Rivista Italiana di Economia Demografia e Statistica *Volume LXXV n. 2 Aprile-Giugno 2021*

RISK FACTORS FOR CONTAGION OF SARS-COV-2: A STATISTICAL COMPARISON

Alessandra Battisti, Chiara Coluccia, Maria Clelia Romano, Simona Staffieri,
Marco D. Terribili[1]

1. Introduction

The outbreak of COVID-19, caused by the novel infectious disease agent *Severe Acute Respiratory Syndrome Corona Virus 2* (SARS-CoV-2), began around December 2019 in Wuhan Hubei Province China. In March 2020, the World Health Organization declared the COVID-19 pandemic. Italy was one of the first Western countries severely affected by the coronavirus pandemic, with the first devastating wave affecting mainly the northern and central regions of the country, and afterwards spreading nationwide.

Since the start of this pandemic, it was clear that the range of disease manifestations and immune responses, which occur after infection with SARS-CoV-2, vary significantly. Many individuals present with either asymptomatic or mild disease. Therefore, the true extent of the pandemic may be underestimated. Globally researchers and public health organizations explored strategies to better understand the spread of SARS-CoV-2 disease, using seroprevalence as a critical measure. A seroprevalence survey uses antibody tests to estimate the percentage of people in a population who, at a specific time point, have antibodies against SARS-CoV-2. The results can tell us how many people in a specific population may have been previously infected with SARS-CoV-2 (Larremore *et al.*, 2021).

Just to estimate the seroprevalence rate in Italy, the Health ministry, and the National Institute of Statistics (Istat) launched a seroprevalence survey of the SARS-CoV-2 virus, carried out between May and July 2020, in collaboration with the Italian Red Cross. In this paper, the aims pursued by this study and the main achieved results will be shown. Through descriptive analysis and statistical models, the factors playing the main roles in having a positive IgG outcome have been studied.

---

[1] The article is the result of a collective work, but M. C. Romano wrote paragraphs (par.) 1 and 2, S. Staffieri par. 3, C. Coluccia par. 3.1, 3.2, 3.3, A. Battisti par. 4 and 6, M. D. Terribili par. 5.

2. Seroprevalence survey in Italy

Considering the urgent need for reliable and complete epidemiological studies, the statistics on the immune status of the population, and to guarantee protection from the health emergency inherent in the SARS-CoV-2 virus infection, the Ministry of Health promoted a population seroprevalence survey, carried out with Istat. The realization of the study was foreseen by the decree-law 10 May 2020 n. 30, which defines the main study's goals and the roles of all the involved institutions.

The main objective of the study was to evaluate the antibody response reached a few months after the start of the pandemic testing a representative sample of the population for the presence of specific SARS-CoV-2 antibodies in serum and finding the fraction of asymptomatic or subclinical infections. The study, therefore, aimed also to:

(i) evaluate the seroprevalence rate for SARS-CoV-2 in the population and the differences by age group, sex, region, economic activity, and other risk factors;
(ii) assess the development of the antibody response following the first wave of the pandemic and the subsequent period;
(iii) have a population biological bank for further evaluations.

The sample consists of 150,000 individuals and it was designed by Istat to guarantee representativeness, both at a national and regional level. The sampling design is a two-stage selection with stratification of both the Primary Stage Units (PSUs) and the Secondary Stage Units (SSUs). PSUs are the municipalities stratified within each province according to the demographic size (about 2,000, almost 25.0% of the Italian municipalities). SSUs are individuals stratified by age, sex and economic activity (Istat, 2020, 2021a).

The survey methodology envisaged the realization of the fieldwork in three successive phases. In the first phase, the survey units have been contacted by telephone for availability to participate in the survey and to fill out a short questionnaire. During the telephone was also set up an appointment for the blood test (second phase) to find the anti-SARS-CoV-2 antibodies (IgG). The third phase concerned the result of the test, the transmission of the relative outcome, and the delivery of the collected samples to the biological bank of the National Institute for Infectious Diseases "L. Spallanzani".

The complexity of the survey joined with a large number of actors involved, and the steady attention to containing the risk of non-sampling error in each phase of data collection, led to adopt some measures as an accurate training of the interviewers' network, a simple questionnaire focused only on a few indispensable variables, fieldwork supervision through a detailed monitoring system. All these adopted measures made it possible to guarantee a high quality of collected data.

The results presented are related to 65,000 respondent individuals, whose blood samples were collected in time with the end of the survey. The survey management in emergency conditions did not allow to fully reaching the whole sample, so total non-response treatment and calibration techniques have been used to correct the bias and inefficiency introduced by non-respondent units (Istat, 2021b).

Total non-response treatment aims to increase the respondents' sampling weight to consider also non-respondents. Then, calibration allows the final estimates to respect totals on the whole population, known for a set of available variables, such as age class, sex, geographical area, educational level.

3. Observed seroprevalence rates

During the first pandemic wave, estimates pointed out that about 1.5 million individuals (2.5% whole resident population) tested positive to the IgG (IgG+). Positive people are six times those officially noticed, and the same occurred in several other countries (Bajema et al., 2020).

There are marked differences among geographical areas: Lombardy is the region with the highest seroprevalence ratio (7.4%). The case of Lombardy is unique: this region alone includes 49.4% of people who developed antibodies. In the regional ranking, after Lombardy follows the Valle d'Aosta (3.7%) and Piemonte (3.5%), and a group of regions around 3%: Trento, Bolzano, Liguria, Emilia-Romagna, and Marche. Eight Regions, all the South except for Abruzzo (1.5%), have a seroprevalence rate of less than 1.0%, with the lowest values in Sicily (0.4 %), Calabria and Sardinia (0.5%).

Looking at the demographic size of the municipality of residence, the lowest seroprevalence rate is recorded both in municipalities with more than 50,000 inhabitants (1.9%) and in the suburbs of the metropolitan area (2.1%). In small municipalities with up to 2,000 inhabitants, the seroprevalence rate doubles, reaching 4.4%. In Lombardy, there are 14.1% of positives people in small municipalities and 4.7% in the municipality of Milan. Men and women were equally affected by SARS-CoV-2, but with slightly higher seroprevalence among women (2.6% versus 2.4% for men). As for age, seroprevalence remains substantially stable, but with the highest value between 50 and 59 years (3%) and among people 60-69 years (2.7%). Similar results emerged also from other studies, such as Stringhini et al. (2020). Concerning the level of education, the lowest rate is observed for people with at least a Bachelor, Master, PhD (2.1%) (Table 1).

Table 1 – *Sars-CoV-2 Seroprevalence rates and confidence intervals by gender, age class, education qualification and employment condition (2020) absolute values and percentages.*

Variables	**IgG positive outcome**				
	% of people with the same characteristics	lower extreme 95% confidence interval	upper extreme 95% confidence interval	% of people tested positive	absolute values
Sex					
Male	2.6	2.3	2.8	52.7	791,407
Female	2.4	2.2	2.7	47.3	710,130
Age classes					
up to 17	2.2	1.7	2.8	13.8	207,105
18-34	2.2	1.8	2.6	15.5	232,864
35-49	2.4	2.0	2.9	20.4	306,371
50-59	3.0	2.6	3.5	19.1	287,165
60-69	2.7	2.3	3.2	13.7	205,809
70 and over	2.4	2.0	2.9	17.5	262,223
Educational qualification					
No qualification, or primary education	2.5	2.2	2.8	26.9	403,582
Lower secondary education	2.6	2.3	2.9	28.7	431,307
Upper secondary education	2.5	2.3	2.7	33.1	497,724
Bachelor, Master, PhD	2.1	1.7	2.6	11.3	168,923
Employment condition (a)					
Employed	2.7	2.5	3.0	52.9	621,317
Other conditions	2.3	2.1	2.6	47.1	699,236
Total	**2.5**	**2.4**	**2.6**	**100.0**	**1,501,530**

(a) data referring to the population aged 15 and over Source: Istat-Ministry of Health, Survey of the seroprevalence survey on SARS-CoV-2, Year 2020

3.1. Contacts and type of relationship

The results of the survey also made it possible to observe the relationship's type of contact with positive persons: 17.4% of those who report having had contact with a person with SARS-CoV-2 then tested positive.

The highest values correspond to cases in which the contacts concerned cohabiting family members. Those who had contact with a cohabiting family member infected with SARS-CoV-2 developed antibodies in 42.1% of cases; the prevalence is lowered to 16.1% if the family member is not cohabiting, however, remaining properly above the average value that characterizes the entire population (2.5%). A substantial increase in prevalence is also observed when there have been contacts with work colleagues affected by the virus or with patients in the same condition (11.2%).

3.2. Symptoms

It was also possible to reconstruct a picture of the main symptoms reported by positive people: 31.3% of the people who developed antibodies did not have any symptoms, in perfect analogy with what has been observed in other countries (Pollán *et al.*, 2020). In addition to the asymptomatic, the remainder of those who have had symptoms is divided between people with one or two symptoms representing 20.3% and 10.5% respectively and people with at least three symptoms (37.9%).

As found also in other studies (Grant *et al.*, 2020), the most common symptoms in symptomatic subjects include fever (34.7%), flu syndrome (27.6%), fatigue (27.3%), cough (27.0%), loss of taste (25.2%). Some symptoms are more associated with positivity in the seroprevalence survey. Out of 100 people who presented the symptom of loss of taste, 27.7% tested positive; similarly, out of 100 people who presented the symptom of loss of smell, 25.5% were positive. This confirms the discriminating power of these symptoms, unlike other more generic ones. For example, only 5.2% of those who experienced a sense of fatigue tested positive.

3.3. Lifestyle

The results were also analyzed concerning some indicators relating to the lifestyle. The seroprevalence rate for overweight adults is 2.7%, compared to 2.5% for the rest of the population.

Having travelled to a foreign country starting from 1 February 2020 has slightly increased the risk of contagion. The seroprevalence rate is equal to 2.8 among those who have made at least one trip since the beginning of the pandemic compared to those who have not travelled (2.5%). For smokers, as in other studies (Wenzl, 2020), the seroprevalence rate stands at 1.5%, rises to 2.8% for non-smokers and 3.0% among former smokers. This data can be better interpreted given the models' results which will be presented in the next paragraph.

4. Multivariate analysis through a Logit model

A logistic model has been applied to detect which factors played the main roles in having a positive IgG outcome.

The logistic regression analysis is used mostly to investigate the relationship between binary or ordinal response probability and P explanatory variables. The method usually fits linear logistic regression models for binary or ordinal response data by the method of maximum likelihood (Agresti, 2018).

In logistic regression, there are many exogenous variables, many predictors and one criterion variable. If P denotes the number of independent variables, then the equation that describes the relationship between the independent variables and the dependent variable can be written as

$$logit(\Pr(Y_i = 1 \mid X = x)) = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_P X_{Pi} + e_i \,, \qquad (1)$$

where $\beta_1, \beta_2, ..., \beta_P$ are the partial regression coefficients and reflect the fact that each of the considered predictors $X_1, X_2, ..., X_P$ provides a partial explanation (or prediction) of the endogenous binary variable Y, observed on the i^{th} statistical unit.

The interpretation of the coefficients β depends on sign and entity. The sign expresses the type of relationship existing between the independent variable and the dependent variable: when the coefficient is positive it indicates a positive relationship, so the variables tend to covariate in the same direction, when it is negative it indicates that the variables have an inverse relationship. The entity is evaluated based on the magnitude of the coefficient.

A stepwise logistic model was applied in this paper to confirm the results described in the previous paragraph. The logistic model has been applied to detect which factors played the main roles in having a positive IgG, the independent variables were educational qualification, economic activity, people with whom he/she have had contact, employment status, body mass index, smoking, number of chronic diseases, type of chronic disease, citizenship, region.

In all regions the probability of having a positive IGG is higher than that of Sicilian citizens: however, while in Lombardy the probability that a citizen had a positive IGG outcome in the period May-July 2020 is approximately 16 times higher than that of Sicilian citizens, the ratio drops to about 8 to 1 for people who lived in Piemonte and Valle D'Aosta.

Living with an infected person means that the probability of being positive is 24 times that of those who are not in the same situation. If it is a non-cohabiting family member, the ratio, while remaining high, drops to 5 to 1. Similar effect, although more subdued in the case of intercourse with other infected people (e.g.: colleagues or patients).

The educational level seems to play also a key role, having achieved a high educational level (Bachelor, Master, PhD) is confirmed to be a protective factor. People with a Bachelor, Master, PhD significantly reduces the risk of getting infected. Individuals who attended compulsory school have a 57% higher probability of testing positive than graduates.

Having foreign citizenship, under the same conditions, doubles the risk of being infected.

Economic activity, employment status, do not have any effect on the probability of having an IGG positive

It has been introduced in the model some variables regarding the health status too: the number and the type of chronic disease. All the variables do not affect the probability of being positive, with the only exception of cancer pathologies which halve the probability of being infected. This result could be due to a more prudential behaviour adopted by people with oncological pathologies.

As a proxy of lifestyle, the body mass index and smoking habit information have been included. Only smoking habits affect being IGG positive. The logistic regression shows a probability of half for smokers to get infected.

The strong regional effect and some of the results listed before suggest that a multilevel model should be applied.

Table 2 - *Logistic model: odds ratio, regression coefficients, std. error and p-value.*

Variables	**Mode**	**Logistic model**			
		Odds ratio	**Coeff.**	**Std. error**	**p-value**
Intercept			-7.37	0.3909	<.0001
Region of residence	Piemonte	8.15	2.10	0.2400	<.0001
	Valle d'Aosta	7.85	2.06	0.5209	<.0001
	Lombardia	16.03	2.77	0.2293	<.0001
	Bolzano	6.04	1.80	0.3390	<.0001
	Trento	6.91	1.93	0.3277	<.0001
	Veneto	4.21	1.44	0.2477	<.0001
	Friuli-Venezia Giulia	2.48	0.91	0.3567	0.0108
	Liguria	7.12	1.96	0.2665	<.0001
	Emilia-Romagna	5.74	1.75	0.2432	<.0001
	Toscana	2.07	0.73	0.2820	0.0098
	Marche	6.15	1.82	0.2754	<.0001
	Lazio	2.56	0.94	0.2585	0.0003
	Abruzzo	3.21	1.17	0.3273	0.0004
	Campania	2.51	0.92	0.2627	0.0005
	Puglia	2.24	0.81	0.2788	0.0038
	Sicilia	Reference mode			

Table 2 - *Logistic model: odds ratio, regression coefficients, std. error and p-value – continued.*

Variables	Mode	Logistic model			
		Odds ratio	**Coeff.**	**Std. error**	**p-value**
Age classes	1-5	Reference mode			
	6-10	3.05	1.12	0.2349	<.0001
	11-17	2.47	0.90	0.2569	0.0004
	18-34	3.78	1.33	0.3154	<.0001
	35-49	4.23	1.44	0.3138	<.0001
	50-59	5.45	1.70	0.3136	<.0001
	60-69	5.27	1.66	0.3154	<.0001
	70-84	4.54	1.51	0.3137	<.0001
	85 and over	4.04	1.40	0.3594	0.0001
Educational qualification	No qualification/Primary and lower secondary education	1.57	0.45	0.0930	<.0001
	Upper secondary education	1.33	0.28	0.0916	0.002
	Bachelor, Master, PhD	Reference mode			
Citizenship	Italian	Reference mode			
	Stranger	1.92	0.65	0.0764	<.0001
Person with whom he/she have had contact	Cohabiting family member-Yes	24.27	3.19	0.1100	<.0001
	Cohabiting family member-No	Reference mode			
	Non-cohabiting family member-Yes	5.23	1.65	0.1407	<.0001
	Non-cohabiting family member-No	Reference mode			
	Colleague-Yes	2.88	1.06	0.1750	<.0001
	Colleague-No	Reference mode			
	Patient-Yes	3.67	1.30	0.1923	<.0001
	Patient-No	Reference mode			
	Another person-Yes	3.96	1.38	0.1227	<.0001
	Another person -No	Reference mode			
Smoke	Yes	0.51	-0.68	0.0836	<.0001
	No, never smoke	Reference mode			
Cancer	Yes	0.53	-0.63	0.1697	0.0002
	No	Reference mode			

5. Introducing the regional effect with the multi-level model

The multilevel model can be considered the proper statistical tool to study a regression model on data hierarchically nested in level, such as students in class, patients in hospitals or, as in our case study, individuals living in Italian regions.

The fundamental idea underlying these models is the introduction of a further intercept β_{0j}, linked to the level j, which could absorb the variability not attributable to the units' characteristics, but to the j^{th} level to which they belong (Snijders and Bosker, 2011).

When the dependent variable Y is binary, we can formalize this kind of model with this generic formula:

$$logit(\Pr(Y_{ij} = 1 \,|\, X = x)) = \beta_{0j} + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \cdots + \beta_P X_{Pij} + e_{ij} \quad (2)$$

Actually, multilevel models are often called random-effects models too, because we could be not interested in the specific value of each intercept, regarding every j^{th} level, but just to their functional distribution.

In our case, we studied a multilevel model to isolate the regional distribution of the independent variables (X) introduced in the model.

Conversely to the simple logistic model previously described, in the table below regional coefficients have not been estimated; their contribution to the Y explanation is considered as intercept instead of covariates.

Table 3 - *Multilevel logistic model: odds ratio, regression coefficients, std. error and p-value.*

Variables	**Mode**	**Logistic model**			
		Odds ratio	**Coeff.**	**Std. error**	**p-value**
Intercept			-6.07	0.4476	<.0000
Age classes	1-5	Reference mode			
	6-10	3.18	1.16	0.2380	0.0000
	11-17	2.49	0.91	0.2616	0.0005
	18-34	3.69	1.31	0.3233	0.0001
	35-49	4.12	1.42	0.3217	0.0000
	50-59	5.44	1.69	0.3215	0.0000
	60-69	5.24	1.66	0.3237	0.0000
	70-84	4.53	1.51	0.3214	0.0000
	85 e più	4.05	1.40	0.3680	0.0001
Educational qualification	No qualification/ Primary and lower secondary education	Reference mode			
	Upper secondary education	0.84	-0.17	0.0674	0.0107
	Bachelor, Master, PhD	0.63	-0.46	0.0981	0.0000

Table 3 - *Multilevel logistic model: odds ratio, regression coefficients, std. error and p-value – continued.*

Variables	Mode	Logistic model			
		Odds ratio	**Coeff.**	**Std. error**	**p-value**
Person with whom he/she have had contact	Cohabiting family member-Yes	24.24	3.19	0.1146	<.0000
	Cohabiting family member-No	Reference mode			
	Non-cohabiting family member-Yes	5.51	1.71	0.1459	<.0000
	Non-cohabiting family member-No	Reference mode			
	Colleague-Yes	3.15	1.15	0.1820	0.0000
	Colleague-No	Reference mode			
	Patient-Yes	3.70	1.31	0.2058	0.0000
	Patient-No	Reference mode			
	Another person-Yes	3.97	1.38	0.1290	<.0000
	Another person -No	Reference mode			
Smoke	No information	Reference mode			
	Yes	0.80	-0.22	0.3100	0.4767
	No, but I have smoke in the past	1.78	0.58	0.3100	0.0627
	No, never smoke	1.62	0.48	0.3100	0.1162
	Not applicable	2.06	0.72	0.3800	0.0577
Cancer	Don't know/No information	Reference mode			
	Yes	0.46	-0.78	0.2698	0.0039
	No	0.88	-0.13	0.2130	0.5552
Citizenship	Italian	Reference mode			
	Stranger	1.94	0.66	0.0772	<.001

Applying a multilevel model, which introduces a regional random effect, the variable regarding smoking habit and having a cancer pathology completely loses its significance. In other words, neutralizing the smokers and the cancer pathologies' distributions among regions, variables have no more a statistically significant effect on the IGG outcome. On the contrary, every other evidence pointed out by the simple logistic model are confirmed.

6. Conclusions and implications

The Seroprevalence survey is referred to a period, from 25 May to 15 July, during which the epidemy had strong territorial characterization. It estimates 1.5 million individuals (2.5% whole resident population) tested positive to the IgG (IgG+), six times those officially intercepted by the Italian National Institute of Health (ISS).

The use of multivariate statistics and in particular the application of the model (logistic and multilevel) helps to clearly interpret the relationship between the variables studied and the seroprevalence rate.

The multilevel model, compared to the logistic one, allowed to control the territory, introducing a random effect of the region. It confirmed the strongest relationships for several variables as type of contact, educational qualification and citizenship and it highlighted the loss of significance for other variables such as smoking and cancer pathologies. Both models confirmed the absence of a relationship with variables related to health conditions and occupational status.

The results of the multilevel model made it possible to have a complete picture of the relationships and in particular to keep under control the variables strongly linked to the territory, characterized by significantly different levels of diffusion of the phenomenon, especially in the reference period of the survey (May-July 2020).

References

AGRESTI A. 2018. *An introduction to categorical data analysis.* New York, NY: John Wiley & Sons.

BAJEMA K.L., WIEGAND R.E., CUFFE K., PATEL S.V., IACHAN R., et al.. 2021. Estimated SARS-CoV-2 Seroprevalence in the US as of September 2020. *JAMA Intern Med.* Vol. 181, No. 4, pp. 450-460.

GRANT M.C., GEOGHEGAN L, ARBYN M., MOHAMMED Z., MCGUINNESS L., et al.. 2020. The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): A systematic review and meta-analysis of 148 studies from 9 countries. *PLoS One.* Jun 23.

ISTAT. 2020. *Primi risultati dell'indagine di sieroprevalenza sul SARS-CoV2*, https://www.istat.it/it/archivio/246156

ISTAT. 2021a. *Indagine di sieroprevalenza sul SARS-CoV2: tavole di dati*, https://www.istat.it/it/archivio/256536

ISTAT 2021b. *Impianto statistico-metodologico dell'indagine di sieroprevalenza sul SARS-CoV2*, Roma https://www.istat.it/it/files//2020/05/Nota-Metodologica_Sieroprevalenza.pdf

LARREMORE D.B., FOSDICK B.K., BUBAR K.M., ZHANG S., KISSLER S.M., et al.. 2021. Estimating SARS-CoV-2 seroprevalence and epidemiological parameters with uncertainty from serological surveys. *Elife*. 2021.

POLLÁN M., PÉREZ-GÓMEZ B., PASTOR-BARRIUSO R., OTEO J., et al.. 2020. ENE-COVID Study Group 2020, Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet*. Vol. 22, No. 396, pp. 535-544.

SNIJDERS T.A., BOSKER R.J. 2011. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage.

STRINGHINI S., WISNIAK A., PIUMATTI G., AZMAN A.S., LAUER S.A., et al.. 2020. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. *Lancet*. Aug 1;396(10247): 313-319.

WENZL T. 2020. Smoking and COVID-19: Did we overlook representativeness? *Tobacco Induced Diseases*, 18 (November), 89. https://doi.org/10.18332/tid/129584

SUMMARY

Risk factors for contagion of SARS-COV-2: a statistical comparison

From 25 May to 15 July 2020, Istat and the Ministry of Health carried out a seroprevalence survey on SARS-CoV-2. The survey aimed to understand how many people developed antibodies to the SARS-CoV-2, even in the absence of symptoms. The survey collected all the necessary information to estimate the infection size in the population and to describe its frequency about several sociodemographic factors.

The sample consists of 150,000 individuals and it was designed by Istat to guarantee representativeness, both at a national and regional level. The results presented in the paper are related to 65,000 respondent individuals, whose blood samples were collected. The survey management in emergency conditions did not allow to fully reaching the whole sample, so post-stratification techniques were used to correct the distortion factors.

Estimates pointed out that about 1.5 million individuals (2.5% whole resident population) tested positive to the IgG (IgG+) having developed antibodies to SARS-CoV-2. The number of positive people is six times those officially intercepted by identification of the RNA viral.

The descriptive analysis allows highlighting the statistically significant relationships between every single variable and the seroprevalence rate.

Logistic and multilevel models have been studied to detect which factors played the main roles in having a positive IgG outcome. The estimation of logistic regression coefficients reveals interesting differences between the different risk factors.

Alessandra BATTISTI, Istat, albattis@istat.it
Chiara COLUCCIA, Istat, coluccia@istat.it
Maria Clelia ROMANO, Istat, romano@istat.it
Simona STAFFIERI, Istat, staffier@istat.it
Marco Dionisio TERRIBILI, Istat, terribili@istat.it