



24-25 October 2016, Rome
FAO Headquarters

Coverage Issues in Agricultural Surveys

Anders Wallgren & Britt Wallgren, ba.statistik@telia.com

Our background

1. Experiences from Statistics Sweden

We worked at the Department for Research and Development 1989 – 2010

We specialized in register statistics – register surveys

2. Experiences from Latin America, LAC

Since 2011 we have worked for IDB and visited many countries in Latin America and the Caribbean

IDB has a long term project for modernizing the statistical systems in LAC by using administrative data

Coverage issues is a neglected and difficult area

If you have register data, where 21 % of the records have completely wrong values on the main variable:

Would you publish statistics based on these data?

If you have register data, where 21 % of the records have completely wrong values on the main variable:

Would you publish statistics based on these data?

Yes, if the alternative is to use the old census, then you must try to do something!

=> We must develop methods to improve quality of our estimates

Our colleagues in LAC face this problem

The problem:

The administrative systems that generate the registers are not good enough

– updating with new information does not work well

Population Register

*A farmer gives up farming
and moves to a city*

Farm Register

Land registration

Business Register

All these *register populations* change – the registers need frequent updating!

The problem with the Population Registers in LAC:

Birth registration, registration of deaths	OK
International migration	OK
Migration within the country	Very bad!

Population Register

*A farmer gives up farming
and moves to a city*

Farm Register

Land registration

Business Register

Statistical registers replace censuses – regional statistics is important. Migration gives coverage errors for municipalities

Problem: Migration within the country causes over- and undercoverage in statistics for municipalities

Example: One Swedish municipality:

Overcoverage, 23 %	619	}	}	2 700, Population according to Registro Civil
	2 081			2 386, Population according to the truth
Undercoverage, 13 %	305			

Population register 2005 = “Registro Civil today” but not updated

Population register 2015 = “The truth today”

We want to test an idea from ISI 2015 with real data:

Calibration of weights in a register

How to estimate the truth today with the badly updated register?

Problem: Migration within the country causes over- and undercoverage in statistics for municipalities

Solution? There are three methods that should be combined:

1. Improve the administrative system – the national registration system of persons. This is necessary, *but will take years*

Example: One country has 4 systems for PIN, id-numbers for persons, 4 ministries/authorities that do not cooperate!
The id-cards have 4 id-numbers

This will generate coverage problems

Example:

One country has 7 institutions that create population registers:

- Registro Civil, national registration of persons
- Authority for id-cards, have their own register

⇒ 2 institutions register the population

The Population Register should handle the following demographic events:

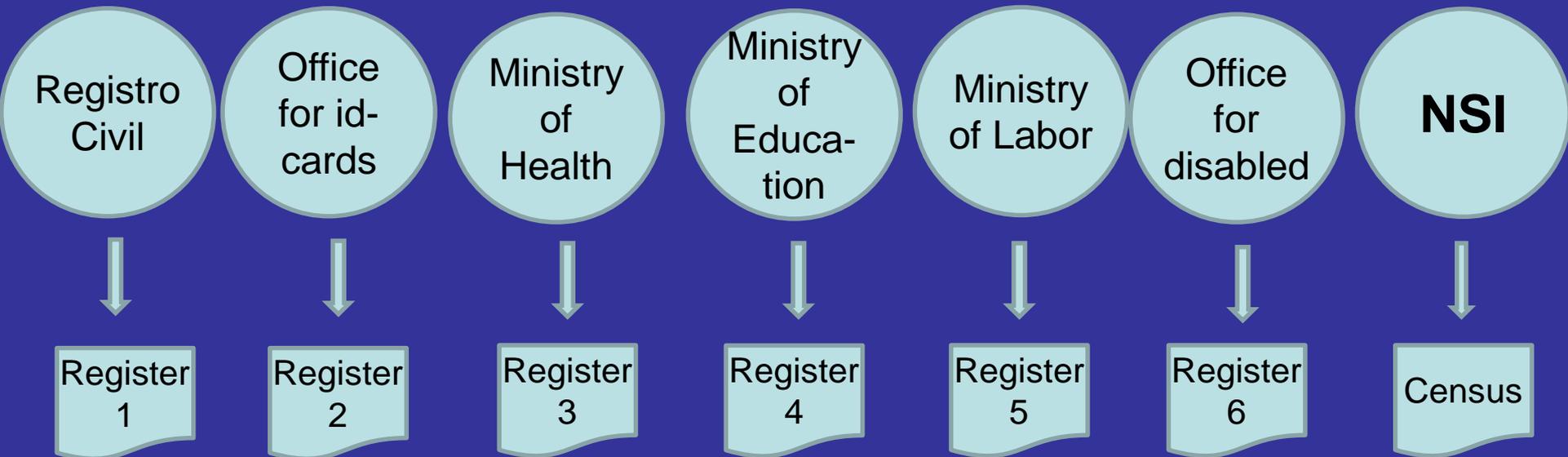
- | | |
|--------------------------------|----------------|
| • Births, | 2 institutions |
| • migration within the country | 0 institutions |
| • international migration | 1 institutions |
| • deaths | 2 institutions |

Demographic events: 5 institutions

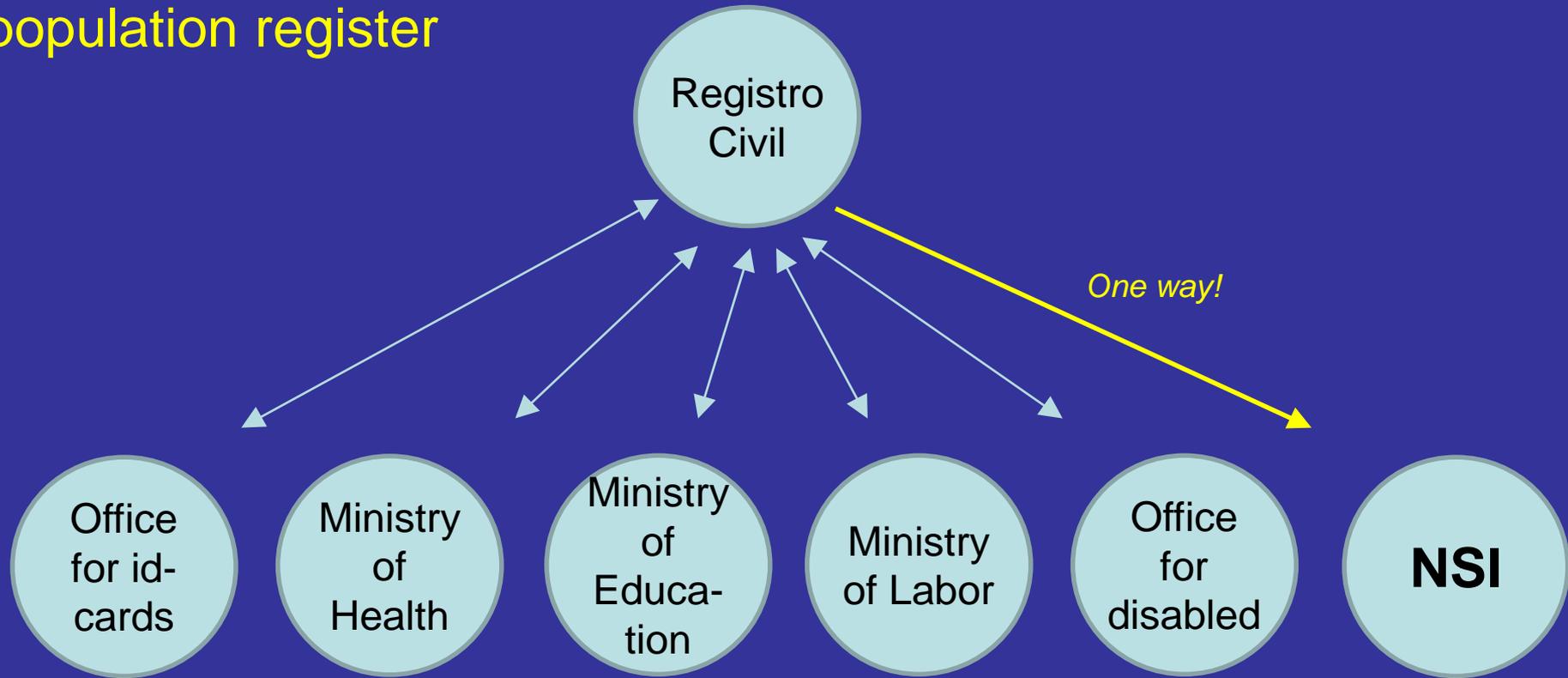
In all, **7 institutions**

Double work, inconsistencies, coverage problems

**In Latin America: No cooperation,
No exchange of microdata
Double work
Inconsistencies
Coverage problems**



In Sweden: Our "Registro Civil" creates *the* national population register



All Ministries, national institutes, ... that need the Population Register get access and can use it in their work and add their variables

When a citizen contacts an institution, on-line check is done, if errors are found, Registro Civil is informed

Problem: Migration within the country causes over- and undercoverage in statistics for municipalities

Solution? There are three methods that should be combined:

1. Improve the administrative system – the national registration system of persons. This is necessary, *but will take years*
2. Combine the *administrative* population register with other sources and create a *statistical* population register with better quality (multiple frames in a register context)
3. Develop estimation method where the statistical register is combined with an area frame based sample survey (dual frame approach in a register context) *Test old idea from ISI 2015*

To test the method we need realistic data. We have used Statistics Sweden's Population Registers (PR) for 2005 and 2015, all persons belonging to these registers both years

⇒ Only migration within Sweden (new municipality of residence)

PR2005 = RC today 21 % wrong municipality PR2015 = Truth today

Registro Civil today				The truth today			
PIN	Age	Sex	Municipality	PIN	Age	Sex	Municipality
1000001	6064	2	2085	1000001	6064	2	2085
1000002	1519	2	1481	1000002	1519	2	1481
1000003	2529	2	1466 Wrong	1000003	2529	2	1480
1000004	1519	1	764	1000004	1519	1	764
1000005	1014	1	1982	1000005	1014	1	1982
1000006	1519	1	662	1000006	1519	1	662
1000007	6064	2	180	1000007	6064	2	180
1000008	2024	2	1283	1000008	2024	2	1283
1000009	2529	1	486 Wrong	1000009	2529	1	180
1000010	3034	2	191	1000010	3034	2	191

Estimator 1: Use Registro Civil as it is:

Muni.code	RC:est1	True values	abs error %	
114	34 027	32 921	3.4	
115	25 248	26 205	3.7	-
117	34 318	33 991	1.0	
120	32 199	33 485	3.8	
123	55 882	55 446	0.8	
125	22 207	21 719	2.2	
126	80 491	77 739	3.5	
127	69 059	62 748	10.1	+
...	

Variable	Mean	Q3	Maximum
abs error %	4.4	6.4	16.8

In all: 290 municipalities

Estimator 2: Combine RC and Sisben (dual frame approach)

PIN	Age	Sex	Muni_RC	Muni:Sisben	Muni:RC+Sb
1000001	6064	2	2085	0	2085
1000002	1519	2	1481	1481	1481
1000003	2529	2	1466	0	1466
...
1000050	2529	2	1435	0	1435
1000051	2024	2	136	1864	1864
1000052	7000	1	1764	0	1764
1000053	7000	1	1491	0	1491
1000054	6064	2	1080	1485	1485
...

Muni.code	RC+Sb:est2	True values	abs error %
114	33512	32921	1.8
115	25728	26205	1.8
117	34153	33991	0.5
120	32768	33485	2.1
123	55589	55446	0.3
125	21904	21719	0.9
126	79115	77739	1.8
127	65938	62748	5.1

Variable		
abs error %		
Mean	Q3	Maximum
2.2	3.2	9.0

Estimator 3: Combine Registro Civil and an area sample (dual frame approach)

Use the area sample to get calibration conditions

The 290 municipalities are grouped into 10 categories, from the most rural to the most urban, these categories is an auxiliary variable we hope is strongly correlated with net migration:

Category	RC, register	Area sample	weight
1	351 107	317 396	0.90399
2	268 733	249 748	0.92935
3	340 143	320 935	0.94353
4	385 406	369 257	0.95810
5	577 247	558 744	0.96795
6	765 929	748 678	0.97748
7	936 860	929 009	0.99162
8	1 011 517	1 014 497	1.00295
9	2 037 960	2 080 777	1.02101
10	1 288 365	1 374 226	1.06664
All	7 963 267	7 963 267	

Calculate 10 estimates with the area sample,

Calculate 10 estimates with Registro Civil:

Population in the 10 categories

Calculate weights

to adjust for under- and overcoverage

			Mun	Cat	Muni	Cat	
PIN	Age	Sex	RC	RC	true	true	weight
1000001	6064	2	2085	5	2085	5	0.96795
1000002	1519	2	1481	7	1481	7	0.99162
1000003	2529	2	1466	4	1480	9	0.95810
1000004	1519	1	764	1	764	1	0.90399
1000005	1014	1	1982	5	1982	5	0.96795
1000006	1519	1	662	1	662	1	0.90399
1000007	6064	2	180	10	180	10	1.06664

The 290 municipalities are grouped into 10 categories, from the most rural to the most urban, *category* is an auxiliary variable we hope is strongly correlated with net migration

Muni.	Cat.	Cal.est:3	True values	abs error %
114	5	32 936	32 921	0.0
115	10	26 931	26 205	2.8
117	7	34 030	33 991	0.1
120	10	34 345	33 485	2.6
123	7	55 414	55 446	0.1
125	6	21 707	21 719	0.1
126	5	77 911	77 739	0.2
127	1	62 428	62 748	0.5
...

Variable	Mean	Q3	Maximum
abs error%	0.6	0.6	6.0

Estimator 4: Combined register Registro Civil + Sisben register and an area sample (multiple frame approach, 3 frames)

Area sample gives calibration conditions: Auxiliary variable, **category**

Estimator 5: Combined register Registro Civil + Sisben register and an area sample (multiple frame approach, 3 frames)

Area sample gives calibration conditions: Auxiliary variables, **category and age**, 7 age classes (in all 70 estimates from the sample):

Calibration exercise:	Mean absolute error %	3 rd quartile	Maximum abs. error, %
1. <u>Registro Civil</u> , not updated for migration, (RC)	4.4	6.4	16.8
2. RC supplemented with Poverty Register (<u>Sisben</u>)	2.2	3.2	9.0
3. RC, Calibrated weights, 10 categories	0.6	0.6	6.0
4. RC and <u>Sisben</u> , calibrated weights, 10 categories	0.4	0.5	3.3
5. RC and <u>Sisben</u> , cal. weights, 10 cat. and age groups	0.4	0.5	3.8

Estimator 5: Combined register Registro Civil Sisben register and an area sample (multiple frame approach, 3 frames)

Use the area sample to get calibration conditions:

Auxiliary variables, *category* and *age*, 7 age classes: From the area sample we need 70 estimates to be compared with corresponding estimates from the combined register Registro Civil + Sisben

Stockholm municipality (category 1)

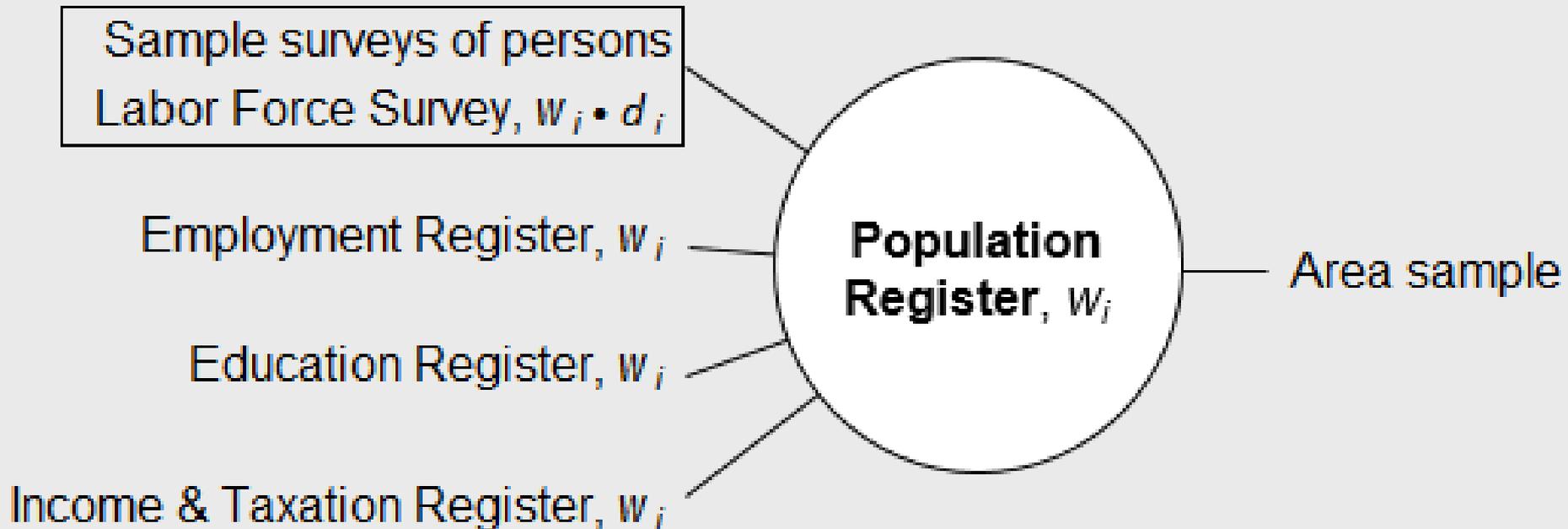
Age	RC	Est. 5	Truth
10-19	11.6	10.6	10.3
20-29	13.4	15.8	17.0
30-39	15.5	17.6	17.9
40-49	17.7	16.3	15.6
50-59	15.3	14.8	14.7
60-69	12.8	12.3	12.1
70+	13.7	12.8	12.5
All	100.0	100.0	100.0
Population	670 474	717 621	718 008
Error %	-6.6	-0.1	

Åsele municipality (category 2)

Age	RC	Est. 5	Truth
10-19	10.7	11.2	11.1
20-29	14.6	10.3	11.2
30-39	10.4	7.8	8.6
40-49	10.5	10.7	10.5
50-59	14.3	15.3	15.1
60-69	16.1	18.2	17.8
70+	23.5	26.6	25.6
All	100.0	100.0	100.0
Population	2 700	2 433	2 386
Error %	13.2	2.0	

Statistics Sweden's production system:

These four registers are perfectly consistent, same populations and no conflicting variable values and the LFS is rather consistent



An integrated system of surveys

When there are no coverage problems, $w_i = 1$

When there are coverage problems, use $w_i \neq 1$

Gives coherent estimates

How to implement this method?

- Attitudes:
Quality of microdata \neq Quality of estimates
- Attitudes:
No tradition on using weights in registers
- Model-based estimation:
Skill is required for building the model (10 categories)

But, you only need to do better than the old census!