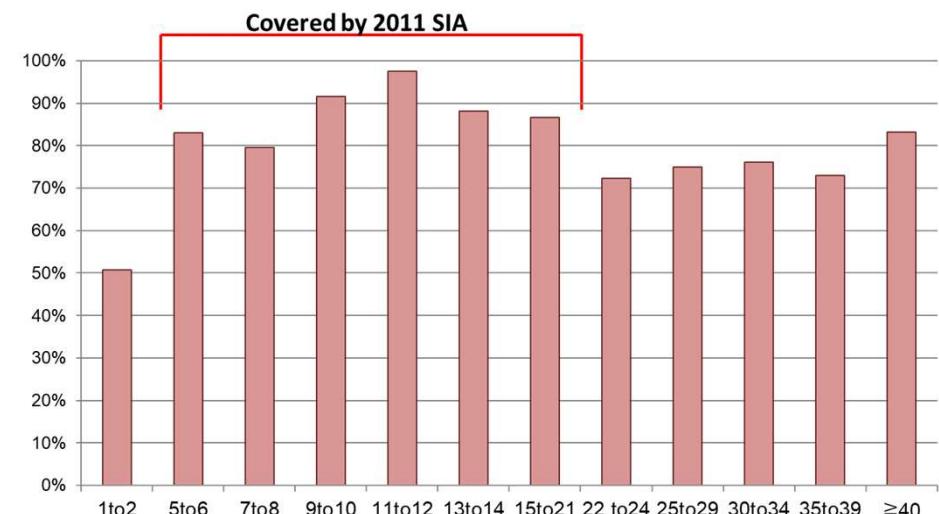
抽出確率を均等化できる。  
→選択バイアスがかかりにくい

# ラオスでの実例 (回答率96.9%)

年齢群別の抗麻疹IgG抗体陽性率



年齢群別の抗風疹IgG抗体陽性率



度重なる接種キャンペーンにも関わらず麻疹抗体の上昇が弱い → 不適切な温度管理によるワクチン活性の低下か  
風疹について、キャンペーン対象外の妊娠可能年齢群の抗体上昇が弱い → 先天性風疹症候群のリスク高

# 日本で実施する場合のポイント

- 何をクラスターとするか
  - 行政区
  - 学校（学生を対象とした調査の場合）
- 何段階でサンプリングを行うか
  - 全国47都道府県で、各7人×30クラスター（市町村）の調査を行う場合
  - 2段階（PSU：市町村）で行う場合、30市町村が47都道府県にばらける
  - 3段階（PSU：都道府県、SSU：市町村）で行う場合、10×3市町村なら、調査地域を10都道府県に絞れる
- 最終段階でのサンプリングをどう行うか
  - 住民台帳は使えるか？ 地図上から行うか？

# 日本で行う際のポイント

- ・人口データ：行政区の統計情報が豊富  
→ 比較的容易に確率比例抽出でのクラスターサンプリングが可能
- ・住民台帳：住民基本台帳には、氏名、住所、生年月日、性別が世帯ごとにまとめられている。
  - 特定の年齢に配慮した個人抽出が可能
  - 利用には大きなメリットがある一方で、制約も大きい
- ・回答率
  - 様々な社会的要因から低くなりがち。

# 住民基本台帳の閲覧要件

## (東京都目黒区の場合、区HPより抜粋)

1. 国、地方公共団体が法令で定める事務の遂行を目的で行うもの
2. 報道機関が世論調査等のために行う場合、その調査結果に基づく報道が行われ、その成果が社会に還元されると認められるもの
3. 学術研究機関が学術研究の用に供するために行う場合、その調査結果またはそれに基づく研究が学会等を通じて公表されることによりその成果が社会に還元されると認められるもの
4. 2、3以外の統計的調査研究の対象者を抽出するために行う場合、その調査結果またはそれに基づく研究が公表されることにより国または地方公共団体における施策の企画・立案や他の機関等における学術研究に利用されることが見込まれるなど、その結果が社会に還元されると認められる特段の事情があると認められるもの
5. 公共的団体が行う地域住民の福祉の向上に寄与する活動のうち、公益性が高いと認められるものを実施するもの
6. 営利以外の目的で行う居住関係の確認のうち、訴訟の提起などの特別な事情によるもので住民基本台帳の写しの閲覧以外には手段がないと認められるもの
7. その他、区長が必要と認める場合

(閲覧に際しての注意事項より抜粋)

閲覧席で、携帯電話等の電子機器の使用ならびに飲食・喫煙はできません。

# 住民台帳が使えない場合(イラクでの実例)

## Mortality before and after the 2003 invasion of Iraq: cluster sample survey

*Les Roberts 1, Riyadh Lafta, Richard Garfield, Jamal Khudhairi, Gilbert Burnham*

*Lancet.* 2004 Nov;364(9448):1857–64. doi: 10.1016/S0140-6736(04)17441-2.

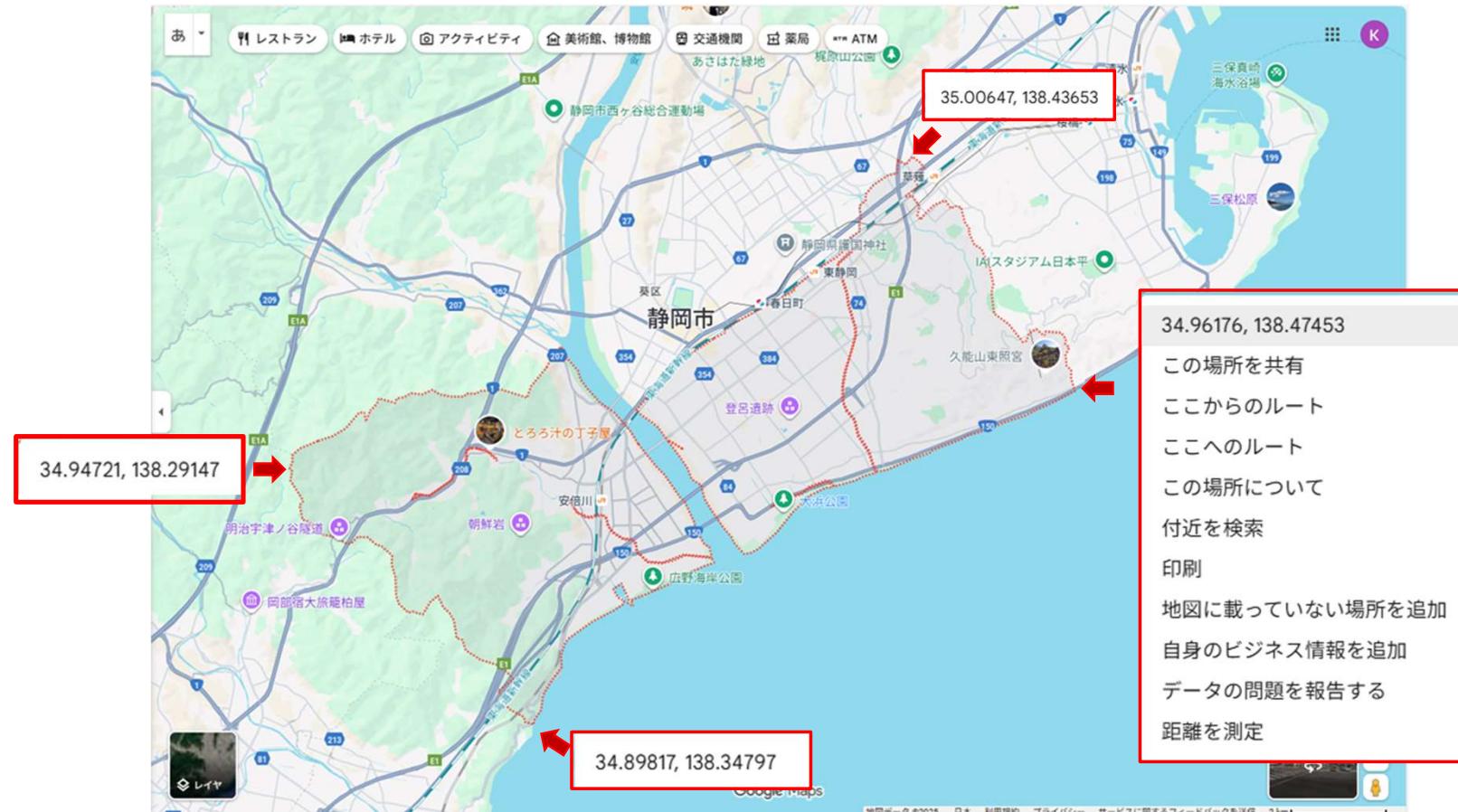
- 目的: 2003年3月のイラク侵攻前後での死亡率の変化を
- 調査方法: イラク全土で33クラスター(各30世帯)を対象に聞き取り調査(2004年9月実施)  
世帯構成、出生、死亡、暴力による死因などを記録
- 主な結果推定: 超過死亡数は少なくとも100,000人以上、死因: 暴力(空爆)が最大

GPS情報を用いて、調査員が現地でクラスター内の最初の世帯を決定。  
最初の世帯を中心に右隣の家を順に訪問し、合計30世帯を調査。

住民リストがないような極限的な状況でも公衆衛生データの収集は可能

# 地図を用いる場合の例(静岡市駿河区)

Google MAP等で、行政区の北端・南端の緯度、東端・西端の経度情報を入手



画像出典: Google マップより(© Google)

# 地図を用いる場合の例(静岡市駿河区)

乱数を取得

緯度: 南端と北端の間

経度: 東端と西端の間

(例: Excelで数式Randbetweenを利用)

当該地点が、該当行政区外にある場合は再度乱数を取得

当該地点から最寄りの住宅を必要数選択

2地点以上を選ぶ場合は、上記作業を繰り返す



画像出典: Google マップより(© Google)

# まとめ

- ・大規模調査では、多段階クラスターサンプリングが現実的
- ・抽出確率をできるだけ均等化することでバイアスを減らすことができる
- ・住民台帳の利用にはメリットも大きいが、制約も大きい
- ・住民台帳が使えない場合は、地図データを利用する方法もある

# グループワーク1

グループ協議(12分)

前方中央5つのテーブル単位でグループを形成

以下について協議

- ① 静岡県レベルで調査を実施すると仮定し、研究テーマを設定
- ② 設定したテーマについて、調査項目・サンプリングデザインを検討
- ③ 発表者を選定

発表(10分)

研究目的、調査項目、サンプリング方法について各グループより発表

# グループワーク2

グループ協議(12分)

グループワーク1でデザインした調査について、  
以下について協議

- ① 実施上の課題と対処方法について協議
- ② 発表者を選定

発表(10分)

想定される課題、考え方について各グループより発表