Sample Design Summary: ESS Round 11

Country:	**Italy** (IT)
NC:	Sveva Balduini (s.balduini@inapp.org)
Other Experts:	Valentina Gualtieri (v.gualtieri@inapp.org)
Survey Institute:	INAPP (inapp.org)
Sampling Expert:	Olha Lysa (olha.lysa@ut.ee)
Country Contact:	May Doušak (May.Dousak@fdv.uni-lj.si)
Reference Survey:	ESS10
Date:	*17 December* 2022

Status:

- [ ] Pre sign-off
- [x] Signed off
- [ ] Post sign-off amendment
- [ ] Final (post-field work)

1.1 Target Population

Number of residents aged 15 or older in the country:	51,506,455
Source and reference date:	Istat (http://dati.istat.it/) , 1 January 2022

1.2 Population Coverage

The frame for selecting persons is the National Register of Resident Population (NRRP), which includes all people resident in Italy. The NRRP is owned by Ministry of the Interior. The information contained in the list and related to each individual are municipality of residence (and the address), date of birth, and sex. Using the date of birth, it is possible to build the sampling frame for all people aged 15 year or more (a person is treated as 15 or older if she or he is 15 at the 1st of July 2023).

2. Summary of the Sample Design

A two-domain sampling design is used.

The first sampling domain consists of the biggest municipalities within Italy (around 15% of the Italian population). Here a one-stage sampling design is used, where persons are sampled by stratified simple random sample from the National Register of Resident Population from each municipality in domain 1. The allocation of the sample size is proportional to target population within the strata.

The second sampling domain consists of all municipalities that are not included in the first domain. Here a two-stage sampling design is used. At the first stage municipalities are selected as Primary Sampling Units (PSUs) by stratified sampling. The stratification is done by crossing two variables, Geographical area (NUTS-1) and demographic size class of the resident population aged 15 and over (4 classes). The allocation of PSUs to strata is proportional to target population within the strata. Within strata, PSUs are selected with probability proportional to size of target population. At the second stage persons are selected from the sampled municipalities using a simple random sample.

3. Sample Design Details

3.1 Domain 1

First Sampling Stage

unit:	Persons
frame:	National Register of Resident Population
size:	443
strata:	12 Municipalities: Roma, Milano, Napoli, Torino, Palermo, Genova, Bologna, Firenze, Bari, Catania, Verona, Venezia
allocation:	Proportional to the target population size within the strata.
	Implicit stratification is done within strata by ordering the strata sampling frame by age and gender before selection.
algorithm:	Systematic sample with frame ordered by gender and age before selection

3.2 Domain 2

First Sampling Stage

unit: Municipalities

frame: List of all 7,891 municipalities within the domain

size: 217

strata: Strata are defined by crossing two variables; Geographical area (NUTS-1 level, namely 5 geographic areas: North-West, North-East, Center, South, Islands) and size of the resident population aged 15 and over (4 classes: less than 2,000; 2,001 - 10,000; 10,001 - 50,000 persons; 50,001 and more), according to the standard classification used by Italian National Institute of Statistics (ISTAT) (see Table: Target Population Size by Strata in Domain 2)

allocation: Proportional to the target population size within the strata (see Table: Allocation of Municipality Sample in Domain 2)

algorithm: PPS Sampford's Method

Second Sampling Stage

unit: Persons

frame: National Register of Resident Population

size: 25 persons per PSU (5425 in total)

strata: NA

allocation: Implicit stratification is done within PSU by ordering the PSU sampling frame by gender and age before selection.

algorithm: Systematic sample with frame ordered by gender and age before selection

Remarks

None

4. Planning the Sample Size

History of Planned and Realised Values

ESS	d	$p.\bar{b}$	$\bar{b}$	$p.\rho$	ρ	$p.\text{Deff}$	Deff	$p.\text{Deff}_c$	Deff_c	$p.\text{Deff}_p$	Deff_p
1		4.0	10.98	0.03	0.108	1.20	1.93	1.09	1.67	1.10	1.16
2		14.6	14.7	0.02	??	1.540	1.529	1.273	1.274	1.21	1.20*
6	1	NA	NA	NA	NA	1.00	1.000	1.00	1.00	1.00	1.000
6	2	14	6.45	0.02	0.092	1.26	1.501	1.26	1.501	1.00	1.000
6	t	NA	NA	NA	NA	1.225	1.434	1.225	1.434	1.00	1.000
8	1	NA	NA	NA	NA	1.00	1.000	1.00	1.00	1.00	1.000
8	2	13.0	14.3	0.06	0.10	1.72	2.326	1.72	2.322*	1.00	1.002
8	t	NA	NA	NA	NA	1.619	2.022	NA	NA	NA	NA
9	1	NA	NA	NA	NA	1.00	1.000	1.00	1.00	1.00	1.000
9	2	14.25	14.59	0.06	0.10	1.795	2.363	1.795	2.351	1.00	1.005
9	t	NA	NA	NA	NA	1.684	1.999	NA	NA	NA	NA
10	1	NA	NA	NA	NA	1.00	1.000	1.00	1.00	1.00	1.000
10	2	12.01	12.01	0.095		2.046		2.046		1.00	1.003
10	t	NA	NA	NA	NA	1.889		NA	NA	NA	NA

Planned and Benchmark Values of the Sample Sizes

ESS	d	$p.\text{rr}$	rr	$p.\text{ri}$	ri	$p.\text{n}_{gross}$	n_{gross}	$p.\text{n}_{net}$	n_{net}	$p.\text{n}_{eff}$	n_{eff}
1		0.68	0.437	0.02	0.08	3,000	3,000	2,000	1,207	1,667	626
2		0.62	0.602	0.052	0.028	2,588	2,613	1,521	1,529	988	1,000
6	1	-	-	NA	NA	329	327	-	128	NA	NA
6	2	-	-	NA	NA	2,451	2,451	-	832	NA	NA
6	t	0.60	0.36	0.101	0.043	2,780	2,778	1,890	960	1,543	669
8	1	0.45	0.398	0.02	0.029	769	770	340	298	NA	298
8	2	0.45	0.506	0.02	0.027	4,727	4,727	2,084	2,328	NA	1,001
8	t	0.45	0.491	0.02	0.027	5,496	5,497	2,424	2,626	1,497	1,299
9	1	0.4	0.491	0.029	0.032	770	770	299	366	NA	366
9	2	0.505	0.524	0.027	0.039	4,727	4,727	2,323	2,379	NA	1,006
9	t	0.490	0.519	0.027	0.038	5,497	5,497	2,621	2,745	1,557	1,372
10	1	0.490	0.548	0.032	0.058	486	486	231	251	231	251
10	2	0.500	0.509	0.039	0.056	4,975	4,975	2,390	2,389	1,168	1,164
10	t	NA	NA	NA	NA	5,461	5,461	2,621	2,640	1,387	1,394

Parameters of the Planned Gross Sample Size

Domain	Achieved interviews per cluster ($\bar{b}$)	Intraclass Correlation Coefficient (ρ)	Design Effect due to Selection Probabilities (Deff_p)	Response Rate (rr)	Ineligible Rate (ri)	Effective Sample Size (n_{eff})	Domain weight (γ)
1	NA	NA	1.00	0.548	0.058	229	0.152
2	12.012	0.095	1.00	0.509	0.056	1,274	0.848
t	NA	NA	1.00		0.056	1,503	

Design Effect: Domain 1

$$\text{Deff}_c = 1.000^*$$

$$\text{Deff}_p = 1.000^*$$

$$\text{Deff} = \text{Deff}_p \times \text{Deff}_c$$
$$= 1.000^*$$

Design Effect: Domain 2

$$\text{Deff}_c = 1 + (\bar{b} - 1) \times \rho$$
$$= 1 + (12.012 - 1) \times 0.095$$
$$= 2.046$$

$$\text{Deff}_p = 1.000^*$$

$$\text{Deff} = \text{Deff}_p \times \text{Deff}_c$$
$$= 2.046$$

*results have been rounded to 3 d.p.s

Design Effect: Overall

$$\text{Deff} = (0.152 \times 1.000) + (0.848 \times 2.046)$$
$$= 1.887^*$$

Gross Sample Size

		Domain 1	**Domain 2**
Min. n_{net}	=	$\text{Deff} \cdot n_{eff}$	
	=	1.000×229	2.046×1.271
	=	229**	2.601**
------------	---	-------------------------------------	-------------------------------------
Min. n_{gross}	=	$\dfrac{n_{net}}{\text{rr} \times (1 - \text{ri})}$	
	=	$\dfrac{229}{0.548 \times (1 - 0.058)}$	$\dfrac{2{,}601}{0.509 \times (1 - 0.056)}$
	=	443**	5.425**
------------	---	-------------------------------------	-------------------------------------
Planned n_{gross}	=	443	5,425
------------	---	-------------------------------------	-------------------------------------
Total n_{gross}		5,868	

** *results have been rounded to 0 d.p.s.*

Remarks

None

5. Sampling Design Data File (SDDF)

Variables to be included in the SDDF

Variable	Description
idno	Personal identifier
cntry	Country (IT)
prob1	Domain 1: Inclusion probability of persons within municipalities. Domain 2: Inclusion probability of municipalities
prob2	Conditional inclusion probability of persons within selected municipalities (only for domain 2)
psu	Domain 1: Respondent identification number. Domain 2: Municipality identification number
stratex1	Domain 1: Municipality identification number. Domain 2: An identification variable of the cross-classification of NUTS2 area and demographic size class (20 values)
stratim1	Order of selection of person within strata (only for domain 1)
stratim2	Order of selection of person within PSU (only for domain 2)
strtval1	Value on first stratification variable (age)
strtval2	Value on second stratification variable (gender)
outcome	Final outcome (1 Data in main data file, 2 Eligible non-respondent, 3 Ineligible)
domain	Sampling domain identifier (values 1 and 2)
BIRTH	Date of birth
CITIZENSHIP	Country of citizenship
MALE	Number of men aged 15 or over in the municipality
FEMALE	Number of women aged 15 or over in the municipality
AGE_15_24	Number of people aged 15-24 in the municipality
AGE_25_34	Number of people aged 25-34 in the municipality
AGE_35_44	Number of people aged 35-44 in the municipality
AGE_45_54	Number of people aged 45-54 in the municipality
AGE_55_64	Number of people aged 55-64 in the municipality
AGE_65	Number of people aged 65 or over in the municipality
DENSITY	Population density of the municipality
EMPLOYEE	Number of employees aged 15 or over in the stratum
SELF_EMPLOYED	Number of self-employed people aged 15 or over in the stratum
UNEMPLOYED	Number of unemployed people aged 15 or over in the stratum
INACTIVE	Number of economically inactive people aged 15 or over in the stratum
EDU_LEV_0	Number of people aged 15 or over not holding a primary level diploma
EDU_PRIMARY_1	Number of people aged 15 or over holding only a primary level diploma

EDU_LOW_SEC_2	Number of people aged 15 or over holding a lower secondary level diploma
EDU_UPP_SEC_3C	Number of people aged 15 or over holding a 2-3 years upper secondary level diploma
EDU_UPP_SEC_3AB	Number of people aged 15 or over holding a 4-5 years upper secondary level diploma
EDU_TER_5	Number of people aged 15 or over holding a tertiary level diploma
FOREIGNERS	Number of people aged 15 or over in the municipality with foreign nationality
ITALIANS	Number of people aged 15 or over in the municipality with Italian nationality

Probabilities of Selection

1. Sampling Domain
    1. $\text{PROB1}_{i|h1} = \frac{n_{h1}}{N_{h1}}$
2. Sampling Domain
    1. $\text{PROB1}_{i|h2} = m_{h2} \frac{N_{ih2}}{N_{h2}}$
    2. $\text{PROB2}_{i|h2} = \frac{q}{N_{ih2}}$

- $\text{PROB1}_{i|h1}$ = Value of *PROB1* for all persons in the h-stratum in domain 1
- $\text{PROB1}_{i|h2}$ = Value of *PROB1* for all persons in the h-stratum in domain 2
- $\text{PROB2}_{i|h2}$ = Value of *PROB2* for all persons in the h-stratum in domain 2
- n_{h1} is the number of persons selected in the h-th stratum in domain 1
- N_{h1} is the size of the target population in the h-th stratum in domain 1
- m_{h2} number of municipalities selected in the h-th stratum in domain 2
- N_{h2} the size of the target population in the h-th stratum in domain 2
- N_{ih2} the size of the target population in the i-th PSU in the h-th stratum in domain 2
- q is the number of persons selected in each PSU in domain 2

Remarks

None

Appendix

Municipalities by Domains and Strata

| Domain | Nuts 1 | Pop. Size classes | | | | Total |
		1-2000	2001-10000	10001-50000	50001 and more	
1	North-west				3	3
	North-east				3	3
	Center				2	2
	South				2	2
	Islands				2	2
2	North-west	1757	1003	215	17	2992
	North-east	464	705	197	19	1385
	Center	393	378	175	22	968
	South	885	628	241	27	1781
	Islands	345	314	94	12	765
	Total	3844	3028	922	109	7903

Target Population Size by Domains and Strata

| Domain | Nuts 1 | Pop. Size classes | | | | Total |
		1-2000	2001-10000	10001-50000	50001 and more	
1	North-west				2449846	2449846
	North-east				799430	799430
	Center				2736917	2736917
	South				1065671	1065671
	Islands				798145	798145
2	North-west	1414948	4459784	4203104	1334593	11412429
	North-east	496557	3362700	3306577	2135061	9300895
	Center	397632	1777692	3562974	1817054	7555352
	South	868643	2796597	4871993	2084534	10621767
	Islands	333305	1447034	1979520	1006144	4766003
	Total	3511085	13843807	17924168	16227395	51506455

Allocation of Municipality Sample by domains

Domain	Nuts 1	Pop. Size classes				Total
		1-2000	2001-10000	10001-50000	50001 and more	
1	North-west				3	3
	North-east				3	3
	Center				2	2
	South				2	2
	Islands				2	2
2	North-west	7	22	21	7	57
	North-east	2	17	16	11	46
	Center	2	9	18	9	38
	South	4	14	24	10	52
	Islands	2	7	10	5	24
	Total	17	69	89	54	229

Allocation of Person Sample in Domain 1

Nr	Municipality	Pop. Size	Proportion	Sample size
1	Torino	749284	9.5%	42
2	Genova	499848	6.4%	28
3	Milano	1200714	15.3%	68
4	Verona	226410	2.9%	13
5	Venezia	226700	2.9%	13
6	Bologna	346320	4.4%	20
7	Firenze	325034	4.1%	18
8	Roma	2411883	30.7%	136
9	Napoli	786411	10.0%	44
10	Bari	279260	3.6%	16
11	Palermo	542098	6.9%	31
12	Catania	256047	3.3%	14
	Total	7850009	100.0%	443