rivista di statistica ufficiale

n. 1
2021

In this issue:

Istat
Istituto Nazionale di Statistica

rivista di statistica ufficiale

n. 1
2021

In this issue:

Istat

Editor:
Patrizia Cacioli

Scientific committee
President:
Gian Carlo Blangiardo

Members:
Corrado Bonifazi, Vittoria Buratta, Ray Chambers, Francesco Maria Chelli
Daniela Cocchi, Giovanni Corrao, Sandro Cruciani, Luca De Benedictis
Gustavo De Santis, Luigi Fabbris, Piero Demetrio Falorsi, Patrizia Farina
Jean-Paul Fitoussi, Maurizio Franzini, Saverio Gazzelloni, Giorgia Giovannetti
Maurizio Lenzerini, Vincenzo Lo Moro, Stefano Menghinello, Roberto Monducci
Gian Paolo Oneto, Roberta Pace, Alessandra Petrucci, Monica Pratesi
Michele Raitano, Maria Giovanna Ranalli, Aldo Rosano, Laura Terzera
Li-Chun Zhang

Editorial board
Coordinator:
Nadia Mignolli

Members:
Ciro Baldi, Patrizia Balzano, FedericoBenassi, Giancarlo Bruno
Tania Cappadozzi, Anna Maria Cecchini, Annalisa Cicerchia, Patrizia Collesi
Roberto Colotti, Stefano Costa, Valeria De Martino, Roberta De Santis
Alessandro Faramondi, Francesca Ferrante, Maria Teresa Fiocca, Romina Fraboni
Luisa Franconi, Antonella Guarneri, Anita Guelfi, Fabio Lipizzi
Filippo Moauro, Filippo Oropallo, Alessandro Pallara, Laura Peci
Federica Pintaldi, Maria Rosaria Prisco, Francesca Scambia, Mauro Scanu
Isabella Siciliani, Marina Signore, Francesca Tiero, Angelica Tudini
Francesca Vannucchi, Claudio Vicarelli, Anna Villa

rivista di statistica ufficiale
n. 1/2021

ISSN 1828-1982
© 2021
Istituto nazionale di statistica
Via Cesare Balbo, 16 – Roma

Unless otherwise stated, content on this website
is licensed under a Creative Commons License - Attribution - 3.0.
https://creativecommons.org/licenses/by/3.0/it/
Data and analysis from the Italian National Institute of Statistics can be copied, distributed, transmitted and freely adapted, even for commercial purposes, provided that the source is acknowledged.
No permission is necessary to hyperlink to pages on this website. Images, logos (including Istat logo), trademarks and other content owned by third parties belong to their respective owners and cannot be reproduced without their consent.
The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.
The Scientific Committee, the Editorial Board and the authors would like to thank the anonymous reviewers (at least two for each article, on a voluntary basis and free of charge, with a double-blind approach) for their comments and suggestions, which enhanced the quality of this issue of the Rivista di statistica ufficiale.

Editorial Preface

The four studies published in the first 2021 issue of the *Rivista di statistica ufficiale* focus on the COVID-19 pandemic, which has been inevitably affecting various areas of scientific research for more than a year.

The opening article by Gian Carlo Blangiardo, the President of the Italian Institute of Statistics – Istat, draws attention to the dynamics of the death rate in the first quarter of 2020, highlighting a shock that led to a much higher annual value than expected. In particular, due to the health emergency, in March 2020 the number of deaths increased dramatically, with an intensity that in some Italian areas reached very high peaks. Despite the positive signals at the beginning of the year, this severe rise in mortality was characterised by both a marked territorial localisation and a particular incidence of deaths among the elderly population, especially among men.

This article illustrates a historical analysis and comparisons with similar situations occurred in the past, making use of a simulation exercise to design different scenarios.

In the second scientific research work, Marco Claudio Traini, Carla Caponi, Riccardo Ferrari and Giuseppe Vittorio De Soci exploit the results of the seroprevalence survey carried out by Istat with the collaboration of the Ministry of Health and the support of the Italian Red Cross.

This survey aims at defining, within the entire population of Italy, the portion of individuals who developed an antibody response against SARS-CoV-2. For the first time an estimate of the asymptomatic infected population is available so as to acknowledge its potential role in the infection spread in Italy, one of the most affected areas in Europe.

The information obtained allow the authors to develop an epidemiological model able to simulate, in a systematic way, the asymptomatic group, whose relevance in the SARS-CoV-2 epidemic has been recently investigated and discussed. The study involves the description of the first epidemic outbreak in Italy as well as the predictive analysis of the second wave. In particular, the possible correction to the data of the serological tests because of their sensitivity and specificity.

The third article (by Chiara Orsi, Daniele De Rocchi, Simona Cinque, Roberta Crialesi, Vincenzo Della Mea, Luisa Frova, Enrico Grande, Stefano Marchetti, Simone Navarra, Marilena Pappagallo, Silvia Simeoni and Francesco Grippo) deals with the use of death certificates for the analysis of COVID-19 complications. Death certificates are indeed considered the most reliable information source for comparing cause-specific mortality across different populations and countries.

However, limited data are available on the complications and comorbidities reported on death certificates with regard to patients who presented COVID-19.

In order to identify these complications, a specific method is developed and applied to deaths occurred in Italy in the period March-April 2020. This method is also tested on other causes of death, specifically on 2018 data concerning diabetes and pneumonia, which represents the most frequent complication of COVID-19, together with some other conditions identified as particularly linked to COVID-19, *i.e.* cardiac complications, shock and infections. The method proves to be valid in distinguishing complications and causes of diseases.

Finally, in the article that closes this issue Roberto Gismondi, Maria Grazia Magliocchi, Filippo Oropallo and Francesco Giovanni Truglia present an economic analysis and an impact assessment of the COVID-19 effects on agritourism farms.

For meeting this goal, they make use of the integration of different data, exploiting statistical surveys and administrative sources. In addition, they carry on a microsimulation based on three hypothetical scenarios for 2020, including both the most recent economic sectoral trends and the survival probability of the agritourism farms.

The recourse to micro-founded analyses, combined with macro trends, is valuable to achieve the evaluation of the extent of the crisis in the agritourism sector, with the related consequences both at territorial level and for different types of businesses.

Patrizia Cacioli — *Editor*

Nadia Mignolli — *Coordinator of the Editorial board*

In memory of Professor Mario Di Bacco, R.I.P.

The *Rivista di statistica ufficiale* with deep sorrow shares the news of the recent death of Professor Mario di Bacco (13th June 1938 – 28th April 2021).

After graduating in Statistics at Sapienza University of Rome, Professor Mario Di Bacco carried out his academic career in the Italian Universities of Venice Ca' Foscari, Padova and Bologna, where he taught Statistical Inference and Calculus of Probability.

During his scientific activity, he has been interested in foundational issues and applications of Statistical Inference in Anthropology, Human Biology and related disciplines.

Very sensitive with regard to the training of young researchers and scholars, to be enhanced through the constant interdisciplinary and multidisciplinary comparisons, within the friendly dialogue that he has always been able to establish, since 1993 Mario Di Bacco has founded and continuously directed the *International Advanced School on Statistics applied to Biology and Human Sciences – BIOSTAT*.

BIOSTAT was initially itinerant and since 2001 has become stable at the Rita Levi-Montalcini University Pole of the municipality of Asti, during the course of which almost 30 editions hundreds of postgraduate students have been trained and specialised in Statistics.

Mario Di Bacco trusted in data and indicators, in the statistical method as a challenge to clearly and comprehensively describe all the aspects of interest under investigation, always with an ethically neutral attitude and being able to look beyond the numbers, up to the people, the other subjects, the events. Also for this reason, he was a great estimator of Istat, and of the *Rivista di statistica ufficiale* to which he gave a relevant support both in the selection phase of some scientific articles, and from the point of view of advice and suggestions for its further improvements.

His open-mindedness and continuous commitment to innovative fields of study, his initiative, keen interest, and his great vitality will always remain and be missed.

Demographic effects of COVID-19: reflections on mortality

Gian Carlo Blangiardo
President of the Italian National Institute of Statistics - Istat
President of the Scientific Committee of the Rivista di statistica ufficiale

Abstract

The dynamics of the death rate in the first quarter of 2020 moved in a very different direction, highlighting a shock that led to a much higher annual value than expected. In particular, due to the COVID-19 epidemic, in March 2020 the number of deaths increased dramatically, with an intensity that in some Italian areas reached very high peaks. This sharp rise in mortality, which occurred during a year that had started with excellent prospects, was characterised on the one hand by a marked territorial localisation, on the other hand by a particular incidence of deaths among the elderly population, and especially among men.
In the light of this sudden and rapid development, this article illustrates a historical analysis, which is essential for evaluating "if and when" something similar occurred in the past.
Finally, a simulation exercise is used to design different scenarios.

Keywords: COVID-19 pandemic; trend in the number of deaths; risk of death; simulation models.

1. The storm after the quiet

In the month of March 2020, due to the COronaVIrus Disease 2019 (COVID-19) pandemic, the number of deaths in Italy increased dramatically, with unprecedented peaks in some territorial areas of the country. Using a subset of 5,069 municipalities as a reference[1], it emerged that the total number of deaths recorded between March 1st and April 4th 2020 is higher by 41% than that observed in the same period of 2019 (such value has been recently updated to 49% for the month of March, on a larger set of 6,866 municipalities). The territorial detail identifies areas (48 municipalities) where the frequency of deaths increased at least tenfold as compared to the previous year; in many others (no less than 140), such frequency was at least five times higher. In 37 small municipalities, that in the period March 1st - April 4th 2019 had not registered any cases, the count of 2020 reached 304 deaths (Istat, 2020*a*).

Comparing gender and age, a 44% growth was recorded among persons aged 65 and over, against 11% among the remaining age groups, with a gap largely unfavourable to men: +56% among men aged 65 and over and +34% among women in the same age group.

It is noteworthy that the sharp increase in mortality in the month of March 2020 occurred in a year that started with excellent prospects: overall, in the 5,069 municipalities under examination the first two months of 2020 recorded a 8% decrease of deaths between January 1st and February 2nd and a 9% decrease of deaths between February 3rd and 29th, as compared to 2019.

1 This subset is included in the National Register of the Resident Population, which groups those municipalities characterised by reliability and timeliness of the data provided. This set is indeed self-selected and therefore is not to be intended as a statistically representative sample of the Italian municipalities, despite the fact that it covers about two-thirds of the total Register.

2. Earlier cases in recent history

Since the end of the Second World War (WWII), relevant increases in mortality have been observed in Italy in at least a couple of occasions (Figure 1). The first in 1956, with about 50,000 extra deaths concentrated during the winter[2] and among the older population. The second time in 2015, with a similar annual increase (+50,000 deaths) also in this case largely due to the action of flu viruses during the wintertime, associated to the lethal effects of a particularly hot summer.

In more detail, 188,132 deaths were recorded in the January-March quarter of 2015, as compared to 164,590 and 166,590 of the same period of 2014 and 2016.

Nevertheless, what we are witnessing today often evokes, at least when emphasising consequences of social and health-related kind, something that occurred long before the end of WWII, one century ago.

It was the global pandemic known as *Spanish Flu*: an event that manifested its most dramatic effects by the end of 1918, in the context – certainly very different from the present – of a population worn out by the Great War and affected by precarious conditions in their physical state, infrastructures and knowledge.

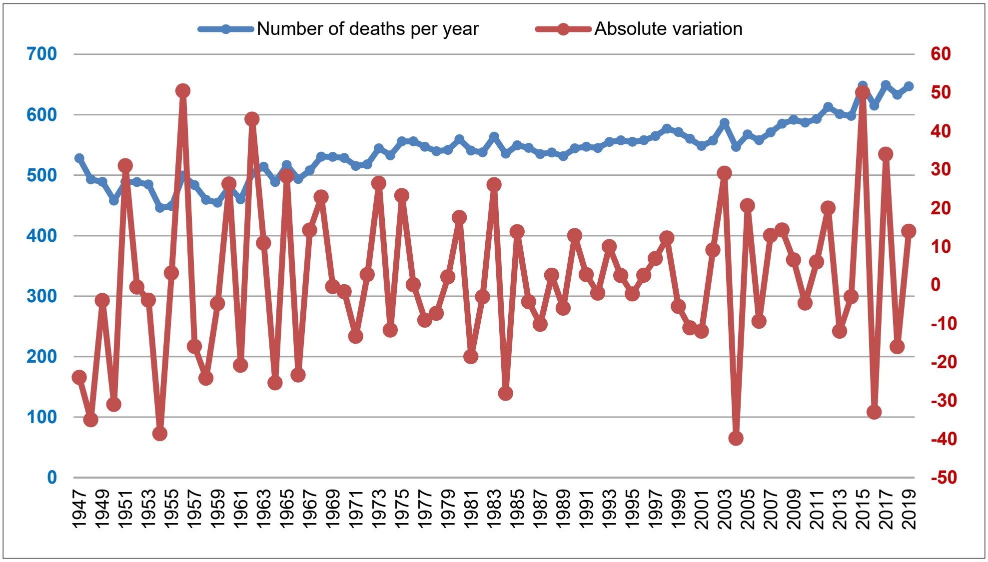
Upon reading the dramatic balance of what occurred in those times, we face figures greatly exceeding the 20-25,000 deaths – even if we should consider those as a default approximation – attributed so far to COVID-19 in 2020 (the current reference is the second half of April).

Quoting a famous Italian scholar of the time, Giorgio Mortara[3]: *"If we sum the extra deaths between August 1918 and March 1919, we obtain 532,457 deaths exceeding the normal figure. If we consider […] that for the invaded municipalities the number of deaths in the official statistics is lower than the real one, it is convenient to round the aforesaid figure up to 600,000 […]"* (Mortara, 1925).

2 From the comparison with the previous year, it is to be considered that the 1956 variation in the number of deaths was by over 80% concentrated in the first quarter of the year. In February 1956 a total of 69,739 deaths were recorded, with an increase of 30,730 and 29,919 cases respectively, as compared to the same month in 1955 and 1957 (Istituto Centrale di Statistica, 1957, 1958 and 1959).

3 Mortara,G., 1925: 121.

Figure 1 – Italy: variations of the number of deaths per year since the end of WWII (in thousands)



Source: Istat, Demographic balance

3. Key issues

As we have seen, figures from the remote past reflect the dramatic nature of an event that had huge consequences, today unconceivable and, we hope, totally non-repeatable.

However, in the light of the current experience, what could realistically be the effect of COVID-19 in terms of mortality? What consequences can we expect on the growth of life expectancy, a trend that we pleasantly came to take for granted? In addition, to what extent the evolution of demographic aging, so far defined by the scholars as “ineluctable”, could show signs of a significant relenting, or even a reversed trend?

In order to respond to such questions, it is necessary to produce some valid estimations of the final balance of the extra deaths caused by the pandemic, as well as of their incidence on different age groups. To this purpose, a simple exercise of simulation may allow us to design a few hypothetic scenarios.

Starting with the age structure of the resident population in Italy on January 1st 2020, and submitting it to the probability of death resulting from the most recent life table produced for Italy (Istat, 2020*b*), we obtain 637,000 deaths, a figure that could be compared – in simulating the global effects of COVID-19 – to what we would obtain by making the same calculation after introducing the appropriate variations of the risk of death, expressed by the probabilities listed in the life table.

For instance, supposing that the COVID-19 effect during a quarter could determine a constant 50% increase of the probability of death for the older age groups – here defined from the 60th birthday on – we would obtain for 2020 an annual number of 710,000 deaths (+73,000). Alongside with that, life expectancy at birth would decrease as far as 82.11 years (-0.87), and that at the 65th birthday from 20.89 to 20.02.

4. Alternative scenarios

If we proceed likewise and introduce, one at a time, alternatives (or variants), both of intensity in risk increase (probability of death) and in its duration, we will obtain the following range of scenarios (results).

Table 1 – Consequences of the increased probability of death of different intensity and duration

Year 2020	Simulation model - Variants							
	I	II	III	IV	V	VI	VI	VIII
Months:	*Percent variations of the basic risk for every age from 60 years*							
January	0	0	0	0	0	0	0	0
February	0	0	0	0	0	0	0	0
March	40	40	40	40	40	40	40	40
April	40	30	30	30	30	30	20	20
May	30	30	30	20	20	10	10	10
June	30	30	20	20	20	10	10	0
July	30	20	20	20	20	10	0	0
August	20	20	20	20	0	0	0	0
September	20	20	20	0	0	0	0	0
October	20	20	0	0	0	0	0	0
November	20	0	0	0	0	0	0	0
December	0	0	0	0	0	0	0	0
	Corresponding total variations for 2020 - Results							
Deaths (in thousands)	+123	+103	+88	+74	+64	+49	+39	+34
(*) e_0	-1.40	-1.19	1.04	-0.87	-0.77	-0.60	-0.48	-0.42
(**) e_{65}	-1.38	-1.18	1.03	-0.86	-0.76	-0.59	-0.47	-0.42
Population aged 65+ (in thousands)	+60	+79	+93	+107	+117	+131	+141	+145
Population aged 85+ (in thousands)	+2	+13	+20	+28	+33	+41	+46	+49

Source: Processing on Istat data
(*) Life expectancy at birth (years and fraction); (**) Life expectancy at age 65 (years and fraction).

It can be seen that increases of mortality due to COVID-19 pandemic are destined to grow, as it is natural to expect, depending on the intensity and duration of the greater risk of death.

Moving from a situation characterised by a persistent high level of increased risk, where the rise, albeit reduced, lasts until November (model I), to the scenario of a relatively rapid containment, with a return to normality within three months (model VIII), the annual frequency of deaths in 2020 would range from a maximum of 123,000 cases to a minimum of 34,000.

Alongside with this, life expectancy at birth would decrease by 1.4 years in the conditions of the most adverse model, and only by 0.42 in the least adverse one (Blangiardo, 2020).

In the two extreme cases, this would mean going back to life expectancy to be found in Italy, respectively, in the 2009-2010 and 2014 life tables calculated considering the total population.

5. Concluding remarks

Some further elements to complete this contribution.

Considering the effect on demographic aging, the models illustrate how the growth of the older population component – with regard to persons aged both 65 years and over and 85 and over – does not seem destined to stop in any case.

It rather results lower: +60,000 and +2,000, respectively, for the two aggregates, in the most adverse survival conditions; while it rises to +145,000 and +49,000 in the least dramatic hypothesis.

References

Blangiardo, G.C. 2020. *Scenari sugli effetti demografici di Covid-19 per l'anno 2020*. Roma, Italy: Istat. https://www.istat.it/it/files//2020/04/Report-Presidente_Finale.pdf.

Istituto Centrale di Statistica. 1960. *Annuario Statistico Italiano - Anno 1959*. Roma, Italy: Istituto Poligrafico dello Stato.

Istituto Centrale di Statistica. 1959. *Annuario Statistico Italiano - Anno 1958*. Roma, Italy: Istituto Poligrafico dello Stato.

Istituto Centrale di Statistica. 1958. *Annuario Statistico Italiano - Anno 1957*. Roma, Italy: Istituto Poligrafico dello Stato.

Istituto Nazionale di Statistica - Istat. 2020a. *#Istat per il Paese. Information on the health emergency and the recovery*. Roma, Italy: Istat. https://www.istat.it/en/archivio/240106; https://www.istat.it/it/archivio/239854.

Istituto Nazionale di Statistica - Istat. 2020b. "Life Tables of the Population by province and region of residence. Years 1974-2020". In *Demography in Figures*. Roma, Italy: Istat. http://demo.istat.it/index_e.php.

Mortara,G. 1925. *La salute pubblica in Italia durante e dopo la guerra*. Bari, Italy: Laterza.

Modelling the SARS-CoV-2 epidemic in Italy making use of the Istat seroprevalence survey

Marco Claudio Traini [1], Carla Caponi [2], Riccardo Ferrari [3], Giuseppe Vittorio De Socio [4]

Abstract

On 3rd August 2020, the Italian National Institute of Statistics – Istat published the preliminary results of the seroprevalence survey on the percentage of individuals affected by COVID-19.
This survey aims at defining (within the entire population of Italy) the portion of individuals who developed an antibody response against SARS-CoV-2. For the first time an estimate of the asymptomatic infected population is available so as to acknowledge its potential role in the infection spread in Italy, one of the most affected countries in Europe.
The information obtained allows a particularly sensitive validation of epidemiological models which include the asymptomatic class.
The present study is devoted to the construction of a model able to simulate, in a systematic way, the asymptomatic group whose relevance in the SARS-CoV-2 epidemic has been recently discussed. The investigation involves the description of the first epidemic outbreak in Italy as well as the predictive analysis of the ongoing second wave. In particular, the possible correction to the data of the serological tests because of their sensitivity and specificity.

Keywords: SARS-CoV-2, asymptomatic group, seroprevalence, infection spread, epidemiological models, sensitivity and specificity of the serological tests, vaccination strategies.

1 Full Professor of theoretical physics; Università degli Studi di Trento, Dipartimento di Fisica, Trento, Italy. Email: marcoclaudio.traini@unitn.it.

2 Physician, MD; Clinica Geriatrica, Azienda Ospedaliero-Universitaria, Perugia, Italy. Email: carla.caponi@ospedale.perugia.it.

3 Manager and Physicist; Bilubah LLC, Sheridan, WY 82801, U.S.A. Email: riccardo.ferrari@bilubah.com.

4 Physician, MD; Clinica Malattie Infettive, Azienda Ospedaliero-Universitaria, Perugia, Italy. Email: giuseppe.desocio@ospedale.perugia.it.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction[5]

Traditionally, epidemiological models have grouped people into two, three or four groups (compartments), usually denoted by Susceptible (*S*), Exposed (*E*), Infected (*I*), and Recovered (*R*). Contact between a member of the infected group (*I*) and another person belonging to the susceptible group (*S*) leads to the latter person becoming infected with a certain probability. Depending on the model, the susceptible person either becomes infected straightaway (SIR model), or enters an intermediate stage called Exposed (*E*) (SEIR model). In the latter scenario, it is assumed that contact between persons belonging to the (*E*) and (*S*) groups does not lead to fresh infections, because members of the (*E*) group do not carry a sufficient viral load to infect others through contact.

However, one of the characteristic features of the coronavirus pandemic is that many of the persons who contract the disease are "asymptomatic" (or group A). In fact the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) throughout the world has been extremely rapid suggesting the hypotheses of a crucial role played by these infected persons who remain asymptomatic even if contagious (Buitrago-Garcia, 2020). A recent paper (Oran, 2020) summarises the available evidences on asymptomatic SARS-CoV-2 infection concluding that "*asymptomatic persons seem to account for approximately 40% to 45% of the SARS-CoV-2 infections, and that they can transmit the virus to others for an extended period of time, perhaps longer than 14 days*". Oran and Topol (Oran, 2020) conclude with the need "t*hat testing programs include those without symptoms*". Other studies estimate the fraction of asymptomatic patients to be more than 50% (Mizumoto, 2020) or as high as 75% (Day, 2020). For this reason, asymptomatic patients remain "hidden" and cannot be identified except through tests of large portions of the population. An example is given in the study by Lavezzo, Franchin, Ciavarella, *et al.* (Lavezzo, 2020) reporting results of a detailed survey in the small town of Vo' (near Padua, Italy) where on 21st February 2020 a lockdown was imposed in the whole municipality as a first outbreak in Italy. After the lockdown they found a prevalence of 1.2% (95% CI:0.8-1.8%). Notably, 42.5% (95% CI:31.5-54.6%) of the confirmed SARS-CoV-2 infections detected across two surveys were asymptomatic (that is did not have symptoms

5 The authors thank Francesca Greselin of the Department of Statistics and Quantitative Methods - University of Milano Bicocca, Italy - for her interest and help in the elaboration of the model used.

at the time of swab testing and did not develop symptoms afterwards). After that a number of studies and projects have been developed to identify the role of asymptomatic in the pandemic infection in various countries (*e.g.* Buitrago-Garcia, 2020; Guerriero, 2020; Peterson, 2020; Pollán, 2020; Snoeck, 2020; Well, 2020; Yanes-Lane, 2020).

Table 1 - Preliminary results of the seroprevalence survey in Italy (a)

Region	IgG positive (95% CI)		Absolute value	Reported cases
	lower limit	upper limit		
Italy	2.3	2.6	1,482,377	244,708

Source: Istat, 2020
(a) The amount of IgG positive tests corresponds to an infected population of (almost) 1.5 million of individuals, when the official reported cases are 244,708 till 27th July 2020.

In August 2020 preliminary results of a seroprevalence survey have been presented for a set of 64,660 persons in Italy (Istat, 2020). The Istat survey aimed at defining (within the entire population) the portion of persons that developed an antibody response against SARS-CoV-2. The survey adopted a methodology allowing the evaluation of the seroprevalence in the population also estimating the fraction of asymptomatic (or subclinical) infections and the differences for age groups, gender, localisation etc. The results are still preliminary because involve a restricted number of tests, in particular those ones whose results have been reported before 27th July 2020. The post-stratified techniques adopted[6] (Little, 1993) allow the production of statistical estimates coherent with epidemic data (both at the international and local level). For the first time there is an estimate of the asymptomatic infected population and its influence in the infection spread in Italy, one of the most affected countries in Europe.

The results of the Istat survey, enlarged to the entire Italian country, are summarised in Table 1. The analysis concludes that 1,482,377 individuals developed IgG antibodies against SARS-CoV-2. A level of seroprevalence of 2.5% (95% CI: 2.3-2.6%) to be compared with 244,708 officially reported cases.

6 Due to intentional and/or unintentional processes the characteristics of a sample may not represent the characteristics of the population of interest. To mitigate this potential bias survey researchers post-stratify base probability of selection survey weights so that sample characteristics match population control totals.

Asymptomatic patients (*A*) differ from exposed patients (*E*) in one important respect. Unlike in traditional epidemiological models, contact between a person in the (*A*) group and another in the (*S*) group does lead to the latter getting infected, with a certain probability. In addition, as in other models, contact between a person in the (*I*) group and another in the (*S*) group also leads to the latter getting infected, with a similar probability. The seminal paper by Robinson and Stilianakis (Robinson, 2013) formulates and analyses a preliminary model that captures the asymptomatic phenomenon (*A*) and can be called SAIR, since the corresponding model including exposed (*E*) group is usually called SEIR. The two models differ in many aspects from a mathematical point of view; see Ansumali, *et al.* for a recent analysis and detailed discussion (Ansumali, 2020).

Moreover it is possible to develop more refined models of the pandemic by introducing additional categories such as Quarantined, Healed, Ailing, Recognised (or Detected), Threatened, etc. (*e.g.* Park, 2020; Giordano, 2020). By introducing more categories, the result is a more realistic model of the disease progression. On the other hand, the number of parameters to be estimated increases drastically. The ideal trade-off between these two conflicting considerations remains and has to be considered on a case by case basis (Jia, 2020*a*; Kinoshita, 2020; Yu, 2020).

The aim of the present paper is the study of asymptomatic compartment population during the SARS-CoV-2 epidemic event in Italy taking advantage of the Istat survey. The model used assumes *eight* compartments (groups) including symptomatic detected and undetected, quarantined, asymptomatic, threatened (hospitalised) and recovered, a model recently proposed to discuss the role of measures against the Italian outbreak (Traini, 2020).

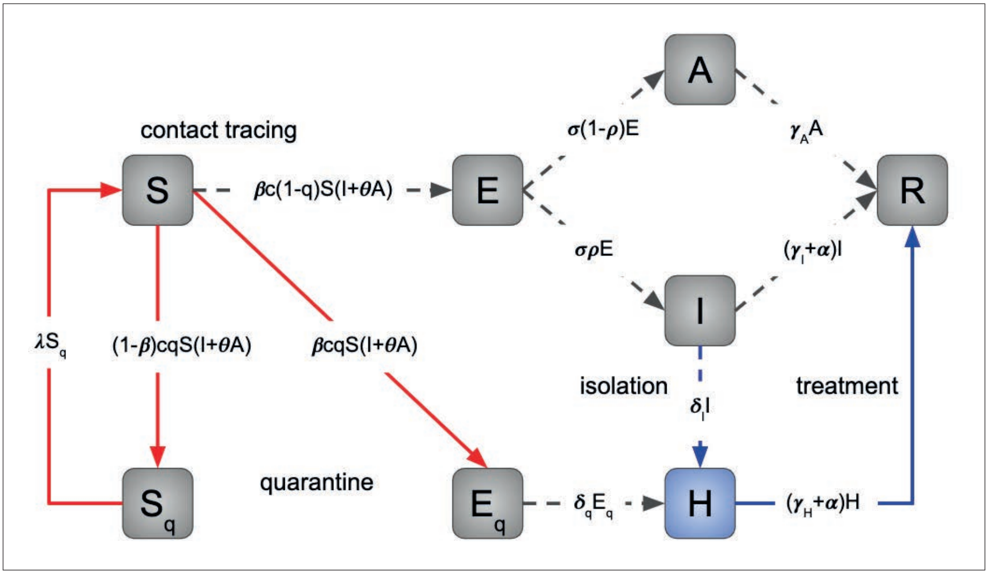
2. Methods and analysis

2.1 A time dependent quarantined model with isolation

The model we use in order to describe the time evolution of the Italian outbreak is an epidemiological model originally proposed by Tang *et al.* in order to study the Wuhan event (Tang, 2020. *J. Clin Med.*; Tang, 2020. *Infect. Dis. Model*). This model incorporates appropriate compartments relevant to intervention such as quarantine, isolation, and treatment. The population is stratified in Susceptible (S), exposed (E), asymptomatic infected individuals (A), infected with symptoms (I), hospitalised in a large sense (detected infected) (H), and recovered (R). Further stratification includes quarantined susceptible (S_q), and isolated exposed (E_q) compartments (see Figure 1) which describe the tracing procedures.

A portion of susceptible, S, get in contact with infected individuals with rate $c\ (I + \theta A)$, where c is the contact rate, I the number of symptomatic infected individuals, A the asymptomatic infected individuals, and θ the contribution of asymptomatic infected to the infection spread. With contact tracing, a proportion, q, of susceptible, S, that get in contact with infected individuals is quarantined. Individuals receive the virus at rate β, which is the transmission probability, and become exposed. On the other side, the exposed individuals identified with contact tracing get quarantined at rate q. Therefore, we have three fluxes of individuals out of S: the quarantined with virus transmission going into E_q, $\beta cqS(I + \theta A)$; the quarantined without virus transmission going into S_q, $(1 - \beta)cqS(I + \theta A)$; the individuals with virus transmission but not identified and not quarantined going into E, $\beta c(1 - q)S(I + \theta A)$. Quarantined individuals without virus transmission are released at rate λ, generating an inbound flux of individuals to S given by λS_q. Exposed, infected, and quarantined individuals move to the hospitalised compartment at rate $\delta_q E_q$. Exposed, infected, and not quarantined individuals become infectious at rate σ. Some of them develop symptoms with a probability of ρ. Then, there are two outbound fluxes of individuals for the compartment E: the infected with symptoms, $\sigma\rho E$; the infected asymptomatic $\sigma(1 - \rho)E$. The infected with symptoms will eventually be detected and hospitalised with a rate of δ_I, which reflects the sanitary system's diagnostic capability. Finally, all the infected

Figure 1 - Diagram of the model simulating the novel Coronavirus (Sars-CoV-2) infection in Italy (a)



Source: Tang, 2020
(a) The population is stratified in Susceptible (S), exposed (E), asymptomatic infected individuals (A), infected with symptoms (I), detected infected (hospitalised in a large sense) (H) and recovered (R), quarantined susceptible (Sq), isolated exposed (Eq) compartments. Interventions like intensive contact tracing followed by quarantine and isolation are indicated.

will recover with rate: γ_A, for the asymptomatic, γ_I for the infected not hospitalised, γ_H for the infected hospitalised. The infected with symptoms, both hospitalised or not, have a mortality rate of α. For disease transmission, these individuals pass to the recovered compartment, as they are no more infectious. Summing the inbound and outbound fluxes at each compartment, we obtain the system of differential equations of the model:

$$\frac{dS}{dt} = -\left[\beta c(t) + c(t) q\,(1-\beta)\right]\, S\,(I + \theta\, A) + \lambda\, S_q; \qquad (1)$$

$$\frac{dE}{dt} = +\beta c(t)(1-q)\, S\,(I + \theta\, A) - \sigma\, E; \qquad (2)$$

$$\frac{dI}{dt} = +\sigma\, \varrho\, E - (\delta_I(t) + \alpha + \gamma_I)\, I \qquad (3)$$

$$\frac{dA}{dt} = \sigma\,(1-\varrho)\, E - \gamma_A\, A; \qquad (4)$$

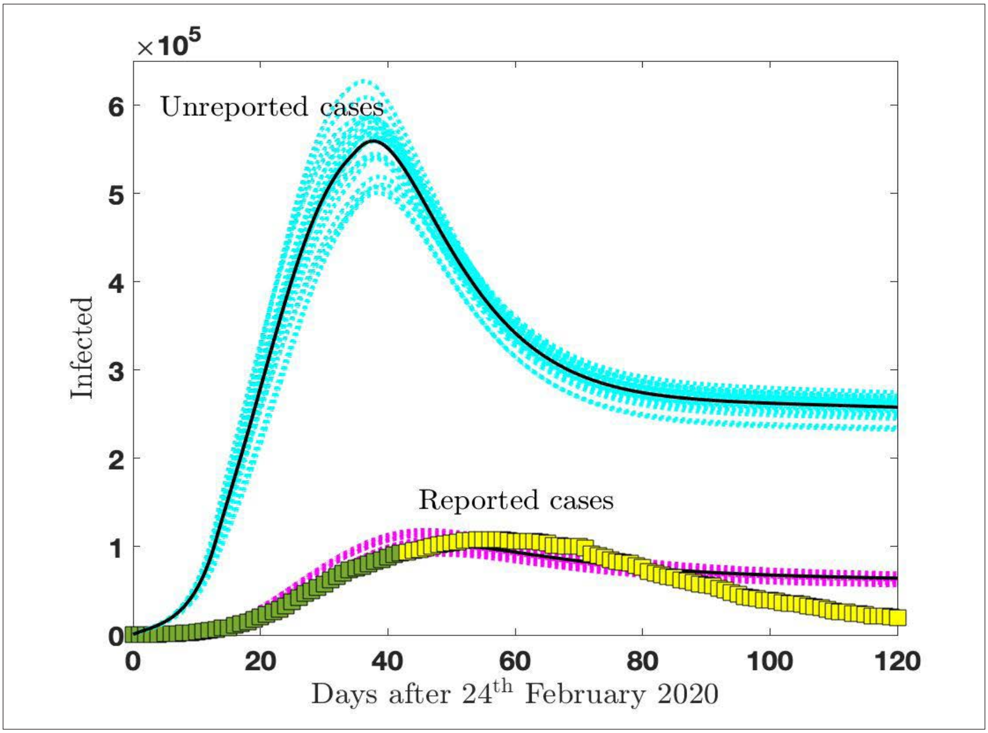
$$\frac{dS_q}{dt} = +c(t)q\,(1-\beta)\,S\,(I+\theta\,A) - \lambda\,S_q; \quad (5)$$
$$\frac{dE_q}{dt} = +\beta c(t)q\,S\,(I+\theta\,A) - \delta_q\,E_q; \quad (6)$$
$$\frac{dH}{dt} = \delta_I(t)\,I + \delta_q E_q - (\alpha+\gamma_H)\,H; \quad (7)$$
$$\frac{dR}{dt} = \gamma_I\,I + \gamma_A\,A + \gamma_H\,H; \quad (8)$$

and the values of the parameters are discussed in the next Section.

2.2 Fixing the model parameters: the outbreak

The method Markov Chain Monte Carlo (MCMC, Brooks, 2011; Hogg, 2018) is used to fit the model on the data of the outbreak in Italy in the period up to 6th April 2020. The procedure is implemented through an adaptive Metropolis-Hasting (M-H) algorithm used for four concatenated runs with 100,000 - 50,000 - 25,000 - 10,000 iterations within the MCMC toolbox for Matlab. Table 2 summarises the parameters. A peculiar aspect of the model is the time-dependence of the key parameters related to the contact rate c and the quarantined rate q. Following the control measures adopted in Italy, the flexibility of the model is such that it is possible to adapt the values of some parameters to the concrete social situation. In particular the values c_0, c_1, c_2 and q_0, q_1, q_2 are fixed on the period of the outbreak 24th February - 6th April 2020.

Figure 2 - A MCMC analysis of the infected individuals (reported and unreported cases) in Italy as a function of time in the period 24th February - 27th July 2020 (a)



Source: The data for the reported-infected individuals are by the Ministry of Health (*Ministero della Salute*, 2020), and are compared with our model predictions (magenta curves) which include statistical uncertainties

(a) The green data points refer to the data set used to fix the model parameters, while the yellow set spans the purely predicted period. The reported cases (magenta curves) are then compared with the predictions for the class of infected (asymptomatic and symptomatic) non reported cases (cyan curves) in the same period of time.

Table 2 - Parameters of the time-dependent model description

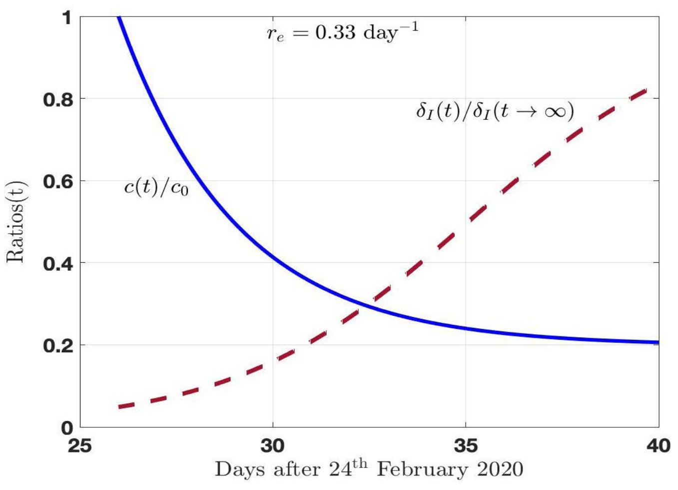
Param.	Definition [dimensions]	Estimated Value	St. Dev.	notes
$c = c_0$	Contact rate $[\text{day}^{-1}]$	4.5248	0.2030	from 24th Feb. to 8th March
$c = c_1$		$c_0/2$	0.2030	from 8th March to 29th March
$c = c_2$		$1.4 \cdot c_1$	0.2030	from 29th March to 6th April
$c = c_3$		c_0	0.2030	from 6th April to 25th Dec.
$c = c_4$		$20 \cdot c_0$	0.2030	after 25th December
$c = c(t)$	see Section 2.3			to discuss the lockdown exit of 4th May
β	Probability of transmission per contact [u]	$1.5851 \cdot 10^{-8}$	$3.360 \cdot 10^{-10}$	
$q = q_0$	Quarantined rate of exposed individuals [u]	$1.155 \cdot 10^{-7}$	$7.720 \cdot 10^{-9}$	from 24th Feb. to 8th March
$q = q_1$		$5.5 \cdot q_0$	$7.720 \cdot 10^{-9}$	from 8th March to 29th March
$q = q_2$		$2 \cdot q_1$	$7.720 \cdot 10^{-9}$	from 29th March to 6th April
$q = q_3$		$2 \cdot q_1$	$7.720 \cdot 10^{-9}$	from 6th April to 25th Sept.
$q = q_4$		$2 \cdot q_1$	$7.720 \cdot 10^{-9}$	from 25th Sept.
σ	Transition rate of exposed individuals to the infected class $[\text{day}^{-1}]$	1/7		
λ	Rate at which the quarantined uninfected contacts were released into the wide community $[\text{day}^{-1}]$	1/14		
ϱ	Probability of having symptoms among infected individuals [u]	0.2344	0.0060	
$\delta_I = \delta_{I_0}$	Transition rate of symptomatic infected individuals to the quarantined infected class $[\text{day}^{-1}]$	0.1086	0.0010	
$\delta_I = \delta_I(t)$	see Section 2.3			to discuss the lockdown exit of 4th May 2020
δ_q	Transition rate of quarantined exposed individuals to the quarantined infected class $[\text{day}^{-1}]$	0.1471	0.0010	
γ_I	Recovery rate of symptomatic infected individuals $[\text{day}^{-1}]$	0.1704	0.0050	
γ_A	Recovery rate of asymptomatic infected individuals $[\text{day}^{-1}]$	0.1177	0.0020	
γ_H	Recovery rate of quarantined infected individuals $[\text{day}^{-1}]$	0.0515	0.0020	
α	Disease-induced death rate $[\text{day}^{-1}]$	$1.2197 \cdot 10^{-5}$	$3.0 \cdot 10^{-7}$	
θ	Relative weight of asymptomatic infections [u]	0.0840	0.0010	
Initial values (24th February 2020)				
N		$6 \cdot 10^7$		
$E(0)$		21308	899	
$I(0)$		201	1.5	
$A(0)$		493	5.6	
$S_q(0)$		50000		
$E_q(0)$		522	1.3	
$H(0)$		221		
$R(0)$		8		

In Figure 2 some relevant predictions drawn from the model (together with some evident failures). The model is parametrised by means of the data on infected (reported cases) in the first time period (green data points in Figure 2) till 6th April 2020 (day 42). It remains consistent with data till end of May (day 90), then it deviates from data. In fact the model was designed to study possible secondary effects after 4th May 2020, when the lockdown in Italy has been gradually mitigated (see next Section 2.3). Its longterm behaviour is discussed in detail (Traini, 2020). Here we want to emphasise an important result that can be derived from the model predictions without formulating a new parametrisation, namely the role of asymptomatic compartment as it emerges from the recent analysis by the Italian National Institute of Statistics – Istat (*Istituto Nazionale di Statistica*). As a matter of fact the model predicts also the global features of the distribution for the asymptomatic compartment illustrated in Figure 2 and discussed in Sections 3.1 and 4.2.

2.3 Predictive power: scenarios for secondary events

On 8th March 2020, the Italian Government announced the implementation of restrictions for controlling the infection. Strong control measures (like convincing all the residents to stay home and avoid contacts as much as possible) have been adopted (lockdown).

Figure 3 - Modelling the variation in time of the contact rate per person $c(t)$ $(c(t=0)=c_0)$ **and of the rate from infected to quarantined classes** $\delta_I(t)$ **(a)**



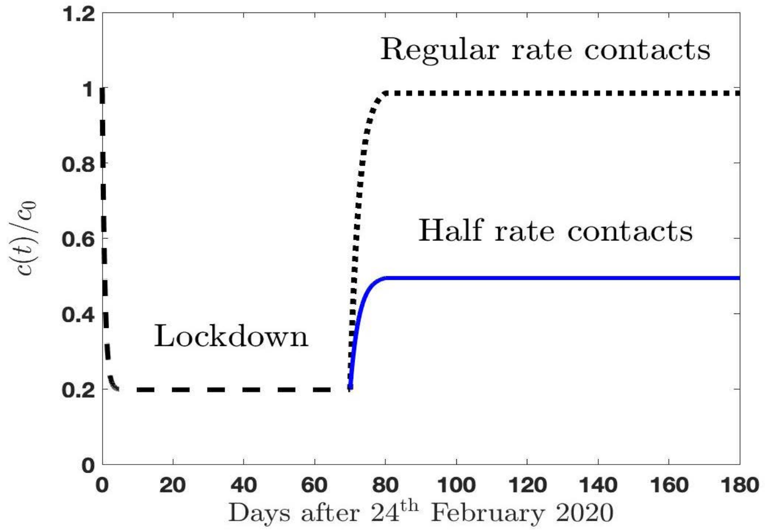
Source: Our processing
(a) See Equations (9) and (10).

From the model perspective, this can significantly contribute to decreasing the contact rate (c) among the persons (Remuzzi, 2020). Besides, the 2019-nCoV tests gradually shortened the time of diagnosis (*i.e.* the value of δ_I

increased gradually). Considering these control strategies, we could tune the model on the concrete Italian conditions.

The equations of the model, shown in Equations (1)-(8), contain parameters explicitly dependent on time. In particular the contact parameter c and the transition rate of declared infected individuals δ_I.

Figure 4 - Time behaviour of the contact rate simulating a possible secondary event in Italy at day=70, after 24th February, when the stringent measures of isolation are hypothetically relaxed (a)



Source: Our processing
(a) Two scenarios are introduced, the scenario of a rapid return to the old style of life (regular rate contacts), and a more realistic scenario where only half of the regular contacts are activated.

2.3.1 Time dependence

The initial Montecarlo analysis to fix the model parameters can be complemented with an accurate analysis of the effects of social measures or technological advances. An example is presented in Figure 3 where the time behaviour of the contact parameters and tracing rate is tuned following specific modelling related to concrete events changing the strength of the parameters involved. The time-dependence is parametrised as

$$c(t > t_0) = (c(t = t_0) - c_a)\, e^{-r_e\,(t-t_0)} + c_a, \qquad (9)$$

where $c_d/c_0 = 0.2$, $t = 0$ is fixed at 24th February and t_0 selects the initial day of a rapid implementation of the isolation measures for the entire population. At the same time the parameter δ_I, tuning the transition rate to quarantine of the infected individuals $(t_I = 1/\delta_I)$, increases because of a decreasing of the testing time:

$$\begin{aligned} & t_I(t) = [\delta_I(t > t_0)]^{-1} = \\ = \; & \left\{[\delta_I(t = t_0)]^{-1} - [\delta_I(t \gg t_0)]^{-1}\right\} e^{-r_e\,(t-t_0)} + \\ + \; & [\delta_I(t \gg t_0)]^{-1}. \end{aligned} \quad (10)$$

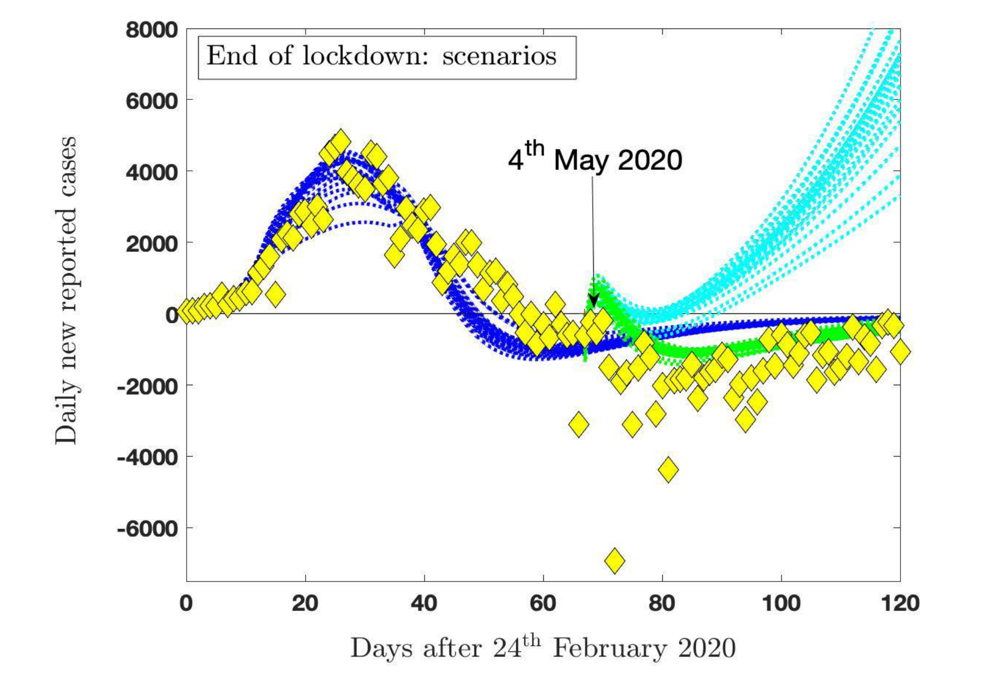
In Figure 3 the behaviour of the two parameters as a function of time. The rate of change $r_e = 0.33$ day^{-1} assumes the same values for the two gradually changing quantities. The parametrisation (10), in particular, is chosen to simulate the behaviour of the screening system to test infected (and asymptomatic) individuals.

2.3.2 Secondary events

Italy has been the first western country involved in the epidemic outbreak SARS-CoV-2. After 39 days (30th March 2020) the evolution was still in a first expanding phase. However, after almost 20 days of lockdown, the perspective of relaxing, at least partially, the social distancing measures appeared in many discussions at different levels and one can verify the predictive power of the model to simulate different scenarios of possible secondary effects. Let us assume that at a given day (4th May 2020 in the present study), the containment measures are relaxed totally or partially. The time evolution of the contact rate could be described as in Figure 4. The isolation value, as discussed in the previous Section, assumes the limiting value $c \approx 0.2$ (isolation), and at day = 70th from 24th February 2020, the isolation is interrupted and the "normality" activated. Two scenarios are introduced: i) a full return to the previous style of life (rather unlike, but it represents a reference point); ii) a more realistic scenario with half isolation.

Assuming the parametrisation of Table 2, one can try to reproduce the data on the reported cases. The good comparison between data and model can be valid for a short time, but we are not interested in reproducing the exact numbers, but the relative effects of a possible secondary event since the first

Figure 5 - The model predictions for the daily new reported cases after 24th February 2020 in Italy, compared with the official data (a)



Source: Our processing using the Italian Ministry of Health data, 2020 (*Ministero della Salute*, 2020)
(a) In addition to the fully locked scenario (blue lines), two other scenarios are included for a mitigation of the lockdown period in Italy starting from May 4th. Scenario A (cyan lines): a social contact rate somewhat similar to the regular rate (only a reduction of a factor of two is included) would result in a violent secondary outbreak. Scenario B (green lines): A stronger reduction of the contact rate which excludes a large part of social events (school, sports, large events, etc.) with the further inclusion of tracing and quarantine.

outbreak did not exhaust its virulence. In addition, guided by the analysis of the effects of the technological developments in the screening phase done in the previous sections, one can try to see their effects in the secondary events and the results of such an investigation are shown in Figure 5 where the comparison between the model predictions and the data is stringent. The data of the daily *new* (reported) cases, from 24th February on, show a large dispersion due to the influence of strategic decisions after the lockdown period (*i.e.* after 4th May 2020). The behaviour of the data follows rather consistently the model predictions of the reported cases of Equation (7) once scenarios for the mitigation of the lockdown are introduced. In particular, the data seem to follow rather consistently a scenario (green lines) where the contact rate excludes a large part of social events (school, sports, large events) and it includes the introduction of tracing and quarantine. The fact that the longterm data (after 4^{th} May 2020) are well reproduced by the model means that distancing and tracing measures proposed in Italy had a positive impact captured by the model. More dramatic scenarios (as described, for instance, by the cyan curves) have been avoided. Particularly relevant the value of $t_I = 1/\delta_I$ reached during the transition time to normality. Lowering the diagnostic time to $t_I \approx 7$ hours, or $t_I \approx 5$ hours has the power to substantially mitigate the secondary events, in particular, if one keeps "half rate contacts" (as schematically designed in Figure 4).

3. The Asymptomatic prevalence: results and discussion

Epidemiological surveillance of COVID-19 cases captures only a portion of all infections because the clinical manifestation of infections with SARS-CoV-2 ranges from severe diseases, which can lead to death, to asymptomatic infection. The official sources of data in Italy (Ministry of Health - *Ministero della Salute*, 2020) do not provide information about the asymptomatic patients thus limiting their usefulness in view of calculating interesting epidemic parameters, not last the lethality rate. Attempts to exploit the existing available data in order to estimate the prevalence and the lethality of the virus in the total Italian population has been proposed by Bassi, Arbia and Falorsi (Bassi, 2020). They used a post-stratification (Little, 1993) of the official data in order to derive the weights necessary for re-weighting the sample results. The re-weighting procedure artificially modify the sample composition so as to obtain a distribution which is more similar to the population. They obtain a prevalence of 9%, a rather large number.

Conversely, a (population-based) seroepidemiological survey can quantify the portion of population which developed antibodies agains SARS-CoV-2 providing information on the exposed individuals and on the remaining susceptible subjects (assuming that antibodies are marker of total or partial immunity). A certain number of surveys of SARS-CoV-2 have been realised or planned (in addition to the already quoted references: Sood, 2020; Valenti, 2020; Stringhini, 2020; Bryan, 2020; Shakiba, 2020; Doi, 2020; Eriks 2020; Wu, 2020; N. Bobrovitz, 2020).

3.1 Model results and Istat survey

The data, elaborated by Istat in their report, have been collected in a period of time (25th May -15th July 2020) running over (almost) two months. They are summarised in the upper and lower limits and the absolute value of Table 1 or (approximately) in the following ratio

$$\text{ratio}|_{\text{survey}} = \frac{\text{absolute value}}{\text{reported cases}} \approx 6. \tag{11}$$

The ratio (11) is a clear sign of the large impact of the asymptomatic prevalence on the total infected population with positive IgG in Italy even if no time dependence is assumed for those data, an assumption that needs some more attention. In fact, the model description we are proposing can be of some help to answer questions on the time dependence of the data. One can presume a roughly constant behaviour of the ratio (11) as a function of time under the following assumptions:

i) The antibodies have a lifetime significantly longer than the time needed for collecting the data.

ii) The social contact rate regime is rather stable during the collection of data. In this way the fraction of infected population is (on average) connected with the transmission probability of the infection which is assumed to be constant. In the present model $\beta = (1.5851 \pm 0.0336) \cdot 10^{-8}$ day^{-1}, (see Table 2).

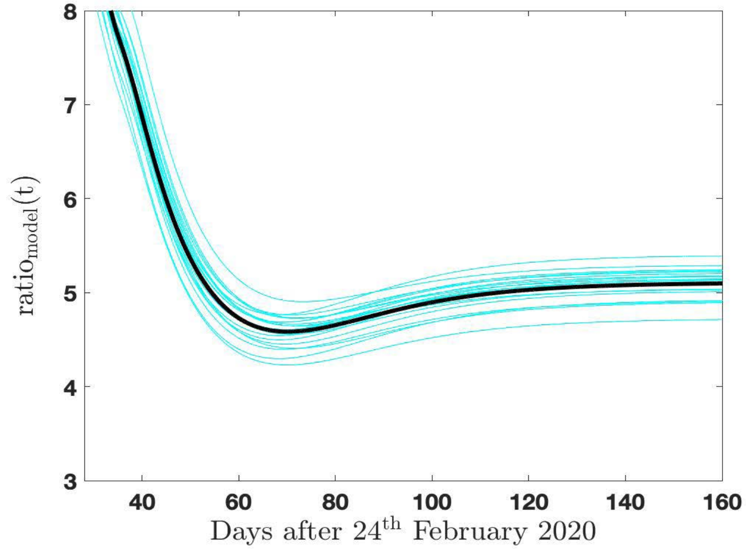
iii) In a (large and representative) sample, the portion of individuals reached by the infection in the previous months is fixed (next discussion will be largely devoted to clarify this point).

The assumption i) sounds reasonable and the assumption ii) can be explicitly checked in our model by defining the following (time dependent) ratio

$$\text{ratio}|_{\text{model}}(t) = \frac{A(t) + I(t) + H(t)}{H(t)}. \tag{12}$$

The numerator, in Equation (12), counts, for each day, the total number of infected (reported and unreported with symptoms plus unreported asymptomatic infected individuals) and the denominator counts the number of reported cases. Both numerator and denominator are largely time dependent functions as already discussed in the previous sections. The model ratio (12) is shown in Figure 6. Despite the rapid decrease of the ratio before 25th May 2020, its value remains, after that date and till the end of July, approximately constant ($\text{ratio}_{\text{model}}(t) \approx 5$). In fact, the contact rate between 24th February -25th May 2020 had large variation due to lockdown and social restriction measures adopted in Italy, while in the period 25th May -31st July the social situation was rather stable. The model captures such variations validating the assumption ii) for that period.

Figure 6 - The model ratio of Equation (12) as a function of time



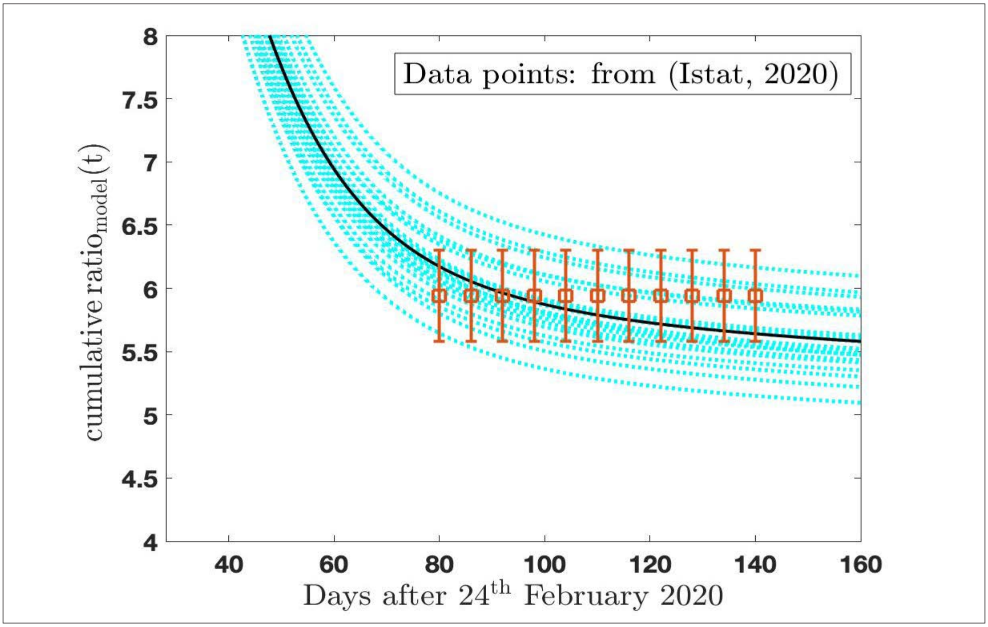
Source: Our processing

However the ratio (12) cannot be compared directly with the results of the Istat survey. As a matter of fact the survey measured the number of individuals who developed antibodies without selecting the period of infection. The serological test cannot answer to time dependent questions only measuring the presence of antibodies which simply count the total amount of infections in the period preceding the test. In order to reproduce, as close as possible, the analysis from Istat, we can calculate the cumulative values of the numerator and denominator of Equation (12) summing (up to a given day (t)) the components. The numerator of Equation (11) (*i.e.* the absolute value) is given by the sum of the Asymptomatic population (A), the (unreported) Infected (I) and the official reported cases (H); the denominator the cumulative value of the official reported cases. One gets:

$$\text{cumulative ratio}\big|_{\text{model}}(t) = \frac{\sum_{t'=0}^{t}\left[A(t') + I(t') + H(t')\right]}{\sum_{t'=0}^{t} H(t')}. \qquad (13)$$

The cyan curves in Figure 7 summarise the model predictions for the ratio (13). The black curve representing the mean value and the dotted lines the normal distributed variations.

Figure 7 - The fixed ratio of Equation (14), elaborated from the results of the seroprevalence survey (data-points) is compared with the model ratio of Equation (13) (cyan curves)



Source: Our processing

The results of the Istat survey can be transformed in an analogous ratio in a simple way: the absolute value corresponds to an average value of the prevalence (2.3 + 2.6)/2 ≈ 2.5 (see Table 1). Therefore the upper and lower limit of the infected population divided by the reported cases is summarised by the ratio

$$\text{ratio}|_{\text{survey}} = 5.94 \pm 0.36, \qquad (14)$$

which now includes the uncertainties estimated in the Istat survey and updates the approximation (11). The data points in Figure 7 give a graphical representation of the ratio (14) assumed, in a first step, to be time-independent on the basis of the assumption i) and ii) already discussed.

From this first analysis one can conclude that:

1. the global comparison in Figure 7 between the estimated values and the evidences of reference (Istat, 2020) is surprisingly successful, in particular when statistical uncertainties are included in the analysis

of the survey and in the model predictions. For a long period of time during the lockdown months, speculations on the Asymptotic population gave strongly divergent estimates. Our results of Figure 7 are obtained within a model fixed in 6th April 2020 and published before the survey (Traini, 2020); data confirm that the epidemiological models can offer predictions rather stable and realistic also for the Asymptomatic compartment.

2. The model results for the cumulative ratio of Figure 7 start with a value ≈ 8 at the beginning of April 2020, reaching an asymptotic *average* value ≈ 5.5 at the end of July 2020. For the whole period of the preliminary survey (25th May -15th July) the ratio of Equation (14), (data-points in Figure 7), is reproduced by the model calculation of Equation (13) within the statistical uncertainties.
3. The conclusions of the Istat survey (*i.e.* "*the seroprevalence data at regional level, to be integrated with the epidemiological surveillance data, are particularly relevant to identify, on one side, the portion of individuals reached by the infection in the previous months, and for programming measures to prevent future possible second waves, on the other side*"[7]) can easily be applied to the relevance of modelling the behaviour of the asymptomatic population, a key ingredient to manage the future of the pandemic event (*e.g.* Gandhi, 2020).

3.2 Sensitivity and specificity of the serological tests

A first correction to the seroprevalence analysis presented in the previous Section is due to the sensitivity and specificity of the serological tests adopted in the survey screenings (for a recent note on the false positive and false negative in diagnosis of COVID-19, see Jia, 2020*b*).

The report by Istat (Istat, 2020: page 9) indicates in *not less than 95%* the specificity of the tests and in *not less than 90%* their sensitivity. The consequent false negative and false positive classified individuals can be

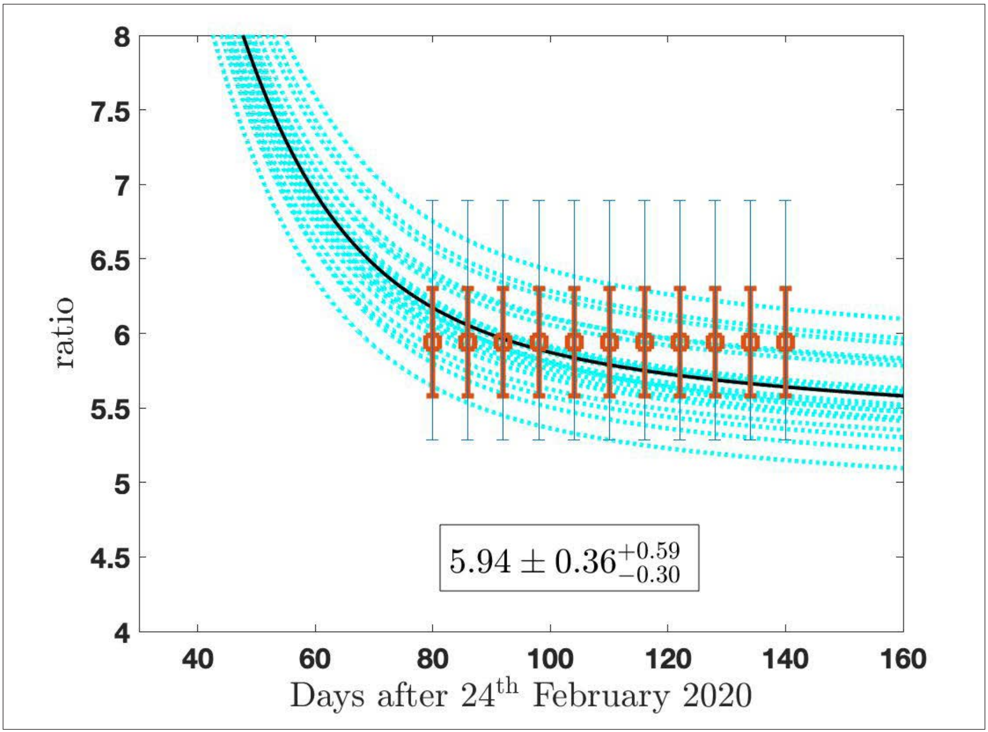
7 Our translation. "*I dati di siero-prevalenza a livello regionale, da integrare con quelli di sorveglianza epidemiologica, sono particolarmente preziosi sia per conoscere la quota di popolazione che è stata infettata nei mesi precedenti, sia per la messa a punto di programmi sanitari al fine di prevenire future ondate dell'epidemia e orientare adeguatamente le politiche sanitarie*" (in Italian). (Istat, 2020: page 1).

taken into account increasing the uncertainties of the ratio (14) of an amount of 10% and 5% in the two directions, assuming the maximum error in the data. One gets:

$$\text{ratio}|_{\text{survey}} = 5.94 \pm 0.36^{+0.59}_{-0.30}, \tag{15}$$

which replaces the value (14) to take into account the (asymmetric) corrections due to false responses. Figure 7 is, consequently, replaced by Figure 8 where the model results are compared with the more elaborated analysis. The asymmetric increase of the error-bars is clearly shown, however, at the same time, the conclusions of Section 3.1 remain basically valid.

Figure 8 - The ratios of Equations (14) and (15) compared with the cumulative model ratio of Equation (13) (cyan curves) (a)



Source: Our processing on Istat 2020 data

(a) The smallest error bars take into account the statistical variation suggested by the Istat report (Istat, 2020). The largest error bars include the modifications induced by the sensitivity and specificity of the IgG tests as indicated by the same Istat report.

4. Critical aspect and limits

The present Section is devoted to the critical aspects of the results and methods used in the present investigation from two points of view: the data and the model. From the point of view of the data an important element is still not well fixed: the time dependence of the data during their collection, and it will be discussed in the next Section 4.1. From the point of view of the model study one cannot remain only with its analysis of the past and in Section 4.2 a study of the present (and future) distribution of the reported cases and of the asymptomatic cases will be discussed. A stringent test for the model.

4.1 Time dependence of antibody tests

A recent meta-analysis by Deeks *et al.* (Deeks, 2020) observed substantial heterogeneity in sensitivities of IgA, IgM and IgG antibodies, or combinations, for results aggregated across different time periods post-symptom onset. They based the main results of the review on the 38 studies that stratified results by time since symptom onset[8].

In particular IgG/IgM all showed low sensitivity during the first week since onset of symptoms (all less than 30.1%), rising in the second week and reaching their highest values in the third week. The combination of IgG/IgM had a sensitivity of 30.1% (95% CI 21.4 to 40.7) for 1 to 7 days, 72.2% (95% CI 63.5 to 79.5) for 8 to 14 days, 91.4% (95% CI 87.0 to 94.4) for 15 to 21 days. Estimates of accuracy beyond three weeks are based on smaller sample sizes and fewer studies. For 21 to 35 days, pooled sensitivities for IgG/IgM were 96.0% (95% CI 90.6 to 98.3). There are insufficient studies to estimate sensitivity of tests beyond 35 days post-symptom onset. Summary specificities (provided in 35 studies) exceeded 98% for all target antibodies with confidence intervals no more than 2 percentage points wide. Assuming as a reference point the results of the meta-analysis by Deeks *et al.*, one must correct further the uncertainties of Figure 8. The questionnaire filled by the people involved in the screening for the Istat survey (Istat, 2020) included questions on the exact period of the symptom onset, a relevant piece of information in order to correct the data and to analyse the effects of the time-dependent sensitivity of

8 The numbers of individuals contributing data within each study each week are small and are usually not based on tracking the same groups of patients over time.

the tests. However information is not available at the moment, consequently it is necessary to introduce more drastic assumptions and corrections. One remains with the simple assumption that sample has no privilege with respect to the time dependence of the sensitivity. As a consequence the single test has to be considered (in average) affected by the average value of the sensitivity among 30.1% (days 1-7), 72.2% (days 8-14), 91.4% (days 15-21), 96% (days 22-35), and 90% for days > 35 (and till the end of the screening period, *i.e.* the remaining 17 days). 90% is indeed the minimum value of the sensitivity proposed by Istat and discussed in Section 3.2:

$$\frac{7 \times 30.1 + 7 \times 72.2 + 7 \times 91.4 + 14 \times 96.0 + 17 \times 90.0}{7 + 7 + 7 + 14 + 17} = \frac{4229.9}{52} \approx 81.3. \qquad (16)$$

The sensitivity of the test decreases from *not less than 90%* to 81.3%, while the specificity remains not less than 95%.

Taking into account the new estimated sensitivity (16), the ratio (15) is replaced by

$$\text{ratio}|_{\text{survey}} = 5.94 \pm 0.36^{+1.11}_{-0.30}, \qquad (17)$$

and the Figure 8 by the Figure 9. The asymmetric increase of the error-bars is again clearly shown, and the conclusions drawn in Section 3.1 become more weak.

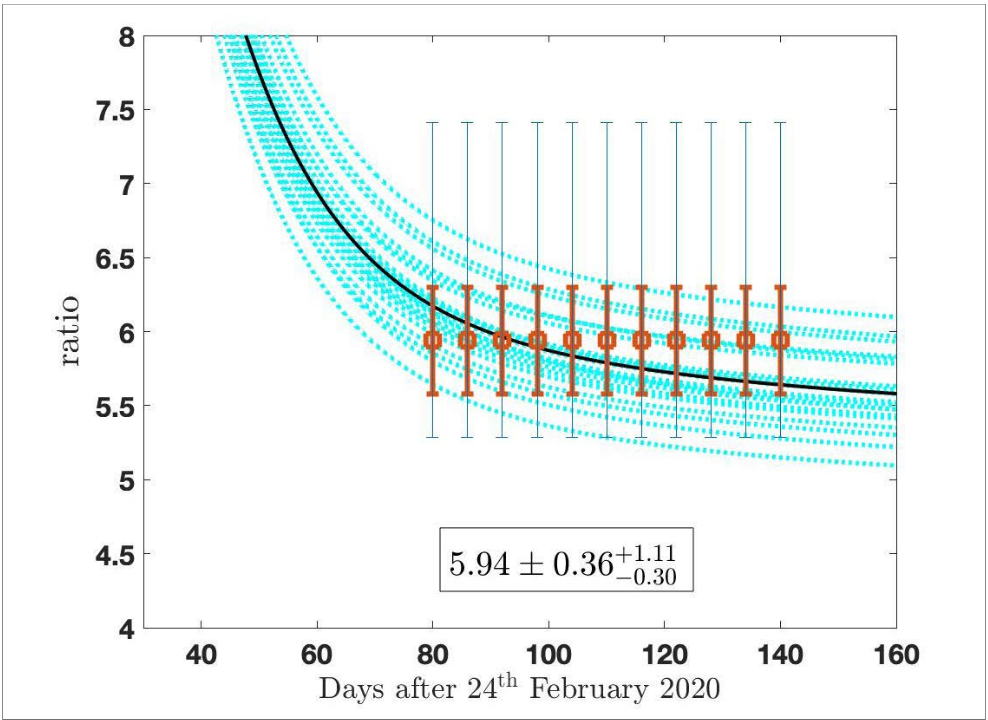
1. The comparison between the model estimates and the data corrected for the time dependence of the sensitivity of the tests as in Figure 9 is less accurate and the data cannot be considered a stringent test for models. The estimate of the asymptomatic prevalence remains valid, but with a larger interval of confidence.
2. Model and corrected data are still consistent for the whole time period.
3. The validity of the model description remains a guide in the interpretation of the unknown asymptomatic distribution within a larger interval of values. A more careful analysis of the Istat data, in particular the reference to the mentioned questionnaire (Istat, 2020) can help in making again the comparison more stringent.

4.2 Longer term study: last months of the year 2020

The present Section is devoted to the validation of the model (and predictions) with the most recent data on the second epidemic wave in Italy (after 25th September 2020), in particular the predictions for the asymptomatic compartment. The parametrisation is again based on values indicated in Table 2, no *ad hoc* modifications are introduced. The model is simply normalised to the value of the reported cases of 25th September (47,718), as indicated in the official site of the Italian Ministry of Health (*Ministero della salute*, 2020).

25th September 2020 is assumed as the beginning of the second wave of infection. The new results are summarised in Figure 10 (analogous to the Figure 5 describing the previous period) and Figure 11 (analogous to Figure 2).

Figure 9 - The ratios of Equations (14) and (17), are compared with the cumulative model ratio of Equation (13) (cyan curves) (a)



Source: Our processing

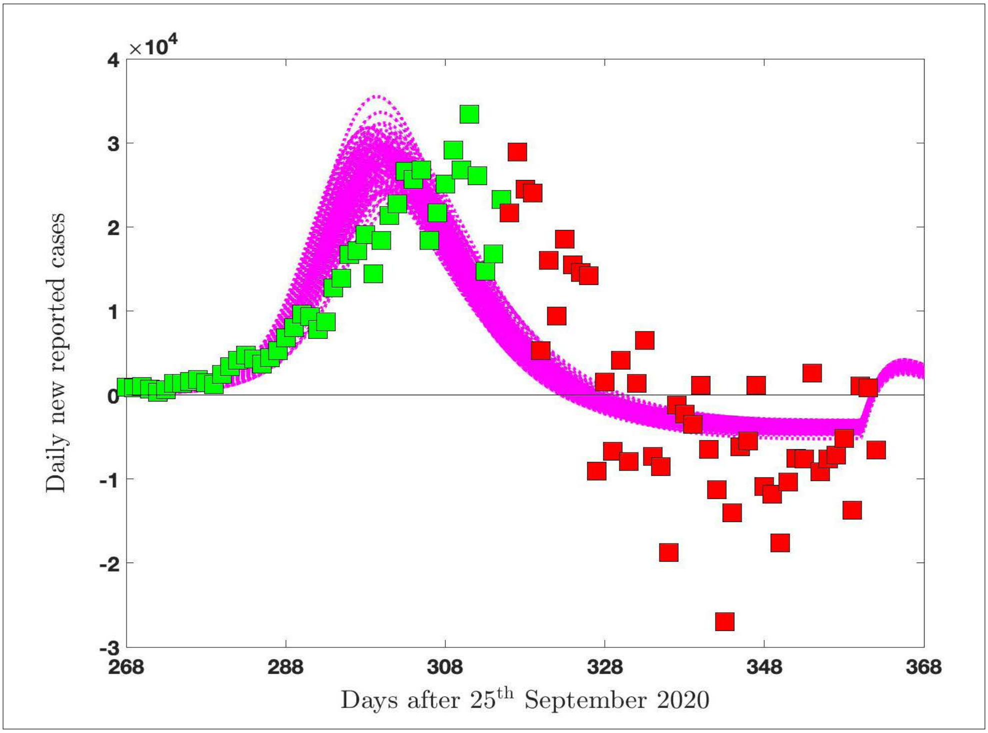
(a) The smallest error bars take into account the statistical variation suggested by the Istat report (Istat, 2020). The largest error bars include the modifications induced by the sensitivity and specificity of the IgG and their time-dependence as elaborated by Deeks *et al.*, 2020.

The social measures adopted in the second period are rather different from the strict lockdown of the first events, in particular the measures assumed are locally differentiate and following the local virulence of the infection. The new approach of the governmental institutions assumes a strategically flexible response to the virus allowing for a possible “coexistence”. The response is therefore more rigid and without rapid variations. A behaviour compatible with the results shown in Figure 10 where the data for the daily variations exhibit a broader aspect in comparison with the model predictions. Despite such a differences, the model remain rather consistent with data of Figure 10 which represent a stringent test since they are related to the derivative of the distributions. In addition the data are submitted to large fluctuations due to data collection variation and local inhomogeneities. The data are presented in two different colors: green (till 13^{th} November 2020) and red afterwards (last update 28^{th} December 2020). The simple raison is that the model has been applied for the first time to the new data in 13^{th} November 2020. After 13^{th} November the additional official data have been added to this Figure, day by day, and no modification or renormalisation is introduced, *i.e.* the model predictions are fixed.

Much more stable are the results of Figure 11 where data points (by Ministero della salute, 2020) for the daily reported cases are shown as a function of time (days), from 25^{th} September on (last update 28^{th} December). Once again the model predictions are compatible with data to a large extent. Such a consistency allows again a risky prediction for the unreported (asymptomatic and symptomatic) cases. The prevalence is rather large (approximately a factor of 4) quite similar to the prevalence emerging from the analogous Figure 2 (factor of 5 roughly, see also Figure 6). Obviously one needs detailed serological tests to have confirmation, however the stability of the virus is assumed by the parametrisation of Table 2 and seems to be consistent with data.

A last piece of information emerges from Figure 11: an indication of the possible effect due to the Christmas period. In Figures 11 and 10 one can see the effect of a foreseeable relaxation of the distancing measures in the behaviour of families and friends group. An average increase of the contact rate (as indicated in Table 2) produces a small wave emerging just after 25^{th} December 2020. A graphical indication of a possible increase of the number of infected.

Figure 10 - The model predictions for the daily new reported cases 25th September - 31st December 2020 in Italy, compared with the official data of the Ministry of Health

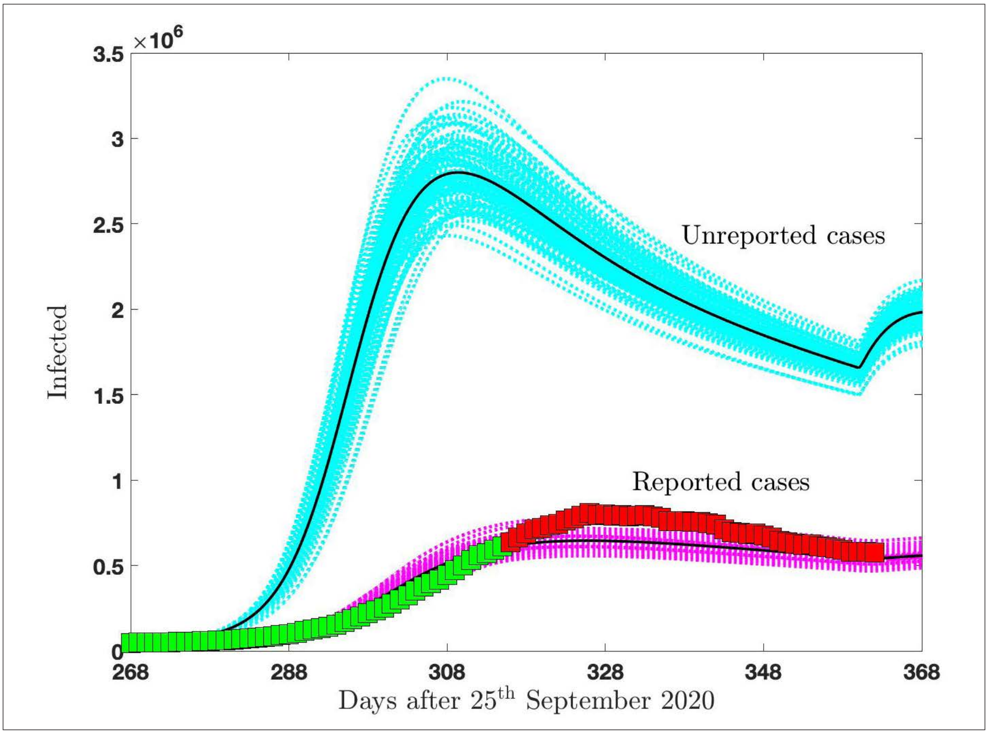


Source: Our processing using the Italian Ministry of Health data, 2020 (*Ministero della Salute*, 2020)

5. A look at vaccination

The present Section is devoted to a burdensome aspect of the nowadays discussion[9] on SARS-CoV-2: vaccination. The United Kingdom started the vaccination campaign earlier, while all the European countries started their own campaign on 27th December 2020. The organisation of a massive historical event is rather heavy and it will need the effort of many institutions.

Figure 11 - The MCMC analysis of the infected individuals (reported and unreported cases) in Italy as a function of time in the period 25th September - 31st December 2020 (a)



Source: The data for the reported-infected individuals are by the Italian Ministry of Health (*Ministero della Salute*, 2020) and are compared with our model predictions (magenta curves) which include statistical uncertainties.
(a) The reported cases are compared with the predictions for the class of infected non reported cases (asymptomatic and symptomatic) in the same period of time (cyan curves).

9 21st December 2020.

5.1 Modelling vaccination: scenarios

A preliminary aspect is the assessment of a possible vaccination scenario. We do not discuss specific strategies, but we want to establish general scenarios to illustrate main advantages and disadvantages within simple assumptions. Basically we will assume that the year 2021 will be devoted to vaccination in Italy and that the order of magnitude of the immunised persons per day is between 40,000 and 80,000 (to be selected with specific social and homogeneity criteria) (*e.g.* Makoul, 2020).

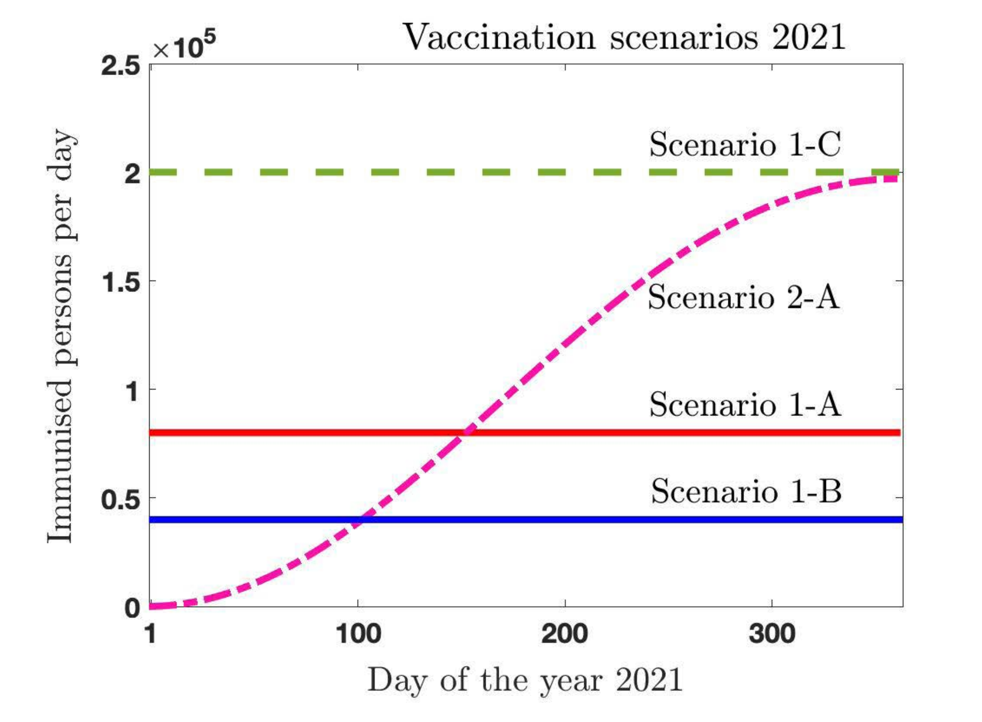
In the following two basic scenarios will be introduced, as illustrated in Figure 12.

i) The first scenario assumes an homogenous distribution, during the 2021, of a fixed number of immunised persons per day: 80,000 (scenario 1-A), (full line in Figure 12), and also 40,000 (scenario 1-B) will be considered. The effort to start with a large portion of population is high and a second scenario is investigated;

ii) The second scenario assumes that the amount of immunised persons during the first period is rather low and the process will accelerate during the year (dotted line in Figure 12, scenario 2-A).

iii) Almost 30 million people are immunised at the end of the year in both scenarios.

Figure 12 - A constant number of person are immunised per day during the year 2021; 80,000, full curve scenario (1-A); 40,000, scenario (1-B); 200,000, scenario (1-C) (a)



Source: Our processing
(a) Alternatively, in scenario (2-A), an increasing number of people are immunised during 2021 allowing for a less strong starting effort. The total number of immunised people results (almost) 30 million for both scenarios, at the end of the year.

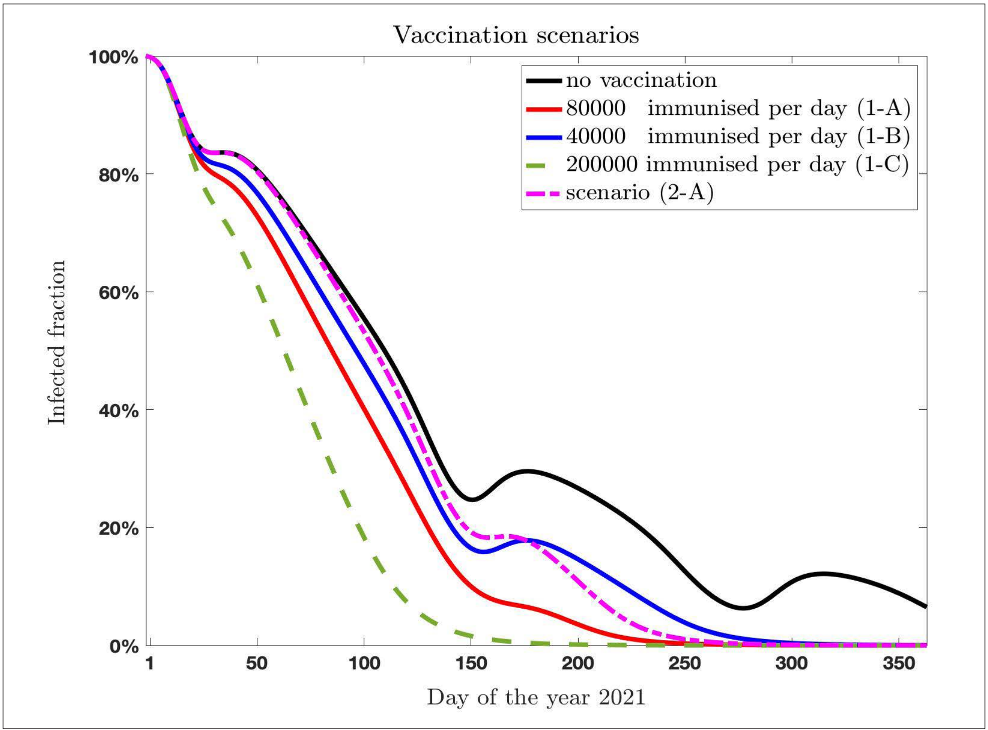
5.2 Year 2021: a perspective

The Section is devoted to a general perspective of the year 2021 offering a macroscopic view of the pandemic event in Italy and the advantages of the vaccination campaign. In Figure 13 one can have an approximated idea of the time-evolution of the SARS-CoV-2 infection in Italy during the year[10] in the case

10 The qualitative intent of the present discussion is put in further evidence by the fact that no-Covid deaths and newborns are not included in the investigation. One is simply assuming that they compensate approximately during the 2021. In addition we will not discuss vaccine efficacy assuming $VE = 95\%$ for two reasons: i) we do not know the real efficacy of the different vaccines; ii) it is rather simple to rescale results for an accepted efficacy (Moghadas, 2020).

of *no vaccination*, an assumption which is ruled out, but it can offer a reference point to appreciate the advantages of the vaccines. Such a scenario is described by the full black line (Figure 13). One can immediately see that the number of the daily reported cases is decreasing during the year since an increasingly large number of "susceptible" individuals get immunised (or dead) after the infected period. The year 2021 will not be sufficient to obtain a vanishing number of infected: in summer they will be around 150,000 and during the next Christmas period roughly 50,000. A rough estimate which takes into account

Figure 13 - The fraction of the reported cases as a function of the day of the year 2021 (a)



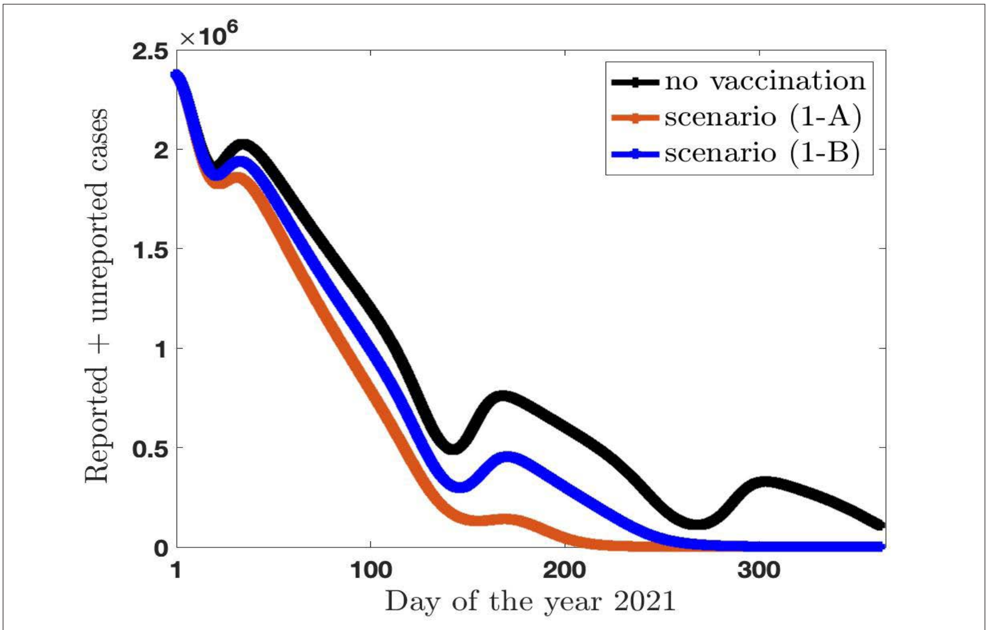
Source: Our processing

(a) On 1st January they were 574,767, equivalent to 100%. The scenario assuming no vaccination (continuous black line) is compared with the scenario (1-A) which assumes a constant rate of vaccines of 80,000 immunised each day (continuous red line). The blu line, scenario (1-B), shows results for 40,000 immunised per day at constant rate. The scenario (2-A), where the rate is not constant and a low starting period is compensated by an accelerated second immunisation period (see Figure 12), is illustrated by the dotdashed (magenta) curve.

oscillations due to possible waves of infections[11] as it emerges from the wave behaviour of the curve. The situation changes largely with the introduction of the vaccination at a rate of 80,000 persons per day (red curve, scenario (1-A)). A reduction of roughly 100,000 cases is already evident in April, while in summer (June) one reaches the 20,000 cases instead of 150,000, in August the virus will give a definitive "*ciao*". As a matter of fact one is gaining several months (from six to ten months) because of vaccination without counting the number of deaths.

The scenario (1-B) (blu line) assumes half immunised persons per day (40,000) and the disadvantages are evident with respect the previous hypothesis of 80,000. One has to wait end of October to eliminate the virus infection and in June one has still 80,000 cases.

Figure 14 - Both reported and unreported cases as a function of the day of the year 2021 (a)



Source: Our processing
(a) The scenario assuming no vaccination (full black line) is compared with the scenario (1-A) which assumes a constant rate of vaccines of 80,000 immunisations each day (full red line). The blu line, scenario (1-B), shows results for 40,000 immunised per day at constant rate.

11 The oscillating behaviour is simulated by an oscillating contact rate $c(t) = c_0 + 10\ c_0 \cdot \cos^2(\Omega t)$, with $0 \leq t \leq 365$ and $\Omega = 4\pi/365$, see Table 2.

Also the scenario (2-A) has been implemented in the model and one can see the effects of a slow rate at the beginning of the year in the lower panel of the same Figure 13. The results of the vaccination are almost invisible till the end of May. They appear rapidly in the second part of the year producing the final result at the end of October. The summer is still a hard period and scenario (2-A) is rather similar to scenario (1-B).

Looking at Figure 13 is not intuitive to accept that the reduction of the cases from 500,000 to zero in one single year can produce the end of the infection in Italy with 60 million of residents.

To partially restore intuition, one has to realise that, after the discussion of the asymptomatic cases of Section 3, a large part of the job is done by the unreported cases. In Figure 14 the behaviour of the total (unreported + reported cases) during the year as indicated by the model. They complement Figure 13 showing that the decreasing behaviour of the reported cases has a specular behaviour for the asymptomatic. The vaccine helps in a strong way the reduction of both.

6. Concluding remarks

The traditional compartmental classes of Susceptible, Exposed, Infected and Recovered which characterise a large fraction of epidemical models, are not sufficient to simulate the coronavirus (SARS-CoV-2) pandemic infection. Many of the people who contract the diseases are Asymptomatic, they are infected and contagious and are often invoked as one of the causes for the rapid spread of the infection. It is hard to estimate the amount of asymptomatic, estimates ranges from 40% to 75% (Buitrago-Garcia, 2020; Oran, 2020; Day, 2020; Muzimoto, 2020) of the total infected population. In a small scale the particular conditions of the Vo' village (near Padua, Italy) allowed for two detailed surveys after a localised lockdown: the analysis found a prevalence of 1.2% (95% CI:0.8-1.8%). Notably, 42.5% (95% CI:31.5-54.6%) of the confirmed SARS-CoV-2 infections were asymptomatic (without any symptoms at the time of positive swab testing, and did not develop symptoms afterwards (Lavezzo, 2020)).

In August 2020, the Italian National Institute of Statistics – Istat presented the preliminary results of seroprevalence survey on the percentage of individuals reached by the SARS-CoV-2 infection in the previous months. The survey aims at estimating, in a methodological precise way, also the asymptomatic infected population and its role in the infection spread in Italy, one of the most affected countries in Europe.

The present study investigates a model description of the entire infected population in Italy considering a eightfold compartmental model which includes Infected, Asymptomatic and Quarantined population in addition to the more classic Susceptible, Recovered, Exposed. The model is illustrated in few examples and used to investigate in detail the asymptomatic prevalence including also sensitivity and specificity of the antibody tests used for the surveys. The results validate the model description and encourage other studies to detail, in a more quantitative way, the role of time-dependence of the sensitivity of the test used for the antibody screening.

Also predictions for the period of the second wave in Italy are presented and discussed, including asymptomatic predictions. Despite the fact that the epidemiological surveillance of the second wave of the epidemic event in Italy is characterised by a strong use of swabs, the model predictions for the amount of unreported cases is of the same order of magnitude of the percentage already seen in the first outbreak. Future antibody screening will verify the present prediction.

To complete modelling, Section 5 has been devoted to the vaccination scenarios. The role played by the introduction of vaccination is clearly shown in reducing in a significative way the time to reach the end of the infection. The strategies of the administration can be simulated and they favour an intense administration from the beginning. Starting with a low rate of administration the risk is to loose a large part of the advantage.

References

Ansumali, S., S. Kaushal, A. Kumar, M.K. Prakash, and M. Vidyasagar. 2020. "Modelling a pandemic with asymptomatic patients, impact of lockdown and herd Immunity, with applications to SARS-CoV-2". *Annual Reviews in Control*, Volume 50: 432-447.

Bassi, F., G. Arbia, and P.D. Falorsi. 2020. "Observed and estimated prevalence of COVID-19 in Italy: How to estimate the total cases from medical swabs data". *Science of the Total Environment*, Volume 764: 142799.

Bobrovitz, N., R.K. Arora, T. Yan, H. Rahim, N. Duarte, E. Boucher, J. Van Wyk, and T. Grant Evans. 2020. "Lessons from a rapid systematic review of early SARS-CoV-2 serosurveys". *medRxiv – The Preprint Server for Health Sciences*.

Brooks, S., A. Gelman, G. Jones, X.-L. Meng. 2011. *Handbook of Markov Chain Monte Carlo*. Milton Park, U.K.: Chapman & Hall/CRC.

Bryan, A., G. Pepper, M.H. Wener, S.L. Fink, C. Morishima, A. Chaudhary, K.R. Jerome, P.C. Mathias, AL. Greninger, and A.J. McAdam (*ed.*). 2020. "Performance Characteristics of the Abbott Architect SARS-CoV-2 IgG Assay and Seroprevalence in Boise, Idaho". *Journal of Clinical Microbiology*, 58: e00941-20.

Buitrago-Garcia, D., D. Egli-Gany, MJ. Counotte, S. Hossmann, H. Imeri, A.M. Ipekci, G. Salanti, and N. Low. 2020. "Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis". *PLoS Medicine*, 17(9): e1003346.

Day, M. 2020. "COVID-19: Identifying and isolating asymptomatic people helped eliminate virus in Italian village". *The BMJ*, 368: m1165.

Deeks, J.J., J. Dinnes, Y. Takwoingi, C. Davenport, *et al.* 2020.

Deeks, J.J., J. Dinnes, Y. Takwoingi, C. Davenport, R. Spijker, S. Taylor-Phillips, A. Adriano, S. Beese, J. Dretzke, L. Ferrante di Ruffano, I.M. Harris, M.J. Price, S. Dittrich, D. Emperador, L. Hooft, M.Mg. Leeflang, A.Van den Bruel, and Cochrane COVID-19 Diagnostic Test Accuracy Group. 2020. "Antibody tests for identification of current and past infection with SARSCoV-2”. *The Cochrane Database of Systematic Reviews*, 6(6): CD013652.

Doi, A., K. Iwata, H. Kuroda, T. Hasuike, S. Nasu, A. Kanda, T. Nagao, H. Nishioka, K. Tomii, T. Morimoto, and Y. Kihara. 2020. “Estimation of seroprevalence of novel coronavirus disease (COVID-19) using preserved serum at an outpatient setting in Kobe, Japan”. *medRxiv – The Preprint Server for Health Sciences*.

Erikstrup, C., C.E. Hother, O.B.V. Pedersen, K. Mølbak, R.L. Skov, D.K. Holm, S.G. Sækmose, A.C. Nilsson, P.T. Brooks, J.K. Boldsen, C. Mikkelsen, M. Gybel-Brask, E. Sørensen, K.M. Dinh, S. Mikkelsen, B.K. Møller, T. Haunstrup, L. Harritshøj, B.A. Jensen, H. Hjalgrim, S.T. Lillevang, and H. Ullum. 2020. “Estimation of SARS-CoV-2 infection fatality rate by real-time antibody screening of blood donors”. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, Volume 72(2): 249-253.

Gandhi, M., D.S. Yokoe, and D.V. Havir. 2020. “Asymptomatic Transmission, the Achilles’ Heel of Current Strategies to Control COVID-19”. *The New England Journal of Medicine*, 382: 2158-2160.

Garcia-Basteiro, A.L., G. Moncunill, M. Tortajada, et al. 2020. “Seroprevalence of antibodies against SARS-CoV-2 among health care workers in a large Spanish reference hospital”. *Nature Communication*, 11: Article n. 3500 (2020).

Giordano, G., F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, and M. Colaneri. 2020. “Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy”. *Nature Medicine*, 26. 855-860.

Guerriero, M, Z. Bisoffi, A. Poli, C. Micheletto, and C. Pomari. 2020. “Prevalence of asymptomatic SARS-CoV- 2- positive individuals in the general population of northern Italy and evaluation of a diagnostic serological ELISA test: a cross-sectional study protocol”. BMJ Open, 10(10): e040036.

Hogg, D.W., and D. Foreman-Mackey. 2018. “Data Analysis Recipes: Using Markov Chain Monte Carlo”. The Astrophysical Journal, Supplement Series, Volume 236: 11.

Istituto Nazionale di Statistica – Istat (Italian National Institute of Statistics), and Ministero della Salute (Ministry of Health). 2020. "Primi risultati dell'indagine di sieroprevalenza sul SARS-CoV-2". Comunicato Stampa. Roma, Italia: Istat. https://www.istat.it/it/archivio/246156.

Jia, X., J. Chen, L. Li, N. Jia, B. Jiangtulu, T. Xue, L. Zhang, Z. Li, R. Ye, and B. Wang. 2020*a*. "Modeling the Prevalence of Asymptomatic COVID-19 Infections in the Chinese Mainland". The Innovation, 1: 100026.

Jia, X., L. Xiao, Y. Liu, 2020*b*. "False negative RT-PCR and false positive antibody tests-Concern and solutions in the diagnosis of COVID-19". *The Journal of Infection*, Volume 82(3): 414-451.

Kinoshita, R., A. Anzai, S.-M. Jung, N.M. Linton, T. Miyama, T. Kobayashi, K. Hayashi, A. Suzuki, Y. Yang, A.R. Akhmetzhano, and H. Nishiura. 2020. "Containment, Contact Tracing and Asymptomatic Transmission of Novel Coronavirus Disease (COVID-19): A Modelling Study". *The Journal of Clinical Medicine*, 9 (10). 3125.

Lavezzo, E., E. Franchin, C. Ciavarella, et al. 2020. "Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'". *Nature*, 584: 425-429.

Little, R.J.A. 1993. "Post-Stratification: A Modeler's Perspective". J*ournal of the American Statistical Association*, Volume 88, N. 423: 1001-1012.

Makoul, M., H.A. Houssein, H. Chemaitelly, S. Seedat, G.R. Mumtaz, S. Al-Omari, and L.J. Abu-Raddad. 2020. "Epidemiological Impact of SARS-CoV-2 Vaccination: Mathematical Modeling Analyses". *Vaccines*, 8 (4): 668.

Ministero della Salute – Italian Ministry of Health. 2020. *Portale: Nuovo Coronavirus*. Roma, Italia: Ministero della Salute. http://www.salute.gov.it/portale/nuovocoronavirus/homeNuovoCoronavirus.jsp

Mizumoto, K., K. Kagaya, A. Zarebski, and G. Chowell. 2020. "Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020". *Eurosurveillance Journal*. Volume 25, Issue 10: 2000180.

Moghadas, S.M., T. N. Vilches, K. Zhang, C.R. Wells, A. Shoukat, B.H. Singer, L.A. Meyers, K.M. Neuzil, J.M. Langley, M.C. Fitzpatrick, and A.P. Galvani. 2020. "The impact of vaccination on COVID-19 outbreaks in the United States". *Clinical Infectious Diseases*: ciab079.

Oran, D.P., and E.J. Topol. 2020. “Prevalence of Asymptomatic SARS-CoV-2 Infection”. *Annals of Internal Medicine*, 174(2): 286-287.

Park, S.W., D.M. Cornforth, J. Dushoff, and J.S. Weitz. 2020. “The time scale of asymptomatic transmission affects estimates of epidemic potential in the COVID-19 outbreak”. *Epidemics*, N. 31: 100392.

Peterson, I., and A. Phillips 2020. “Three Quarters of People with SARS-CoV-2 Infection are Asymptomatic: Analysis of English Household Survey Data”. *Clinical Epidemiology*, Volume 12: 1039-1043.

Pollán, M., B. Pérez-Gómez, R. Pastor-Barriuso, J. Oteo, M.A. Hernán, M. Pérez-Olmeda, J.L. Sanmartín, A. Fernández-García, I. Cruz, N. Fernández de Larrea, M. Molina, F. Rodríguez-Cabrera, M. Martín, P. Merino-Amador, J.L. Paniagua, J.F. Muñoz-Montalvo, F. Blanco, and R. Yotti. 2020. “Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study”. *The Lancet*. Volume 396, Issue 10250: 535-544.

Remuzzi, A., and G. Remuzzi. 2020. “COVID-19 and Italy: what next?”. *The Lancet*, Volume 395, Issue 10231: 1225-1228.

Robinson, R., and N.I. Stilianakis. 2013. “A model for the emergence of drug resistance in the presence of asymptomatic infections”. *Mathematical Biosciences*, 243(2): 163-177.

Salje, H., C.T. Kiem, N. Lefrancq, P. Bosetti, J. Paireau, A. Andronico, N. Hozé, J. Richet, C.-L. Dubost, Y. Le Strat, J. Lessler, D. Levy-Bruhl, A. Fontanet, L. Opatowski, P.-Y. Boelle, and S. Cauchemez. 2020. “Estimating the burden of SARS-CoV-2 in France”. *Science*, Volume 369, Issue 6500: 208-211.

Shakiba, M., M. Nazemipour, A. Salari, F. Mehrabian, S.S.H. Nazari, S.M. Rezvani, Z. Ghasempour, A. Heidarzadeh, and M.A. Mansournia. 2020. “Seroprevalence of COVID-19 virus infection in Guilan province, Iran”. *Emerging Infectious Diseases*, 27(2): 636-638.

Snoeck, C.J., M. Vaillant, T. Abdelrahman, V.P. Satagopam, et al. 2020. “Prevalence of SARS-CoV-2 infection in the Luxembourgish population - the CON-VINCE study”. *medRxiv – The Preprint Server for Health Sciences*.

Sood, N., P. Simon, P. Ebner, D. Eichner, J. Reynolds, E. Bendavid, and J. Bhattacharya. 2020. "Seroprevalence of SARS-CoV-2-specific antibodies among adults in Los Angeles County, California, on April 10-11, 2020". *JAMA*, 323(23): 2425-2427.

Stringhini, S., A. Wisniak, G. Piumatti, A.S. Azman, S.A. Lauer, H. Baysson, D. De Ridder, D. Petrovic, S. Schrempft, K. Marcus, S. Yerly, I.A. Vernez, O. Keiser, S. Hurst, K.M. Posfay-Barbe, D. Trono, D. Pittet, L. Gétaz, F. Chappuis, I. Eckerle, N. Vuilleumier, B. Meyer, A. Flahault, L. Kaiser, and I. Guessous. 2020. "Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study". *The Lancet*, Volume 396, Issue 10247: 313-319.

Tang, B., X. Wang, Q. Li, N.L. Bragazzi, S. Tang, Y. Xiao, and J. Wu. 2020. "Estimation of the Transmission Risk of the 2019-nCoV and Its Implication for Public Health Interventions". *Journal of Clinical Medicine*, Volume 9(2): 462.

Tang, B., N.L. Bragazzi, Q. Li, S. Tang, Y. Xiao, and J. Wu. 2020. "An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov)". *Infectious Disease Modelling*, Volume 5: 248-255.

Traini, M.C., C. Caponi, R. Ferrari, and G.V. De Socio. 2020. "A study of SARSCoV-2 epidemiology in Italy: from early days to secondary effects after social distancing". *Infectious Diseases*, Volume 52, Issue 12: 866-876.

Valenti, L., A. Bergna, S. Pelusi, F. Facciotti, A. Lai, M. Tarkowski, A. Berzuini, F. Caprioli, L. Santoro, G. Baselli, C. della Ventura, E. Erba, S. Bosari, M. Galli, G. Zehender, and D. Prati. 2020. "SARS-CoV-2 seroprevalence trends in healthy blood donors during the COVID-19 Milan outbreak". *medRxiv – The Preprint Server for Health Sciences*.

Wells, P.M., J.K. Doores, S. Couvreur, R.M. Nunez, J. Seow, C. Graham, S. Acors, N. Kouphou, S.J.D. Neil, R.S. Tedder, P.M. Matos, K. Poulton, M.J. Lista, R.E. Dickenson, H. Sertkaya, T.J.A. Maguire, E.J. Scourfield, R.C.E Bowyer, D. Hart, A. O'Byrne, K.J.A. Steel, O. Hemmings, C. Rosadas, M.O. McClure, J. Capedevilla-Pujol, J. Wolf, S. Ourselin, M.A. Brown, M.H. Malim, T. Spector, and C.J. Steves. 2020. "Estimates of the rate of infection and asymptomatic COVID-19 disease in a population sample from SE England". *The Journal of Infection*, Volume 81, Issue 6: 931-936.

Wu, X., B. Fu, L. Chen, and Y. Feng. 2020. "Serological tests facilitate identification of asymptomatic SARS-CoV-2 infection in Wuhan, China". *Journal of Medical Virology*, Volume 92, Issue 10: 1795-1796.

Yanes-Lane, M., N. Winters, F. Fregonese, M. Bastos, S. Perlman-Arrow, J.R. Campbell, and D. Menzies. 2020. "Proportion of asymptomatic infection among COVID-19 positive persons and their transmission potential: A systematic review and meta-analysis". *PLoS ONE*, Volume 15(11): e0241536.

Yu, X. 2020. "Modeling return of the epidemic: Impact of population structure, asymptomatic infection, case importation and personal contacts". *Travel Medicine and Infectious Disease*, Volume 37: 101858.

Analysing complications of COVID-19 from death certificates: which ones kill most?

Chiara Orsi [1], Daniele De Rocchi [2], Simona Cinque [1], Roberta Crialesi [1], Vincenzo Della Mea [3], Luisa Frova [1], Enrico Grande [1], Stefano Marchetti [1], Simone Navarra [1], Marilena Pappagallo [1], Silvia Simeoni [1], Francesco Grippo [1]

Abstract

Death certificates compiled by physicians are the basis for cause-of-death statistics and contain the whole sequence of diseases leading to death. This information is useful for understanding pathways from disease to death. Several complications are associated to COVID-19, nevertheless limited data are available on the frequency of these conditions leading to death. For identifying complications of COVID-19, a method was developed and applied to deaths occurred in Italy in March-April 2020. In order to test the method on other causes, analyses have been carried-out also for pneumonia and diabetes related deaths occurred in 2018. Pneumonia is the most frequent complication of COVID-19 together with some other respiratory conditions; besides those, some other conditions were identified as linked to COVID-19 related deaths such as cardiac complications, shock and infections. The method proves to be powerful in distinguishing complications and causes of diseases.

Keywords: COVID-19, comorbidity, causes of death, death certificates.

1 Italian National Institute of Statistics - Istat.
2 National Research Council - Institute for Research on Population and Social Policies (CNR - IRPPS).
3 Università degli Studi di Udine, Italy.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction[4]

COVID-19 is characterised by a high rate of lethality. In patients infected by SARS-CoV-2, death can occur as the consequence of various complications, especially pneumonia and other respiratory system conditions. In addition, the presence of comorbidities, as hypertension, diabetes, cardiovascular diseases, and chronic respiratory conditions, can increase vulnerability to the development of these complications (Zheng *et al.*, 2020; Palmieri *et al.*, 2020).

Death certificates are considered the most reliable source of information for comparing cause-specific mortality across different populations and countries. However, limited data are available on the complications and comorbidities reported on death certificates of patients who presented COVID-19. The UK Office of National Statistic conducted some analyses of conditions existing prior to contracting COVID-19 (ONS, 2020), but the complications of COVID-19 on death certificates are less explored.

Cause-of-death (CoD) data are based on the information reported by physicians on a standardised death form recommended by the World Health Organization (WHO international format of death certificate). Data collected with the death certificates are coded according to provisions of the International Classification of Diseases, 10th revision (ICD10) (WHO, 2019): the corresponding ICD10 code is assigned to each condition reported on the certificate. Despite certificates generally report several causes, CoD statistics are traditionally based on the concept of underlying cause (UC), *i.e.* the disease or condition that initiated the train of events leading to death (WHO,

4 The method described was developed in the frame of a project aimed at evaluating the effort needed for the implementation of the International Classification of Diseases, 11th revision (ICD11) in the automated coding systems Iris (www.Iris-institute.org).
Chiara Orsi and Francesco Grippo prepared the draft and contributed to the analysis of data and results interpretation. Francesco Grippo, Daniele De Rocchi and Chiara Orsi conceived the method and contributed to its development. Daniele De Rocchi and Francesco Grippo developed the SAS programme; Simona Cinque, Enrico Grande, Francesco Grippo, Simone Navarra, Chiara Orsi and Silvia Simeoni contributed to ICD10 data coding of 2020 mortality data. All authors contributed to the revision of the text and agreed to this published version of the manuscript.
All authors thank the members of the Istat task force on cause-of-death analysis during COVID-19 pandemic, in particular: Tania Bracci, Giulia Marcone, Simona Pennazza, Paola Rocchi and Alessia Scuro for their contribution to ICD10 data coding; Donatella Braca Sbaccanti, Marina Cecchini and Stefano Petrolo for their contribution to the timely collection of certificates and to the verification and correction of demographic variables; Claudia Bianchi, Giancarlo Gualtieri and Patrizia Incutti for the relations with local authorities and service provider companies.

2019). Since several years, in Italy, as in other countries, automated coding systems are in use for CoD coding and selecting the UC. In Italy the software Iris is used, which is the most widely used. This software follows rules and provisions established by ICD10.

Death certificates consist of two parts: in part 1 the certifier reports the sequence of diseases or events leading directly to death, in part 2 he lists other significant conditions contributing to death. Part 1 of the death certificate includes four different lines: conditions reported on each line are due to what is reported on the line below. The condition that initiated the sequence leading to death should be reported in the lowest used line. Therefore, filling the death certificates, certifiers provide information on causal relations among the conditions leading to death.

1.1 Objectives

This paper presents a method to individuate causal relations among conditions reported on death certificates. In particular, the method allows identifying, for a given condition, which are its complications or its antecedent causes in the death certificate. The paper describes the procedure in detail and shows the results of its application to identify the complications of COVID-19 as well as to provide a measure of the frequency of such complications in deaths occurred in Italy during the first wave of the pandemic (March-April 2020). To show the potential general applicability to other conditions, we also performed analyses on two selected causes of death: pneumonia and diabetes.

2. Methods

The method presented is an advancement of a procedure developed as part of a project for the new version of ICD (ICD11) implementation in automated coding systems (www.iris-institute.org; Orsi *et al.*, 2020).

2.1 Causal relations in death certificates

Figure 2.1 shows an example of filled international death certificate. The person died for a cardiac arrest due to a congestive heath failure due to an ischemic heart disease, and was affected by diabetes and hypertension. These latter conditions both contributed to death, but were not part of the sequence leading directly to death; for this reason, the certifier reported them in part 2. UC is the ischemic heart disease, which is the condition that initiated the train of events leading to death.

In general, given a condition, which we will hereafter name *A*, all conditions reported in lines above (we will name other conditions *S*) can be considered due to *A*. On the other hand, *A* can be considered due to conditions reported below. In example 1 of figure 2.2[5], with regard to the causal relations of congestive heart failure (condition A corresponding to the ICD10 code I500), cardiac arrest (condition S_1, I469) is due to congestive heart failure, while congestive heart failure is due to ischemic heart disease (S_2, I259).

However, it should be taken into account that sometimes the order reported on death certificates does not correspond to a real causal connection between diseases. For instance, in example 2 of figure 2.2, the certifier misplaced liver metastasis (ICD10 code C787) on a row that indicates this condition as due to pneumonia (J189) but this is an incorrect causal sequence.

Provided that the death certificates could contain two given conditions reported in both directions (*A due to S* as well as *S due to A*), we developed a method for the identification of recurrent causal patterns in multiple cause-of-death data.

5 For the examples shown in figure 1.2, to each condition is given a corresponding ICD10 code, so in this document, the terms condition and code are used with the same meaning.

Figure 2.1 - Example of filled international death certificate

FRAME A: ▸ Medical data: Part 1 and 2			
		▸ Cause of death	▸ Time interval from onset to death
1. Report disease or condition directly leading to death on line a Report chain of events in due to order (if applicable) State the underlying cause on the lowest used line	a	Cardiac arrest	
	b	Congestive heart failure	
	c	Ischemic heart disease	
	d	Due to:	
2. Other significant conditions contributing to death (time intervals can be included in brackets after the condition)		Diabetes, hypertension	

Figure 2.2 - Examples of filled death certificates

Example 1		ICD10 codes	**Example 4**		ICD10 codes
Part 1	a. Cardiac arrest	I469	Part 1	a. ARDS	J80
	b. Congestive heart failure	I501		b. Covid-19 pneumonia	U071 J189
	c. Chronic ischemic heart disease	I259		c.	
	d.			d.	
	e.			e.	
Part 2	Diabetes, hypertension	E149 I10	Part 2	Hypertensive heart disease	I119
*Iris multiple cause string: I469 I501/I259*E149 I10*			*Iris multiple cause string: J80/U071 J189*I119*		
Example 2		ICD10 codes	**Example 5**		ICD10 codes
Part 1	a. Liver metastasis	C787	Part 1	a. Sepsis	A419
	b. Pneumonia	J189		b. Pneumonia	J189
	c.			c.	
	d.			d.	
	e.			e.	
Part 2	Lung cancer	C349	Part 2	Covid19	U071
*Iris multiple cause string: C787/J189*C349*			*Iris multiple cause string: A419/J189*U071*		
Example 3		ICD10 codes	**Example 6**		ICD10 codes
Part 1	a. Acute respiratory failure	J960	Part 1	a. Covid19	U071
	b. Interstitial pneumonia, congestive heart failure	J849 I500		b.	
	c. Covid-19	U071		c.	
	d. Ischemic heart disease	I259		d.	
				e.	
Part 2			Part 2	Pneumonia and respiratory failure	J189 J969
Iris multiple cause string: J960/J849 I500/U071/I259			*Iris multiple cause string: U071*J189 J969*		

2.2 Overview of the method

The input data is the multiple cause string produced by the automated coding system Iris (shown in figure 1.2 below each certificate). This string consists of ICD10 codes and separators representing all conditions reported on the death certificate and their relative position.

The method focusses only on part 1 of certificates and, in summary, involves two steps:

1. causal relations: analysis of causal relations reported in part 1; a chi-square test is carried out in order to identify the preferred causal order between two conditions;
2. associations between conditions: analysis of associations between conditions in part 1; a second chi-square test identifies if two conditions are reported together in part 1 more frequently than expected. This step is used as a confirmation of the first.

The following paragraphs describe the two steps in detail.

2.3 Causal relations

This step is aimed at identifying the preferred causal order between two conditions reported by certifiers: given two conditions, *A* and *S*, a test is performed to verify if the pattern *A due to S* is significantly more frequent than *S due to A*.

For each pair of codes (*A* and *S*), certificates jointly mentioning the two codes in different lines of part 1 were selected, excluding those in which they are reported on the same line. The frequencies of certificates reporting the pattern *A due to S* and *S due to A* were calculated. Afterwards, these observed frequencies were compared with the expected ones calculated under the null hypothesis of equal probability for the two patterns, assuming that 50% of total certificates would have the pattern *A due to S* and 50% *S due to A* (table 2.1). The null hypothesis was tested by means of chi-square test (double tailed with 1 degree of freedom). Based on the results, we can distinguish the following cases.

- *A due to S* (positive due to relation between the two codes): the frequency of certificates reporting *A due to S* is significantly ($p<0.05$) higher than expected.
- *S due to A* (negative due to relation between the two codes): the frequency of certificates reporting *A due to S* is significantly ($p<0.05$) lower than expected.
- No due to relation between *A* and *S*: there is not a significant difference between observed and expected frequencies ($p>=0.05$).

Table 2.1 - Observed and expected frequencies of certificates for the analysis of causal relations between conditions (a)

Pattern of disposition of *A* and *S*	Observed	Expected
A due to S	$N_{AduetoS}$	$N_{AduetoS}^{exp} = \frac{1}{2} N_{AS}^{*}$
S due to A	$N_{SduetoA}$	$N_{SduetoA}^{exp} = \frac{1}{2} N_{AS}^{*}$
Total	N_{AS}^{*}	N_{AS}^{*}

(a) Subscripts $AduetoS$ and $SduetoA$ indicate respectively the presence of *A due to S* and *S due to A* on the death certificate.
N_{AS}^{*} indicates the number of certificates reporting *A* and *S* on different lines.

2.4 Association between conditions

This step serves as a confirmation of the previous one and it is aimed at verifying if two codes are associated, *i.e.* reported in part 1 more frequently than expected under the hypothesis that the two codes are reported independently from each other. Certificates are cross-tabulated by presence/absence of the codes under analysis (*A* and *S*) as shown in table 2.2. Observed and expected distributions are compared by a chi-square test (double tailed with 1 degree of freedom). For each possible pair of codes we define that there is association between the two conditions when the pair is mentioned significantly more than expected: the frequency of certificates reporting both the codes is higher than expected and the chi-square test highlights a statistically significant ($p<0.05$) association.

In the other possible cases, when either the pair is mentioned on certificates significantly ($p<0.05$) less than expected or the chi-square is not significant ($p>=0.05$), the two conditions are not considered associated.

The method was developed using the software SAS Studio (Release: 3.6 Enterprise edition).

Table 2.2 - Observed and expected frequencies of certificates for the analysis of association between conditions (a)

Observed frequencies

		Code *S* Mentioned	Not mentioned	Total
Code *A*	Mentioned	N_{AS}	$N_{A\bar{S}}$	N_A
	Not mentioned	$N_{\bar{A}S}$	$N_{\bar{A}\bar{S}}$	$N_{\bar{A}}$
	Total	N_S	$N_{\bar{S}}$	N

Expected frequencies

		Code *S* Mentioned	Not mentioned	Total
Code *A*	Mentioned	$N_{AS}^{exp} = N_A * N_S/N$	$N_{A\bar{S}}^{exp} = N_A * N_{\bar{S}}/N$	N_A
	Not mentioned	$N_{\bar{A}S}^{exp} = N_{\bar{A}} * N_S/N$	$N_{\bar{A}\bar{S}}^{exp} = N_{\bar{A}} * N_{\bar{S}}/N$	$N_{\bar{A}}$
	Total	N_S	$N_{\bar{S}}$	N

(a) N indicates absolute frequencies.
Subscripts A and $\bar{A}$ indicate respectively the presence or the absence of the condition *A* on the death certificate.
Subscripts S and $\bar{S}$ indicate respectively the presence or the absence of the condition *S* on the death certificate.
N_{AS} indicates the number of certificates reporting both *A* and *S*, where $N_{AS}^* \leq N_{AS}$ used in table 1.2, since N_{AS}^* refers only to certificates that have the condition *A* and *S* in different lines of part 1, excluding cases where these two conditions are on the same line.

2.5 Definitions of precipitating and antecedent conditions, complications and causes

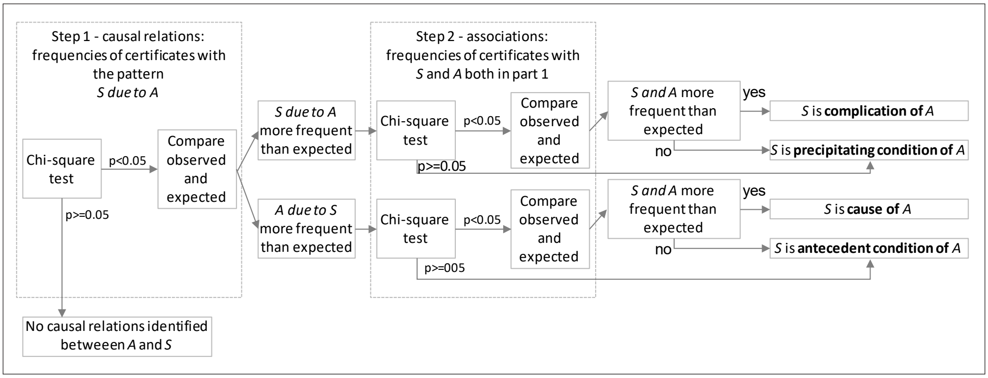
Based on the results of the first step, we define a condition *S* precipitating condition of *A* if *S due to A* is significantly more frequent than expected; we define *S* antecedent condition of *A* if *A due to S* is significantly more frequent than expected.

Based on the results of the second step, we define i) complication of *A* a precipitating condition *S* associated with *A* and ii) cause of *A* an antecedent condition *S* associated with *A*.

Conditions *S* not associated with *A* in the analysis of causal relations (first step), *i.e.* conditions that are not complications nor causes nor precipitating causes nor antecedent conditions of *A*, were not considered and therefore not shown in the results.

Figure 2.1 represents the process of the method.

Figure 2.2 - Representation of the method's process



2.6 Analysis performed on mortality data

We applied this method to data from the national CoD register, managed by the Italian National Institute of Statistics - Istat and referring to all deaths occurred in Italy among people aged 1 year and above. Provisional data of March-April 2020 were used in order to identify the most relevant complications and precipitating conditions (corresponding to the conditions *S* mentioned in the methods Section) of COVID-19 (ICD10 codes U071-U072, corresponding to the condition *A*). To show the potential general applicability of the method to other conditions, we applied it to Italian CoD data for the year 2018 by evaluating the causal relations of pneumonia (ICD10 codes J12-J18, condition *A*) and diabetes (ICD10 codes E10-E14, condition *A*) with other diseases (conditions *S*). Data were processed with Iris according to ICD10 provisions and the rejects of Iris (about 20% of total deaths) were revised by expert nosologists. Pneumonia and diabetes were chosen, as they are both common conditions on death certificates. On the other hand, they have different characteristics: pneumonia is a complication of many diseases, while diabetes is a disease whose onset does not generally result as a complication of other conditions. Moreover, these diseases are both related to COVID-19, but in different ways. Pneumonia is a common manifestation of COVID-19, while diabetes increase frailty of individuals and consequently the risk of dying for COVID-19.

We first carried out an explorative analysis using all ICD10 codes (third digit level). Successively, we grouped codes with similar behaviour in the explorative analysis, taking also into account the nosological similarity in ICD10. Groups studied are shown in table 2.3. Final analysis was performed on these groups instead of single ICD10 codes. In case of duplications of code-group on the same certificate, only the one on the lowest line of part 1 has been retained.

For the analysis of COVID-19 the method was applied on the whole sample of certificates, as well as by age groups to evaluate if the pattern of COVID-19 complications could depend on the age at death. We carried out the analysis on four age groups (1-49, 50-64, 65-79, 80 and more) but we did not observe differences between 0-49 year old and 50-64 and between 65-79 and 80 and more. For this reason, we focus only on two age groups: 1-64 years and 65 years and more.

Table 2.3 - Groups of diseases used for the analyses

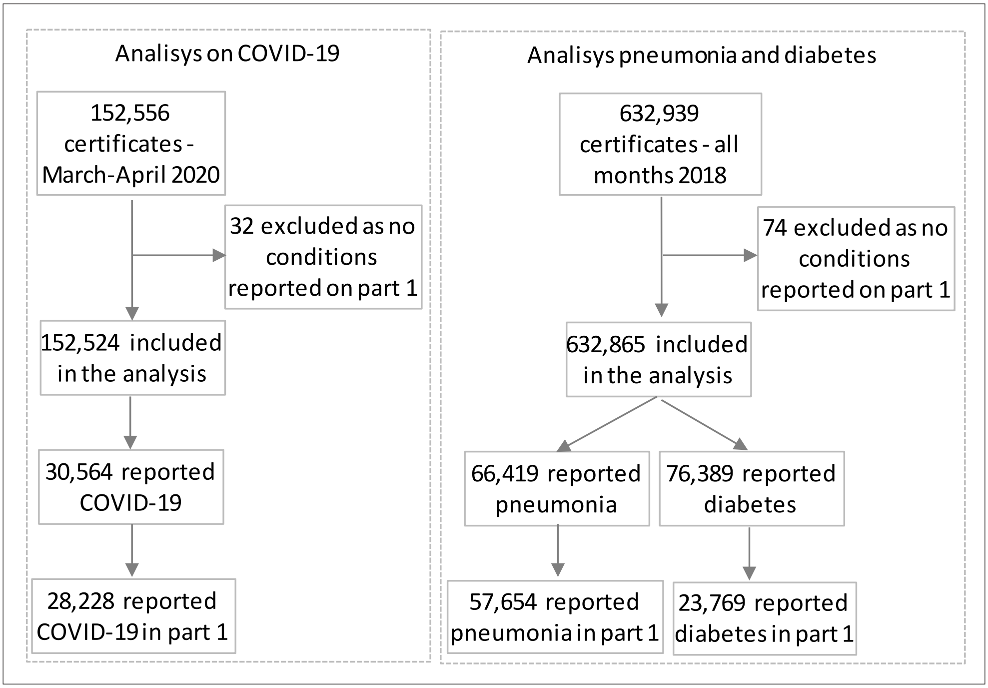
ICD10 codes	Description
A40-A41, B37, B49, B99	Sepsis and infections of unspecified site
Other codes in A00-B99	Other infectious and parasitic diseases
C00-D48	Neoplasms
D60-D64	Aplastic and other anaemias
D84-D87	Specified disorders of the immune mechanism
D89	Unspecified disorders of the immune mechanism
E10-E14	Diabetes mellitus
E86-E87	Volume depletion and other fluid disorders
Other codes in E00-E99	Other metabolic and endocrine diseases
F00-F03	Dementia
Other codes in F00-F99	Other mental and behavioural disorders
G00-G09, G93	Inflammatory d. of the central nervous system
G20, G22-G26	Extrapyramidal and movement disorders
G30-G31	Alzheimer disease
G61-G64	Other neuropathies
Other codes in G00-G99, H60-H99	Other disorders of nervous system
H00-H59	Diseases of the eye and adnexa
I10-I13	Hypertensive heart disease
I15	Secondary hypertension
I20-I24	Acute ischemic heart diseases
I25	Chronic ischemic heart disease
I26	Pulmonary embolism
I27-I28	Pulmonary heart disease and diseases
I48	Atrial fibrillation
I50-I51	Heart failure and other cardiac diseases
I00-I09, I30-I45, I47	Other heart diseases
I60-I66, I670, I672-I679	Cerebrovascular accident
I671, I69	Sequela of cerebrovascular diseases
I70	Atherosclerosis
I73-I79	Other peripheral vascular diseases
I00-I09, I90-I99	Other circulatory diseases
J12-J18, J849	Pneumonia (including interstitial pulmonary disease)
J40-J47	Chronic lower respiratory diseases
J80	Adult respiratory distress syndrome (ARDS)
J81	Pulmonary oedema
J960, J969	Respiratory failure
J961	Acute respiratory failure
J00-J11, J30-J39, J60-J70, J82-J848, J85-J99	Other diseases of the respiratory system
Other codes in K00-K99	Other diseases of the digestive system
Other codes in L00-L99	Other diseases of the skin
L89	Pressure ulcer
M00-M99	Diseases of the musculoskeletal s. and connective
N17, N19	Renal failure, acute and unspecified
N18	Chronic renal failure
Codes in N00-N29 (excluding N17-N19)	Other renal diseases
Q00-Q99	Congenital anomalies and chromosomal defects
R04-R09	Symptoms and signs involving the respiratory systems
R570-R571, R573-R579	Shock
R65	Systemic inflammatory response syndrome (SIRS)
U071-U072	COVID-19
V00-Y39, S00-T79	External causes

3. Results

At the time of this paper was drafted, 152,556 death certificates were collected and coded for March-April 2020, referring to deaths of people aged above 1 year. The analysis was performed on 152,524, since 32 had no conditions reported on part 1. COVID-19 was mentioned on 30,564 death certificates and in 28,228 (92.4%) was reported in part 1 (Figure 3.1).

Data used for pneumonia and diabetes, referring to all months of 2018, contained 632,939 death certificates and 632,865 had part 1 completed. Pneumonia was reported in 66,419 death certificates and was in part 1 in 57,654 cases (86.8%), while diabetes was found in 76,389 certificates of which in 23,769 in part 1 (31.1%) (Figure 3.1).

Figure 3.1 - Data analysed



3.1 COVID-19

Table 3.1 shows the results for COVID-19. The analysis allowed to distinguish conditions found associated in the first step of the analysis into complications, precipitating conditions, causes, and antecedent conditions of COVID-19.

Pneumonia, respiratory failure, respiratory symptoms, ARDS (adult respiratory distress syndrome), and SIRS (systemic inflammatory response syndrome) were complications of COVID-19. Overall, pneumonia was reported in part 1 in 40,580 certificates, of which 56% together with Covid-19. ARDS was together with COVID-19 more than 85% of the times it was mentioned (1,823 out of 2,133). Several other conditions, mainly non-respiratory, were precipitating conditions of COVID-19. Among these, the most frequent were heart failure, sepsis, shock, and renal failure.

Neoplasms, chronic lower respiratory diseases, cerebrovascular accident, hypertensive heart disease, and dementia resulted antecedent conditions of COVID-19. Nevertheless, none of them was cause of COVID-19, since none of them showed a significant association with it. Actually, all these conditions represent pre-existing diseases placed by certifiers in a lower line compared to COVID-19.

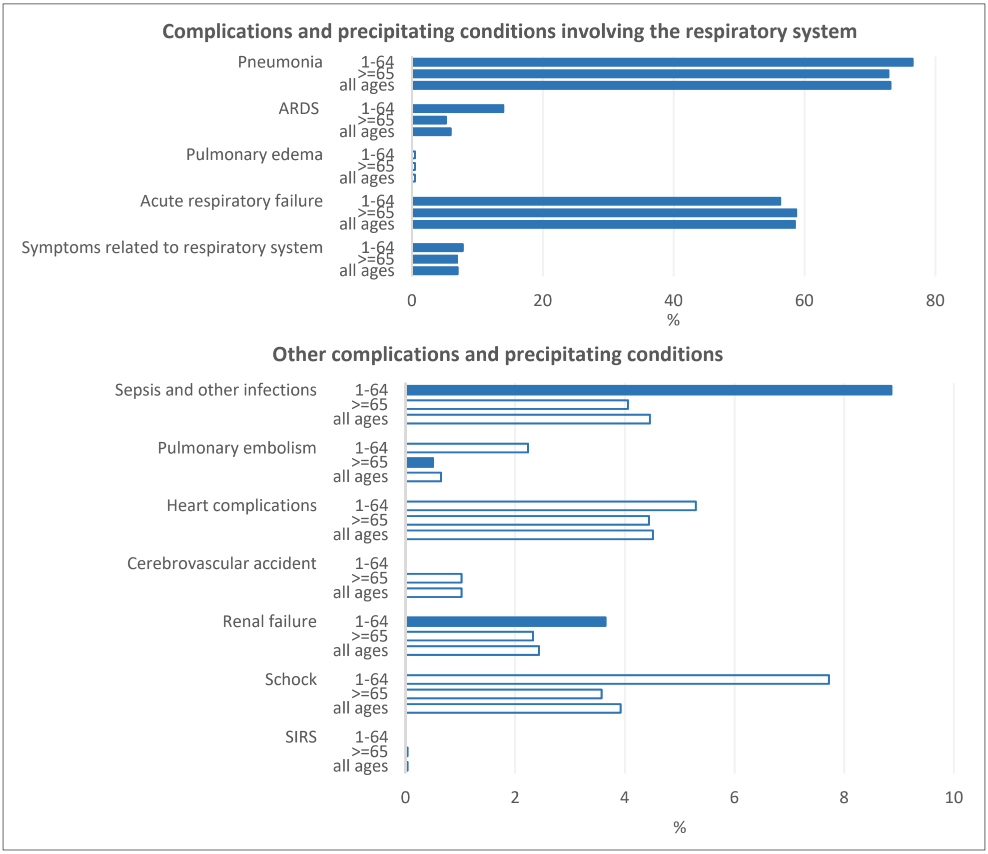
Figure 3.1 shows the results of the analysis by age group. Only complications and precipitating conditions of COVID-19 are shown, represented by blue and white bars respectively. Several respiratory conditions were found to be complications of COVID-19 in all age groups, except pulmonary oedema, which was a precipitating condition.

Non-respiratory conditions were mainly precipitating conditions of COVID-19, except for SIRS, which was a complication. However, we could highlight different patterns for the different age groups. Sepsis and renal failure were complications of COVID-19 for the age group 1-64 years, while they were precipitating conditions in the older age group. Pulmonary embolism was a complication of COVID-19for the age group 65+, while it was a precipitating condition in the younger age group.

The length of the bars in figure 3.2 represents, for each condition, the proportion of certificates reporting the condition due to or on the same line of

COVID-19 among certificates with mention of COVID-19, *i.e.* the proportion of cases where we cannot rule out that COVID-19 played a role in causing the condition. Overall, respiratory conditions were far more frequent than non-respiratory ones, with pneumonia reaching about 73% of COVID-19 related deaths and respiratory failure 58%. ARDS, overall reported in 6% of COVID-19 related deaths, was more frequent in 1-64 years old. Among non-respiratory conditions, some differences among age groups were observed. Sepsis, pulmonary embolism, heart complications, renal failure, and shock were more common among younger people. Cerebrovascular accidents and SIRS were observed only among older people.

Figure 3.2 - Complications (blue bars) and precipitating conditions (white bars) of COVID-19 by age group (a)



Source: Istat, National register of cause of death
(a) Precipitating conditions are represented only if the frequency is above 1%.

Table 3.1 - Conditions in a causal relation with COVID-19

Condition (a)	Certificates with both Covid-19 and *S* in part 1							Expected certificates with COVID-19 and *S*	Certificates with *S* (with or without COVID-19) in part 1
	COVID-19 *due to S*		*S due to* COVID-19		COVID-19 and *S* on the same line		Total (observed certificates with COVID-19 and *S*)		
S	N	%	N	%	N	%	N	N	N
Complication of COVID-19									
Pneumonia	659	2.9	9,348	40.6	13,003	56.5	23,010	7,510	40,580
Respiratory failure	767	4.1	17,395	93.2	496	2.7	18,658	7,771	41,989
Symptoms and signs involving the respiratory sys.	206	8.8	2,072	88.6	60	2.6	2,338	1,898	10,255
Adult respiratory distress syndrome (ARDS)	10	0.5	1,735	95.2	78	4.3	1,823	395	2,133
Systemic inflammatory response syndrome (SIRS)	0	0.0	10	90.9	1	9.1	11	5	26
Precipitating conditions of COVID-19									
Heart failure and other cardiac diseases	449	24.6	1,341	73.4	38	2.1	1,828	5,192	28,053
Sepsis and infections of unspecified site	147	9.7	1,303	86.3	59	3.9	1,509	1,846	9,977
Shock	29	2.4	1,189	96.9	9	0.7	1,227	2,093	11,308
Renal failure, acute and unspecified	144	16.2	739	83.2	5	0.6	888	1,581	8,544
Other diseases of the respiratory system	126	38.1	173	52.3	32	9.7	331	780	4,212
Volume depletion and other fluid disorders	42	16.6	201	79.4	10	4.0	253	489	2,641
Acute ischemic heart diseases	50	21.6	176	76.2	5	2.2	231	1,005	5,429
Pulmonary embolism	29	12.8	194	85.8	3	1.3	226	257	1,388
Other infectious and parasitic diseases	62	32.3	118	61.5	12	6.3	192	248	1,342
Other circulatory diseases	46	26.9	123	71.9	2	1.2	171	395	2,132
Pulmonary oedema	18	11.3	136	85.0	6	3.8	160	782	4,223
Antecedent conditions of COVID-19									
Neoplasms	578	88.8	62	9.5	11	1.7	651	5,312	28,703
Chronic lower respiratory diseases	472	82.2	78	13.6	24	4.2	574	1,218	6,580
Cerebrovascular accident	366	66.8	176	32.1	6	1.1	548	2,214	11,963
Hypertensive heart disease	407	85.9	64	13.5	3	0.6	474	2,627	14,197
Dementia	411	91.7	34	7.6	3	0.7	448	1,212	6,550
Chronic ischemic heart disease	356	81.1	80	18.2	3	0.7	439	2,008	10,851
Diabetes mellitus	271	77.0	78	22.2	3	0.9	352	1,074	5,805
Atrial fibrillation	186	59.6	123	39.4	3	1.0	312	1,134	6,126
Alzheimer disease	206	91.2	19	8.4	1	0.4	226	654	3,533
Chronic renal failure	135	63.4	76	35.7	2	0.9	213	702	3.791

Source: Istat, National register of cause of death
(a) Precipitating and antecedent conditions with less than 150 cases are not represented.

3.2 Pneumonia

Table 3.2 shows the results of the analysis on pneumonia. Conditions more frequently reported with pneumonia were respiratory failure (24,882 certificates), sepsis and infections of unspecified site (12,197 certificates), heart failure and other cardiac diseases (10,469 certificates). Generally, conditions were reported on the same line of pneumonia in a small percentage of cases; other diseases of the respiratory system were instead reported in the same line of pneumonia in 27.1% of certificates.

Complications of pneumonia are mainly diseases related to respiratory system (respiratory failure, symptoms and signs involving the respiratory systems, other diseases of the respiratory system, acute respiratory failure, pulmonary heart disease, ARDS), but include also sepsis and volume depletion and other fluid disorders. Precipitating conditions were mainly disease of circulatory system, but include also shock, renal failure, and pulmonary oedema. The following conditions were causes of pneumonia: chronic lower respiratory diseases, extrapyramidal and movement disorders, specified and unspecified disorders of the immune mechanism. Several conditions were found to be antecedent, mainly involving the circulatory system.

3.3 Diabetes

Table 3.3 shows the results for diabetes. Conditions more frequently reported with diabetes were hypertensive heart disease (6,829 certificates), heart failure and other cardiac diseases (5,545 certificates), chronic ischemic heart disease (4,924 certificates). Hypertensive heart disease and chronic lower respiratory diseases were frequently reported on the same line of diabetes (more than 31% of cases), followed by other metabolic and endocrine diseases, other mental and behavioural disorders, diseases of the eye and adnexa, and other neuropathies (between 21 and 26% of cases).

A large number of conditions were complications of diabetes, mainly diseases of circulatory system, followed by diseases of the genitourinary system. Other complications include dementia and other mental and behavioural disorders, volume depletion and other fluid disorders and other metabolic and endocrine diseases, diseases of the musculoskeletal system and connective tissue, pressure ulcer and other diseases of the skin, diseases of the eye and adnexa, and other neuropathies. The most frequent precipitating conditions were neoplasms and shock. Congenital anomalies and chromosomal defects were antecedent conditions of diabetes, while causes of diabetes were not found.

Table 3.2 - Conditions in a causal relation with pneumonia

Condition (a)	Certificates with both pneumonia and *S* in part 1							Expected certificates with pneumonia and *S*
	Pneumonia *due to S*		*S due to* pneumonia		Pneumonia and *S* on the same line		Total (observed certificates with pneumonia and *S*)	
S	N	%	N	%	N	%	N	N
Complications of pneumonia								
Respiratory failure	1,557	6.3	22,385	90.0	940	3.8	24,882	8,898
Sepsis and infections of unspecified site	836	6.9	10,860	89.0	501	4.1	12,197	5,008
Symptoms and signs involving the respiratory systems	343	6.4	4,769	89.6	209	3.9	5,321	4,020
Other diseases of the respiratory system	974	30.9	1,322	42.0	852	27.1	3,148	1,976
Acute respiratory failure	459	25.9	1,247	70.4	66	3.7	1,772	657
Volume depletion and other fluid disorders	238	14.7	1,289	79.6	93	5.7	1,620	1,321
Pulmonary heart disease and diseases	154	26.9	388	67.8	30	5.2	572	360
Adult respiratory distress syndrome (ARDS)	12	3.0	354	89.8	28	7.1	394	89
Precipitating conditions of pneumonia								
Heart failure and other cardiac diseases	2,415	23.1	7,263	69.4	791	7.6	10,469	11,563
Shock	89	2.1	4,156	97.4	20	0.5	4,265	5,495
Renal failure, acute and unspecified	566	17.3	2,541	77.8	158	4.8	3,265	3,835
Pulmonary oedema	130	7.4	1,547	88.0	81	4.6	1,758	2,296
Causes of pneumonia								
Chronic lower respiratory diseases	3,163	72.6	781	17.9	413	9.5	4,357	2,483
Extrapyramidal and movement disorders	789	92.4	46	5.4	19	2.2	854	725
Specified disorders of the immune mechanism	172	90.1	16	8.4	3	1.6	191	50
Unspecified disorders of the immune mechanism	11	91.7	1	8.3	0	0.0	12	7
Antecedent conditions of pneumonia								
Neoplasms	6,264	90.7	457	6.6	182	2.6	6,903	16,269
Cerebrovascular accident	2,599	82.7	474	15.1	70	2.2	3,143	5,526
Dementia	1,971	92.5	130	6.1	30	1.4	2,131	2,360
Chronic ischemic heart disease	1,423	75.1	373	19.7	98	5.2	1,894	4,782
Hypertensive heart disease	1,339	76.9	293	16.8	109	6.3	1,741	5,544
Other heart diseases	882	51.4	770	44.9	64	3.7	1,716	3,172
Atrial fibrillation	824	51.7	693	43.5	76	4.8	1,593	2,417
External causes	1,144	87.8	111	8.5	48	3.7	1,303	2,214
Other diseases of the digestive system	766	58.9	398	30.6	136	10.5	1,300	2,690
Alzheimer disease	1,097	92.9	68	5.8	16	1.4	1,181	1,371

Source: Istat, National register of cause of death
(a) Precipitating and antecedent conditions with less than 1,500 cases are not represented.

Table 3.3 - Conditions in a causal relation with diabetes

Condition (a)	Certificates with both diabetes and *S* in part 1							
	Diabetes *due to S*		*S due to* diabetes		Diabetes and *S* on the same line		Total (observed certificates with diabetes and *S*)	Expected certificates with diabetes and *S*
S	N	%	N	%	N	%	N	N
Complications of diabetes								
Hypertensive heart disease	2,102	30.8	2,592	38.0	2,135	31.3	6,829	2,286
Heart failure and other cardiac diseases	641	11.6	4,441	80.1	463	8.3	5,545	4,767
Chronic ischemic heart disease	1,149	23.3	3,196	64.9	579	11.8	4,924	1,972
Cerebrovascular accident	445	15.5	2,236	77.9	189	6.6	2,870	2,278
Renal failure, acute and unspecified	259	9.3	2,367	84.9	161	5.8	2,787	1,581
Chronic renal failure	255	11.1	1,705	74.5	330	14.4	2,290	666
Acute ischemic heart diseases	136	6.6	1,907	91.9	33	1.6	2,076	1,135
Sequela of cerebrovascular diseases	313	21.0	1,080	72.4	99	6.6	1,492	825
Chronic lower respiratory diseases	305	22.2	638	46.5	428	31.2	1,371	1,024
Atrial fibrillation	291	23.1	772	61.3	196	15.6	1,259	997
Other circulatory diseases	67	5.7	963	82.1	143	12.2	1,173	406
Other peripheral vascular diseases	65	6.1	917	85.6	89	8.3	1,071	243
Dementia	210	20.3	698	67.4	128	12.4	1,036	973
Other metabolic and endocrine diseases	78	10.9	476	66.6	161	22.5	715	244
Atherosclerosis	51	7.3	588	84.4	58	8.3	697	288
Volume depletion and other fluid disorders	79	13.0	486	80.1	42	6.9	607	545
Diseases of the musculoskeletal s. and connective	65	24.3	155	58.1	47	17.6	267	207
Other mental and behavioural disorders	88	33.2	120	45.3	57	21.5	265	193
Pressure ulcer	16	6.5	215	87.4	15	6.1	246	169
Other diseases of the skin	20	8.2	196	80.0	29	11.8	245	78
Other renal diseases	27	14.7	128	69.6	29	15.8	184	113
Diseases of the eye and adnexa	7	8.0	58	66.7	22	25.3	87	14
Other neuropathies	6	7.8	54	70.1	17	22.1	77	15
Secondary hypertension	7	11.7	46	76.7	7	11.7	60	11
Precipitating conditions of diabetes								
Neoplasms	823	37.9	1,087	50.0	262	12.1	2,172	6,707
Shock	58	2.8	2,044	97.0	5	0.2	2,107	2,266
Antecedent conditions of diabetes								
Congenital anomalies and chromosomal defects	12	70.6	3	17.6	2	11.8	17	50

Source: Istat, National register of cause of death
(a) Precipitating conditions with less than 1,500 cases are not represented.

4. Discussion and concluding remarks

The present analysis used the complete information provided by physicians on death certificates, widely considered the most reliable source of information to compare cause-specific mortality across populations. The described method, based on the simple application of chi-square tests, is able to identify causal relations between diseases leading to death. This is an important information for public health, since it provides data on the major complications to be tackled for specific diseases.

In the first step of the method, causal pattern are taken into account, while the second step confirms the results. This latter step verifies if there is an association between conditions within causal sequences, *i.e.* if the two conditions under exam are found in causal sequences more frequently than expected. If this association is not found, it cannot be concluded that one condition is a cause or a complication of the other even if a causal pattern is found in the first step.

With the introduction of automated coding systems, several countries have been able to produce statistics on all causes reported on certificates. This kind of data, referred to as multiple causes of death, have been used to recalculate mortality levels attributed to a given condition, and to determine the most frequent associations of causes involving such condition (Desesquelles *et al.*, 2012). From the analysis of multiple causes reported in the sequence of part 1, additional information can be withdrawn concerning pathways of diseases, for instance which are the most frequent complications of underlying conditions. Our study fits into this context. However, it investigates an aspect, that of causes and complications in the process leading to death, which has not been studied so far.

Results of the first step on COVID-19 deaths show that respiratory conditions are the most common complications leading to death. Non-respiratory conditions, for instance cardiac complications, are generally precipitating conditions of COVID-19 but they cannot be considered direct complications of it. This could mean that, although these conditions are reported due to COVID-19, the causal relation might be not direct, but mediated by other conditions associated to COVID-19. An example of this scenario is shown in example 3 of figure 1.2, where congestive heart failure is reported due to both COVID-19 and ischemic heath disease.

Some conditions are antecedent of COVID-19, but none of them is a cause of COVID-19, meaning that they are actually underlying conditions that increase the risk of death but cannot be seen as direct causes of COVID-19. Nevertheless, certifiers wrongly place them in a causal relation with COVID-19 as shown in example 3 of figure 1.2. The second step of the procedure seems to be essential for the interpretation of the results.

Results on antecedent conditions confirm the knowledge on risk factors for critically ill COVID-19 as well as the role of SIRS as a complication of this disease. Our results on the precipitating conditions also confirm the indirect role of cardiovascular complications described in literature and linked to the pre-existing health status of the patients. Nevertheless, some rare complications such as kidney injury, coagulation disorders, thromboembolism, and vascular inflammation appear in our results only as precipitating conditions (Gao *et al.*, 2021; Zheng *et al.*, 2020; Chang *et al.*, 2021). This could be due to the rarity of these complications but also to the lack of quality in cause-of-death reporting.

Different patterns of non-respiratory conditions by age were found, especially for sepsis, renal failure, and pulmonary embolism. The first two conditions are complications of COVID-19 in younger subjects (<65 years), while they are precipitating conditions in older (>=65 years). Pulmonary embolism is a complication of COVID-19 in older deaths and a precipitating condition in younger deaths. Since these conditions do not have a common pattern for all age groups, the presence of them as complications of COVID-19 might reflect the presence of different underlying conditions across the different age groups. These patterns should be further analysed.

The first step of this method was used for measuring the frequencies of complications of deaths of SARS-CoV-2 infected patients (Grippo *et al.*, 2020). Changes of these frequencies in different periods of the pandemic waves have also been analysed, showing an increase of non-respiratory complications in the transition phase (May-September 2020), when COVID-19 mortality was lower (Grippo *et al.*, 2021). Results of the current paper are consistent with those of previous ones. In the previous researches the second step of the analysis was not performed, therefore the method described here represents an improvement of the previous work.

In order to validate the method, we carried out the analysis on pneumonia, a common manifestation of COVID-19 and a complication of many diseases, and on diabetes, a completely unrelated disease whose onset does not generally result as a complication of other conditions. We chose to run the analysis on 2018 data since 2020 data would have been affected by the presence of COVID-19, especially for pneumonia.

Findings on pneumonia confirm the strength of the method. Complications found for pneumonia are very similar to those of COVID-19 but, differently from COVID-19, pneumonia is a complication of some conditions: chronic lower respiratory diseases, extrapyramidal and movement disorders, specified disorders of the immune mechanism and unspecified disorders of the immune mechanism. On the other hand, diabetes is not a complication of any disease. It has only one group of antecedent conditions: congenital malformations and chromosomal anomalies. Several conditions were found to be complications of diabetes, especially cardiovascular complications known to be the leading cause of morbidity and mortality in diabetic patients (Zheng *et al.*, 2018). The analysis on pneumonia and diabetes allows to highlight the potentialities of the method. Some conditions, for instance pneumonia and sepsis, can be due one to each other. Both the causal relations are plausible on a medical point of view. The method allows identifying the pattern of the causal relation on death certificates: in the context of deaths, sepsis is a complication of pneumonia, and not vice-versa.

In summary, the method succeeds in highlighting patterns between conditions and it is useful to document complications of diseases as well as causative conditions. Moreover, it has been proven helpful for describing characteristics of COVID-19 related deaths. For its simplicity, it can be applied to all settings where multiple causes of death are available providing comparable results, since cause-of-death data are highly comparable.

Nevertheless, the patterns identified by the method include only relations between two conditions at the time, without taking into account more complex relations involving multiple conditions. As a matter of fact, with the aging population, multimorbidity at death is common and complications of conditions directly leading to death may depend on the presence of different comorbidities. The method does not allow to take into account the presence of such comorbidities reported on the death certificate.

Our study highlights some differences in complications for different age groups, further analysis by smaller age groups and gender should be performed.

Moreover, the method presented does not take into account part 2 of the death certificate. It could be interesting to take into account also this part in order to better characterise antecedent and precipitating conditions, distinguishing conditions for which there is an association on the whole death certificate from those completely not associated, when the complete year of data will be available.

Some limits linked to the nature of data used should also be taken into account. The method relies on cause-of-death reporting and the quality of the information reported by the certifying physicians may vary greatly: both under and over-reporting may occur and only in rare cases the causes are confirmed by autopsy or other pathological findings (Anderson, 2011). However, findings from previous research support the accuracy of COVID-19 mortality surveillance using official death certificates (Gundlapalli *et al.*, 2021). Moreover, causal sequences are sometimes misreported by certifiers, who, instead of reporting conditions in causal order, report them by chronological or importance order. Nevertheless, the second step of the method provides additional hints on the association between conditions.

Overall, the results show that the proposed method is reliable and can be of great utility to understand the role of a condition in the morbid process and to reduce its lethality; strategies for preventing deaths should focus not only on the condition itself but also on the complications which lead to death.

References

Anderson, R.N. 2011. "Coding and classifying causes of death: "Trends and international differences". In Rogers, R.G., and E.M. Crimmins (*eds.*). *International Handbook of Adult Mortality. International Handbooks of Population*, Volume 2: 467-489. Dordrecht, The Netherlands: Springer.

Chang, W.-T., H.S. Toh, C.-T. Liao, and W.-L. Yu. 2021. "Cardiac Involvement of COVID-19: A Comprehensive Review". *The American Journal of the Medical Sciences*. Volume 361, N.: 14-22.

Desesquelles, A.F., M.A. Salvatore, M. Pappagallo, L. Frova, M. Pace, F. Meslé, and V. Egidi. 2012. "Analysing Multiple Causes of Death: Which Methods For Which Data? An Application to the Cancer-Related Mortality in France and Italy". *European Journal of Population*, 28: 467-498.

Gao, Y.-D., M. Ding, X. Dong, J.-J. Zhang, A. Kursat Azkur, D. Azkur, H. Gan, Y.-L. Sun, W. Fu, W. Li, H.-L. Liang, Y.-Y. Cao, Q. Yan, C. Cao, H.-Y. Gao, M.-C. Brüggen, W. van de Veen 2, M. Sokolowska, M. Akdis, and C. A. Akdis. 2021. "Risk factors for severe and critically ill Covid-19 patients: A review". *Allergy. European Journal of Allergy and Clinical Immunology*, Volume 76, Issue 2: 428-455.

Grippo, F., E. Grande, A. Maraschini, S. Navarra, M. Pappagallo, S. Marchetti, R. Crialesi, L. Frova, C. Orsi, S. Simeoni, A. Carinci, G. Loreto, C. Donfrancesco, C. Lo Noce, L. Palmieri, X. Andrianou, A.M. Urdiales, G. Onder, G. Minelli, and The Italian National Institute of Health COVID-19 Mortality Group. 2021. "Evolution of Pathology Patterns in Persons Who Died From COVID-19 in Italy: A National Study Based on Death Certificates". *Frontiers in Medicine* 8: 645543.

Grippo, F., S. Navarra, C. Orsi, V. Manno, E. Grande, R. Crialesi, L. Frova, S. Marchetti, M. Pappagallo, S. Simeoni, L. Di Pasquale, A. Carinci, C. Donfrancesco, C. Lo Noce, L. Palmieri, G. Onder, G. Minelli, and The Italian National Institute of Health COVID-19 Mortality Group. 2020. "The Role of COVID-19 in the Death of SARS-CoV-2-Positive Patients: A Study Based on Death Certificates". *Journal of Clinical Medicine - JCM*, Volume 9, Issue 11: 3459.

Gundlapalli, A.V., A.M. Lavery, T.K. Boehmer, M.J. Beach, H.T. Walke, P.D. Sutton, and R.N. Anderson. 2021. "Death Certificate-Based ICD-10 Diagnosis Codes for COVID-19 Mortality Surveillance - United States, January-December 2020". Centers for Disease Control and Prevention - CDC, *Morbidity and Mortality Weekly Report – MMWR*, Volume 70(14): 523-527.

Office for National Statistics - ONS. 2020. *Measuring pre-existing health conditions in death certification - deaths involving covid-19: March 2020. A method for deciding which pre-existing condition mentioned on death certificates is the main pre-existing condition*. UK: ONS.

Orsi, C., D. De Rocchi, M. Popescu, F. Heuser, S. Weber, L. Frova, V. Della Mea, and F. Grippo. 2020. "Implementing ICD-11 for mortality: translation of decision tables embedded in the automated coding system Iris". *Rivista di statistica ufficiale*, N. 1/2020. Roma: Istat.

Palmieri, L., N. Vanacore, C. Donfrancesco, C. Lo Noce, M. Canevelli, O. Punzo, V. Raparelli, P. Pezzotti, F. Riccardo, A. Bella, M. Fabiani, F.P. D'Ancona, L. Vaianella, D. Tiple, E. Colaizzo, K. Palmer, G. Rezza, A. Piccioli, S. Brusaferro, G. Onder, and The Italian National Institute of Health COVID-19 Mortality Group. 2020. "Clinical characteristics of hospitalized individuals dying with COVID-19 by age group in Italy". *The Journals of Gerontology, Series A*, Volume 75, Issue 9: 1796–1800.

World Health Organization - WHO. 2019. *International Statistical Classification of Diseases and Related Health Problems. 10th Revision. 2019 Edition*. Geneva, Switzerland: WHO. https://icd.who.int/browse10/2019/en#/

Zheng, Z., F. Peng, B. Xu, J. Zhao, H. Liu, J. Peng, Q. Li, C. Jiang, Y. Zhou, S. Liu, C. Ye, P. Zhang, Y. Xing, H. Guo, and W. Tang. 2020. "Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis". *Journal of Infection*, Volume 81, Issue 2: e16 - e25.

Zheng, Y., S.H. Ley, and F.B. Hu. 2018. "Global aetiology and epidemiology of type 2 diabetes mellitus and its complications". *Nature Reviews Endocrinology*, Volume 14, Issue 2: 88-98.

Integration of agritourism farms' microdata: economic analysis and impact assessment of the COVID-19 effects

Roberto Gismondi, Maria Grazia Magliocchi, Filippo Oropallo, Francesco Giovanni Truglia [1]

Abstract

The paper presents the results of the integration of different data, collected by statistical surveys and administrative sources, in order to estimate the economic results of the Italian agricultural holdings and agritourism farms. It also proposes an assessment of the consequences of COVID-19 on farmhouses through a micro-simulation, based on three hypothetical scenarios for 2020, including both the most recent economic sectoral trends and the survival probability of the agritourism farms. According to the analysis, reduction in turnover could range between -14.3% and -23.4%, while losses of value added should range between -16.9% and -27.8%. Micro-founded analysis, coupled with macro trends, allowed the evaluation of the extent of the crisis in the agritourism sector and the consequences both at territorial level and for different types of businesses. The incomes reduction is stronger for small farms in the South and Islands area of Italy, while the biggest ones show lower losses due to the diverse and larger activities.

Keywords: Micro-integration, agritourism farms, economic indicators, microsimulation.

1 Roberto Gismondi (gismondi@istat.it); Maria Grazia Magliocchi (magliocchi@istat.it); Filippo Oropallo (oropallo@istat.it); Francesco Giovanni Truglia (truglia@istat.it), Italian National Institute of Statistics – Istat.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction[2]

From the beginning of the Coronavirus pandemic, more than 2.7 million people have died. In order to contain this health emergency, at the beginning of 2020 the social and economic interconnections were affected by a sudden stop. Consequently, as affirmed by the United Nations World Tourism Organization (UNWTO) Secretary General Zurab Pololikashvili *"[...] tourism has been hit hard, with millions of jobs at risk in one of the most labour-intensive sectors of the economy"*.

The Istat survey "*Situation and prospects of enterprises in the health emergency* COVID-19" provided a first empirical evidence of the consequences of the pandemic. This survey[3] had the purpose of collecting assessments directly from enterprises about the effects of the health emergency and the economic crisis on their business. The survey attests that during phase 1 of the health emergency (between 9 March and 4 May 2020), 45.0% of units with 3 and more employees (458 thousand, which cover 27.5% of employees and 18.0% of turnover) suspended operations. At sectoral level, considering construction and service the majority of the enterprises suspended their activities: 58.9% and 53.3% respectively, compared to 36.0% of industry in the strict sense and 30.3% of trade. Moreover, 38.0% (with 27.1% of employees) reported operational and sustainability risks of their business and 42.8% requested liquidity and credit support (DL 18/2020 and DL 23/2020). More than half of the enterprises (37.8% employed) have forecasted a lack of liquidity to meet the expenses planned until the end of 2020. Over 70% of units (representing 73.7% of employment) reported a reduction in turnover in the two-month period of March-April 2020 compared to the same period of 2019.

2 The assessments expressed in this article are based on the situation referring to the end of March 2021 and on data updated to December 2020. Moreover, they do not consider further government measures that could limit the accommodation activity in 2021.

3 Istat, 2020*f*. The target population is consistent with that defined in the permanent census of enterprises: active enterprises operating in industry, commerce and services have been considered according to the classification of economic activities "ATECO 2007": sections "B" to "N" and "P" to "R", divisions "S95" to "S96". Enterprises with at least 2.5 average employees in the year were included and enterprises with an average number of employees of 99.5 or more were included as a census. The overall sample, interviewed with self-compilation CAWI, included 90,461 units, of which 46.9% have provided answers.

On this point, the Survey on "*Agritourism season 2020 – COVID-19 impact*" conducted under the Programme of the National Rural Network 2014-2020, has underlined how the 86% of the agritourism farms has suffered a loss of revenues (ISMEA, 2020*b*; 2021).

Starting from this context, our analysis aims at estimating the impact of the lockdown on the economic performance of the agritourism farms (farmhouses), on which the domestic and international demand collapse had a deep impact. To this purpose, we used an integrated database that gathers different sources to estimate the economic results of agricultural enterprises and agritourism farms with a bottom-up approach. This work is organised in four sections. The next one describes the whole population of agritourism farms (AFs) using data from the Istat census survey[4] and the extended farm register[5] (FR2) for the estimation of economic results at unit level. The extended farm register integrates the basic farm register[6] (FR) with economic data derived from administrative and statistical sources such as social security contribution and tax returns data, financial statements and foreign trade data. This allows to estimate the values of the income statement and to build a pre-pandemic scenario at a micro level.

The central part (third and fourth paragraph) refers to the resume of main analyses and evaluations about the effects of the pandemic situation on AFs and to the first simulation of possible scenarios, based on electronic invoicing data at sectoral level and tourism statistics.

The final part (fifth paragraph) reports the microsimulation analysis. Three alternative scenarios have been defined, based on different hypotheses as regards the collapse of demand for accommodation and catering services. The main purpose was the estimation of a range of variation of the revenues at micro level, given the knowledge of sector trends based on data from electronic invoicing, tourism statistics and the distribution of the farms survival rate by province.

4 Istat, 2019*b*. Istat has been carrying out the survey on agritourism farms since 2007. It is an annual census survey which detects the main structural characteristics and the kind of activities carried out by AFs. Regions and Autonomous Provinces collect data.

5 Oropallo, 2021.

6 Istat, 2019*c*.

2. Agritourism farms in Italy: a pre-pandemic snapshot

The capacity of farmhouses to combine innovation and tradition (Palmi and Lezzi, 2020) could be one the reasons of the significant signs of growth shown in the recent years. Between 2007 and 2018, in fact, the number of AFs increased by 33%, while in the same period the value added (VA) of AFs increased from 1.08 to 1.46 billion euros7. Between 2011 and 2018, farmhouses recorded an increase of +15.7% (equal, in absolute terms, to 3,202 units), closing 2018 with 23,615 AFs authorised to carry out agritourism activities (Table 1 and Figure 1). Among the regions with the highest growth, Puglia (+139.3%) is characterised by a high variability around this trend (the Coefficient of Variation – C.V.[8] is 33%).

Concerning the southern regions and the Islands (10.5%), different trends characterise Abruzzo (-22.6%), Campania (-15.2%), Calabria (-3.3%) and Sardegna (-3.3%). In 2018, Toscana alone accounted for almost 20% of the total of national AFs, followed by 15.5% in Trentino-Alto Adige.

It is interesting to note that a greater diffusion of AFs is associated to longer average life (Figure 1); it happens in Toscana, Emilia-Romagna and Trentino-Alto Adige. In other regions, the picture is more heterogeneous and partly affected by administrative changes, as in the case of Sardegna.

The spatial-temporal configuration of the AFs highlights two macro-areas with high longevity, described by an average age of 15 years (the national average age is 10). This survival analysis will be also integrated in the microsimulation analysis to calibrate the risk of market exit. The first macro-area extends from the eastern side of Liguria to the south-west border of Umbria and includes all Toscana. The second one covers the regions of the Northeast. These two areas are joined by a "corridor" given by the municipalities that cross Emilia-Romagna from the north (Mirandola) to the south (Camugnano), which gives territorial continuity to these two 'high longevity' macro-areas.

7 The estimation is based on National Accounts of agricultural sector (Istat, 2020*b*).
8 The Coefficient of Variation is the ratio between the standard deviation of longevity and the average life.

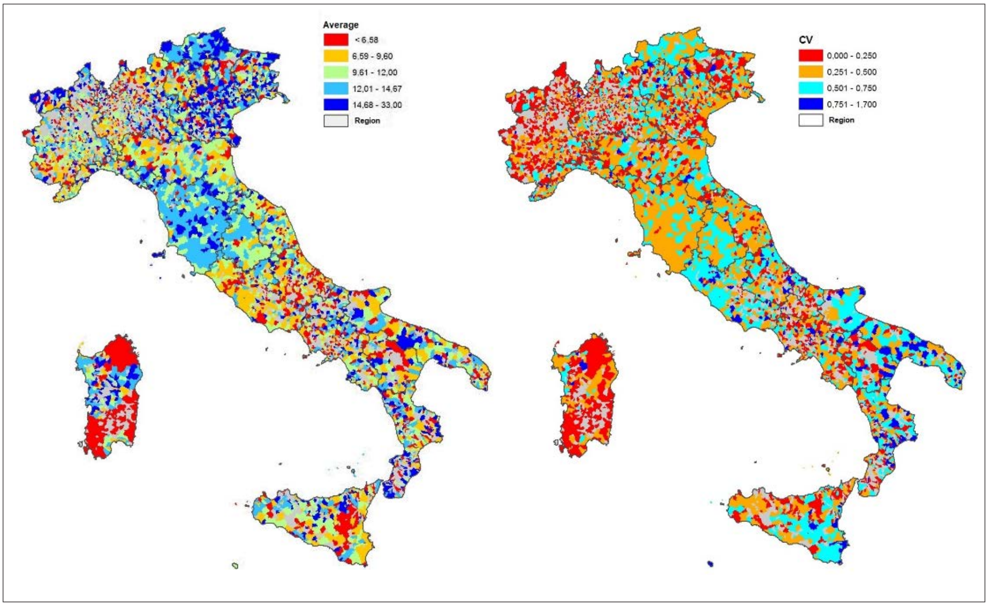
Table 1 - Active agritourism farms by region and geographical division. Years 2011-2018 (absolute values, % change 2011-2018 and Coefficient of Variation)

REGIONS AND GEOGRAPHICAL AREAS	2011	2012	2013	2014	2015	2016	2017	2018	% change	C.V.
North-west	*3,001*	*3,176*	*3,361*	*3,481*	*3,576*	*3,596*	*3,656*	*3,705*	*23.5*	*0.07*
Liguria	478	543	567	588	624	621	652	656	37.2	0.10
Lombardia	1,361	1,415	1,521	1,565	1,588	1,614	1,637	1,673	22.9	0.07
Piemonte	1,110	1,164	1,220	1,271	1,305	1,300	1,305	1,316	18.6	0.06
Valle d'Aosta/*Vallée d'Aoste*	52	54	53	57	59	61	62	60	15.4	0.06
North-east	*6,300*	*6,391*	*6,675*	*6,794*	*6,870*	*6,877*	*6,904*	*6,940*	*10.2*	*0.03*
Emilia-Romagna	1,030	1,036	1,106	1,133	1,187	1,156	1,167	1,166	13.2	0.05
Friuli-Venezia Giulia	566	588	614	632	643	656	661	670	18.4	0.05
Trentino-Alto Adige/*Südtirol*	3,366	3,391	3,506	3,570	3,550	3,581	3,651	3,648	8.4	0.03
Veneto	1,338	1,376	1,449	1,459	1,490	1,484	1,425	1,456	8.8	0.03
Centre	*6,935*	*7,076*	*7,152*	*7,274*	*7,642*	*7,777*	*8,264*	*8,382*	*20.9*	*0.07*
Lazio	811	841	884	940	950	947	1,253	1,278	57.6	0.17
Marche	786	788	880	1,005	1,030	1,060	1,070	1,082	37.7	0.12
Toscana	4,125	4,185	4,108	4,052	4,391	4,518	4,568	4,620	12.0	0.05
Umbria	1,213	1,262	1,280	1,277	1,271	1,252	1,373	1,402	15.6	0.05
South	*2,760*	*2,395*	*2,257*	*2,731*	*2,651*	*2,858*	*2,917*	*3,050*	*10.5*	*0.09*
Abruzzo	730	774	653	790	601	575	575	565	-22.6	0.13
Basilicata	131	145	112	131	135	162	180	187	42.7	0.17
Calabria	609	610	577	544	521	605	608	589	-3.3	0.05
Campania	831	407	458	523	572	648	677	705	-15.2	0.22
Molise	93	104	104	105	135	136	125	128	37.6	0.13
Puglia	366	355	353	638	687	732	752	876	139.3	0.33
Islands	*1,417*	*1,436*	*1,452*	*1,464*	*1,499*	*1,553*	*1,665*	*1,538*	*8.5*	*0.05*
Sardegna	828	834	819	799	794	794	807	801	-3.3	0.02
Sicilia	589	602	633	665	705	759	858	737	25.1	0.12
Italy	*20,413*	*20,474*	*20,897*	*21,744*	*22,238*	*22,661*	*23,406*	*23,615*	*15.7*	*0.05*

Source: Processing based on Istat data - Survey on agritourism farms

In most of the municipalities of the first geographical area (North-west), the C.V. varies between 0.06 and 0.10, showing a low variability of the AFs' life years within the municipalities. The values of this statistical indicator for the second macro-area (North-east) are even lower. In general, the AFs of these two areas are not only longer-lived, but this longevity is not due to the presence of outliers. On the contrary, the municipalities with average values of less than 10 years also have a higher C.V. and, therefore, show greater volatility of the permanence of these structures on the market.

Figure 1 - Mean and C.V. of the number of years of agritourism farms' activity at the municipal level (Years 2011- 2018)



Source: Processing based on Istat data - Survey on agritourism farms

2.1 Economic results of the agritourism farms

The estimation of economic results in the context of the extended farm register (FR2) made it possible to obtain a measure of the economic dimension of the agritourism sector with a bottom-up approach. The FR2 expands the information content of the basic farm register[9] through the integration of administrative and statistical sources. The basic register, available since 2014, includes the structural information of the farms such as the unit type, economic activity and technical-economic orientation (OTE) with standard production, the main crops (in terms of utilised agricultural area), livestock, the size of the farms and the location. The additional variables of the FR2 involve inputs of self-employed agricultural work, employees and their characteristics, labour costs, income statement variables such as sales and other revenues, changes in stocks, value of production, cost of goods and services, leasing cost, other charges. These variables allow us to compute

9 Istat, 2019*c*.

the VA and the gross operating margin of the agricultural activity for each farm. Other information is structure cost, investments, import-export at a detailed level, belonging to groups and level of control. For this purpose, the administrative and statistical sources have been integrated with the farm register (FR) data for each year:

1) National Social Security Institute (INPS) declarations relating to self-employed agricultural workers (AUTAGR) and agricultural labour (DMAG).
2) Tax return declarations (sole proprietorships, partnerships, corporates and non-commercial entities) and VAT returns.
3) Financial statements of corporates.
4) Istat Foreign Trade Data (Coe).
5) Structural business data (SBS Frame)[10];
6) Data from the employment register and the register of business groups.

The matching of the units was carried out using the identifying code of the integrated system of microdata (SIM), which translates the tax code of the production unit or of the natural person into a unique anonymous code. The first results of integration with data from administrative sources confirm the presence of a significant portion of units without tax obligations whose economic size is minimal (about two of three units). In these cases, they are small farms that employ less than half a person-year and derive from agricultural production less than 7 thousand euros a year, which represents the threshold for exemption from tax obligations. To estimate the economic variables, the first step concerns the estimate of the self-employed and family work input based on the declarations of the 350 thousand farms from INPS-AUTAGR source. Other farms are not obliged to submit an INPS declaration relating to family labour. To this end, a logarithmic model that relates the observed independent workdays with the characteristics of the basic register has been estimated: utilised agricultural area (UAA), adult bovine units (UBA), specialisation (OTE), unit type (enterprise with farm, secondary activities of SBS enterprises, farms of public institutions and informal farms of individuals) and farms location. The estimated parameters applied to the

10 Luzi and Monducci, 2016; Istat, 2019*a*.

smaller farms made it possible to calculate the annual days of self-employed for the management of farm activities.

The analysis of the coverage with regard to all the sources used in the integration process allows the stratification of all farms based on both a dimensional criterion and the availability of economic data:

- Smaller units (below the threshold) with missing revenue data (equal to 66.7%). They are widespread, but predictable and not very influential from an economic point of view. They constitute a significant part of the whole population, with the possibility of estimation from a mass-imputation model, using all the auxiliary variables of the FR and considering the behaviour of farm near to the revenue threshold.
- Micro and small-medium units: farms with 2-99 employees above the exclusion threshold, with percentages of valid data in the income statement of approximately 97%. These constitute the less problematic subset with little missing data.
- Large farms with 100 employees or more with complex organisation. For these units it is important to consider the type and legal status, in order to identify farms belonging to public institutions (regions and other public bodies). Out of 139 large farms, 23 are public and private institutions (equal to 16%); this share rises to 30% for farms with 500 employees or more.

As far as data are complete, the economic value of agricultural activity was calculated using as a proxy the share of revenues from agricultural activities on the total turnover from VAT returns. The imputation procedure is the multiple hot-deck technique with stratified selection of donors (Kim, J. K., and Shao, J., 2014).

Among the 24 thousand AFs, 51% belongs to the subset of businesses with prevalent agricultural activity, 26% is associated with SBS enterprises (with prevalent activities in industry or services) and the remaining 23% concerns less structured farms. Total annual jobs of AFs are about 57 thousand, with 26,404 employees. The annual turnover reaches 3.2 billion euros in 2018. The VA of these units is about 1.6 billion and the remuneration of the labour factor is about 641 million.

The Table 2 shows the breakdown of farms by size and the main variables of the income statement together with the export values. More than two thirds of the farms employ one people-year (class 1), while 26% of the farmhouses employ between 2 and 9 workers. Small-size units (between 10 and 19 employees) are 442 (1.9%), while medium-large enterprises with at least 20 employees are 185 (0.8%). For this segment there are the highest values of turnover (981 million euros) and of employees (7,774) and these are units with more articulated organisation and with several kind of activities connected with the agritourist one; they also record high values of exports 410 million (approximately 32% of revenues). Among the micro and small farms, those with 3-4 workers employ the most part of employees and show the highest value added.

Table 2 - Agritourism farms and main economic variables by class of workers. Year 2018

CLASS OF WORKERS	Farms	Total jobs	Employees	Turnover (mil€)	Value added (mil€)	Labour cost (mil€)	Gross operating margin (mil€)	Export (mil€)
1	12,986	12,133	526	379.6	211.6	10.1	201.5	3.1
2	4,661	9,188	2,200	339.3	209.5	44.6	164.9	4.9
3-4	4,345	15,300	6,454	679.3	414.2	136.1	278.1	18.9
5-9	996	6,846	4,818	405.5	217.1	108.5	108.6	32.1
10-19	442	5,460	4,631	400.5	185.8	111.3	74.4	38.4
20 and more	185	8,067	7,774	981.3	398.8	230.3	168.5	312.4
Total	*23,615*	*56,994*	*26,404*	*3,185.5*	*1,637.1*	*640.9*	*996.2*	*409.8*

Source: Processing based on both Istat and administrative data - Extended Farm Register (FR2)

In Toscana, about 13 thousand workers are active in AFs, which produce a turnover of about 944 million and about 456 million of VA (Table 3). Bolzano/*Bozen* and Lombardia follow in the ranking. In the South, Puglia produces the highest VA. The economic results in 2018 represent the basis of the microsimulation analysis to evaluate the impact of pandemic made in the fifth paragraph.

Table 3 - Agritourism farms and main economic variables by Region. Year 2018

REGIONS	Farms	Total jobs	Employees	Turnover (mil€)	Value added (mil€)	Labour cost (mil€)	Gross operating margin (mil€)	Export (mil€)
Piemonte	1,316	3,067	1,298	176.7	95.8	30.3	65.5	24.2
Valle d'Aosta/ *Vallée d'Aoste*	60	107	28	2.9	1.3	0.6	0.7	0.0
Lombardia	1,673	5,242	2,679	412.5	199.2	71.1	128.1	13.6
Bolzano/*Bozen*	3,185	6,564	959	309.1	230.4	19.7	210.7	1.6
Trento	463	1,319	578	77.9	40.4	14.1	26.4	0.8
Veneto	1,456	3,933	1,684	270.9	139.5	38.4	101.1	11.7
Friuli-Venezia Giulia	670	2,066	1,063	104.2	61.5	24.8	36.7	10.0
Liguria	656	1,045	303	45.1	24.6	6.7	17.9	1.6
Emilia-Romagna	1,166	3,764	2,038	228.8	114.1	50.8	63.3	6.7
Toscana	4,620	13,058	7,690	944.0	445.9	222.4	223.5	310.2
Umbria	1,402	2,586	1,077	104.8	42.6	25.1	17.5	3.7
Marche	1,082	1,854	638	63.0	28.9	14.1	14.8	7.9
Lazio	1,278	2,532	1,227	113.5	56.3	26.7	29.7	5.3
Abruzzo	565	845	249	24.9	12.2	4.6	7.7	1.8
Molise	128	228	99	11.1	2.5	1.8	0.8	0.0
Campania	705	1,248	577	36.3	15.9	9.8	6.1	0.6
Puglia	876	2,599	1,728	103.0	47.9	32.2	15.6	1.4
Basilicata	187	344	158	8.9	3.9	2.4	1.5	4.2
Calabria	589	1,342	795	26.4	11.4	13.7	-2.3	0.4
Sicilia	737	1,726	976	67.0	32.8	19.4	13.4	3.6
Sardegna	801	1,525	561	54.6	29.8	12.3	17.6	0.4
Italy	*23,615*	*56,994*	*26,404*	*3,185.5*	*1,637.1*	*640.9*	*996.2*	*409.8*

Source: Processing based on both Istat and administrative data - Extended Farm Register (FR2)

From the point of view of performance indicators, we observe a nominal labour productivity increasing with the size of the enterprises and increasing profitability[11] by size (Table 4). The ratio between average profitability and VA is 15.5%. An average share of exported turnover is also estimated at 12.9%, which rises to 31.8% for farms with 20 employees or more. Nominal labour productivity is the highest in Lombardia, Bolzano/*Bozen*, Veneto and Toscana (Table 5). The share of exports on turnover is the highest in Toscana (32.9%) and Basilicata (47.4%). Profitability is negative for many farms in the South: Calabria (-103%) and Molise (-62%).

11 Gross profitability (r) is equal to the adjusted EBITDA (Gross Operating Margin net of the cost of independent labour) on the value added. The cost competitiveness indicator (Compet) is equal to the ratio of value added per capita to unit labour costs (ULC) and represents the value created by the business unit for 100 euros of labour costs. The gross profitability r = 1-1 / c can also be derived from it.

Table 4 - Economic performance indicators by class of workers. Year 2018

CLASS OF WORKERS	Average size	Turnover per worker (000€)	Value added per worker (000€)	Unit Labour Cost (000€)	Compet %	Profit %	Export share%
1	0.9	31.3	17.4	19.2	90.9	-10.0	0.8
2	2.0	36.9	22.8	20.3	112.5	11.1	1.4
3-4	3.5	44.4	27.1	21.1	128.4	22.1	2.8
5-9	6.9	59.2	31.7	22.5	140.9	29.0	7.9
10-19	12.4	73.3	34.0	24.0	141.5	29.3	9.6
20 and more	43.6	121.6	49.4	29.6	166.9	40.1	31.8
Total	*2.4*	*55.9*	*28.7*	*24.3*	*118.3*	*15.5*	*12.9*

Source: Processing based on Istat data and administrative data - Extended Farm Register (FR2)

Table 5 - Economic performance indicators by Region. Year 2018

REGIONS	Average size	Turnover per worker (000€)	Value added per worker (000€)	Unit Labour Cost (000€)	Compet %	Profit %	Export share%
Piemonte	2.3	57.6	31.2	23.4	133.8	25.3	13.7
Valle d'Aosta/ *Vallée d'Aoste*	1.8	26.5	12.2	21.8	55.9	-78.9	1.0
Lombardia	3.1	78.7	38.0	26.5	143.2	30.2	3.3
Bolzano/*Bozen*	2.1	47.1	35.1	20.5	170.9	41.5	0.5
Trento	2.8	59.1	30.6	24.3	126.1	20.7	1.0
Veneto	2.7	68.9	35.5	22.8	155.6	35.8	4.3
Friuli-Venezia Giulia	3.1	50.4	29.7	23.3	127.7	21.7	9.6
Liguria	1.6	43.2	23.6	22.1	106.7	6.3	3.6
Emilia-Romagna	3.2	60.8	30.3	24.9	121.5	17.7	2.9
Toscana	2.8	72.3	34.1	28.9	118.1	15.3	32.9
Umbria	1.8	40.5	16.5	23.3	70.6	-41.6	3.6
Marche	1.7	34.0	15.6	22.1	70.5	-41.9	12.5
Lazio	2.0	44.8	22.2	21.7	102.3	2.2	4.7
Abruzzo	1.5	29.4	14.5	18.3	79.0	-26.5	7.1
Molise	1.8	48.7	11.2	18.1	61.8	-61.7	0.4
Campania	1.8	29.1	12.8	17.0	74.9	-33.5	1.8
Puglia	3.0	39.6	18.4	18.7	98.7	-1.3	1.4
Basilicata	1.8	25.9	11.4	15.1	75.5	-32.4	47.4
Calabria	2.3	19.7	8.5	17.2	49.3	-102.9	1.4
Sicilia	2.3	38.8	19.0	19.9	95.5	-4.7	5.4
Sardegna	1.9	35.8	19.6	21.8	89.5	-11.7	0.6
Italy	*2.4*	*55.9*	*28.7*	*24.3*	*118.3*	*15.5*	*12.9*

Source: Processing based on both Istat and administrative data - Extended Farm Register (FR2)

3. The effects of the pandemic: legislative measures and some assessments

Farms with agritourism activities suffered more than others from the shocks of the pandemic. The study of Mastronardi and Giaccio (2011) highlights different performance between farms with and without farmhouses. In fact, the former have a sort of advantage on the social and environmental level, while the latter have better performance on the economic side. AFs have lower profitability than other agricultural holdings, due probably to the high incidence of farm costs, in particular labour costs. In fact, they are characterised by a greater use of labour, especially of an extra-family nature, irrespective of the size of the farm.

To tackle the pandemic, several regulatory measures have taken place since spring 2020.

In the so-called *Relaunch Decree* (*Decreto Rilancio*), it was planned to grant farmhouses an extraordinary contribution for each estimated failure of overnight stays of costumers, determined by the difference between the actual attendance of the period January - June 2019 and that of the same period of 2020. Missed attendance had to be quantified based on communications made to the competent Quaestors pursuant to public safety regulations. The budget for 2020 was 80 million euros and the draft version of the Decree provided that farms could be granted an advance payment.

In support of the process of containment of losses resulting from the pandemic, several regions, including Friuli-Venezia Giulia and Sicilia, prepared specific economic measures, aimed at farmhouses and educational farms. As a rule, all beneficiaries who fulfilled the eligibility conditions were in the condition to be eligible for funding: the holdings had to be active on the date of submission of the application for support, in order to be included in the register of enterprises of the Chamber of Commerce and in the lists of farmhouses or educational farms.

The further additional Decree (*Decreto Ristori*), published on the Official Gazette of 28 October 2020, provided for a non-repayable contribution for active entrepreneurs, with VAT number, on 25 October 2020 and who carried out one of the activities listed in the decree. Among these, the activities of agritourism (accommodation or catering) and other related to entertainment, sport and tourism. The contribution was subject to the condition that the revenues for April 2020 were lower than 2/3 of the revenues for April 2019[12] (this requirement was not necessary for those who started after 1 January 2019).

As for the procedure, there were two modalities: those who have already benefited from the contribution provided for by the previous Decree (*Decreto Rilancio*) would receive a direct contribution on their own bank account from the Revenue Agency; the other subjects, on the other hand, should have electronically submitted a specific request. The amount of the contribution would be calculated by applying a specific coefficient related to the ATECO activity code (for farmhouses with accommodation was 150%, for those with catering to 200%) to the amount due according to the rules of the *Decreto Rilancio*.

The decree also cancelled the second tax duty IMU 2020, due within 16 December 2020, with reference to the properties and related appliances in which they exercised the agritourism activities and rent rooms, bed and breakfast, holiday homes. To benefit from the facility, it was necessary that the owners of the real estate should also be directly managers of the activities carried out there.

For the private employers of the AFs with accommodation and catering, the terms relating to the payment of social security contributions, social security premiums and compulsory insurance premiums due for the month of November 2020 were suspended. Such payments were due without the application of penalties and interest, in a single instalment by 16 March 2021 or in tranches up to a maximum of four equal monthly tranches, with the payment of the first one within 16 March 2021.

The new anti-COVID decree (end of 2020), which stated further restrictions until 15 January 2021, has continued to limit the activities of AFs, contextualising them to the possibility of take-away and home delivery, with an estimated loss of overnight stays in AFs, during the Christmas and New Year period, equal to about 2 million. The take-away and home delivery activities had a positive outcome during the Christmas holidays, although

12 According to estimates based on the electronic invoicing data (introduced in section 5), the revenues for April 2020 were quite lower than 2/3 of the revenues for April 2019.

even until the weekend 9 and 10 January 2021 Italy remained in the "orange" zone, so with about 24 thousand Italian farmhouses still closed even at lunch.

The views expressed by the farm managers themselves represent one of the most up-to-date measurement of how the sector operators are metabolising the consequences of the pandemic. However, there are no quantitative measurements on actual losses that could have characterised the farmhouse sector in the year 2020 based on objective data and not only on estimates.

According to ISMEA (2020), one of the sectors that should pay a high price to the COVID-19 pandemic is the tourism sector. ISMEA estimates that the losses of the farmhouse sector reached one billion in 2020. After the record figure of 13.4 million overnight stays in AFs in 2018, ISMEA estimates predicted, for 2020, a total loss for the sector of about 970 million euros, equivalent to 65% of turnover, mainly derived from the collapse of international demand.

We should also consider the fall in domestic demand because of the lockdown, for which both the Easter period and the holiday bridges of 25 April and 1 May were characterised by low tourist demand, with effects estimated at a loss of about 200 million euros, equal to 40-50% of the annual quota deriving from Italian guests. In addition, there was a further lost income of about 70 million euros. This results from the cancellation of the visits to the 1,500 farms that are also educational farms (*i.e.* used by school and families). These visits are mainly concentrated in the months of April and May.

Subsequently, in July 2020, the Confagricoltura Study Center (Baccino, 2020) estimated a drop in turnover of more than two-thirds compared to 2019, with a 71% reduction in overnight stays. In details, it was estimated that the sector's turnover fell slightly below 1.5 billion euros (1,460 million), 62% higher than the 900 million euros (970 million, including educational activities) estimated in May by ISMEA. Therefore, the turnover of accommodation and catering in the farmhouse sector should have fallen just below 600 million euros (597 million), less than a third of the turnover of 2019. Possible further deterioration could have occurred in the autumn period because of the dreaded "second wave" of infections.

According to the CIA (Italian Farmers' Confederation)[13], the situation of inland areas was particularly problematic, fragile by nature, because they still lacked adequate physical and digital infrastructure and services. In this perspective, the Recovery Fund is a great opportunity to give the right space and projects aimed at promoting the interior areas, involving primarily the farmhouses.

In the last months of 2020, ISMEA carried out a special survey regarding the evaluation of the effects of the pandemic on the performance of Italian farmhouses (in addition, ISMEA launched a portal dedicated to farmhouses[14]). The first results showed that, based on the views expressed by the sample of farmhouses that participated in the survey[15], because of the pandemic:

- the revenues decreased for 85% of the farmhouses, resulting in growth in only 6% of cases;
- during the lockdown, 15% of the structures thought to stop the activity, 45% to limit the damages and to wait for the return to a situation of normality and 29% to relaunch the enterprise with new strategies;
- almost half of the structures (47%) said, however, that they saw positive prospects for the future, 21% uncertain and only 9% negative, confirming how the sector analysed can be able to adapt even to unforeseen and complex situations such as the pandemic, which is still ongoing.

Following the entry into force of the *Decreto Ristori*, the main trade associations, including CIA and COLDIRETTI, have underlined the difficult situation. They highlighted how it continued to be "*[...] Impossible to work for the 24 thousand Italian agritourism farms because of the mandatory stop in the "red" and "orange" Italian zones and the night limitations in the "yellow" areas [...] Aids do not solve much [...] because structural interventions are needed, starting with energy concessions for at least 80% of the costs*" (Amato, 2020).

13 Ranerelli, 2020.
14 www.agriturismoitalia.gov.it/it/homepage.
15 The average age of the holder of the farm that participated in the survey is 51 years. In half of the cases (49%) the management is family type (with employees). The 45% of the farmhouses has a utilised agricultural area between 6 and 20 hectares. In 78% of cases, the farms surveyed have been in business for at least 10 years.

4. A first assessment of the effects of the pandemic: three scenarios

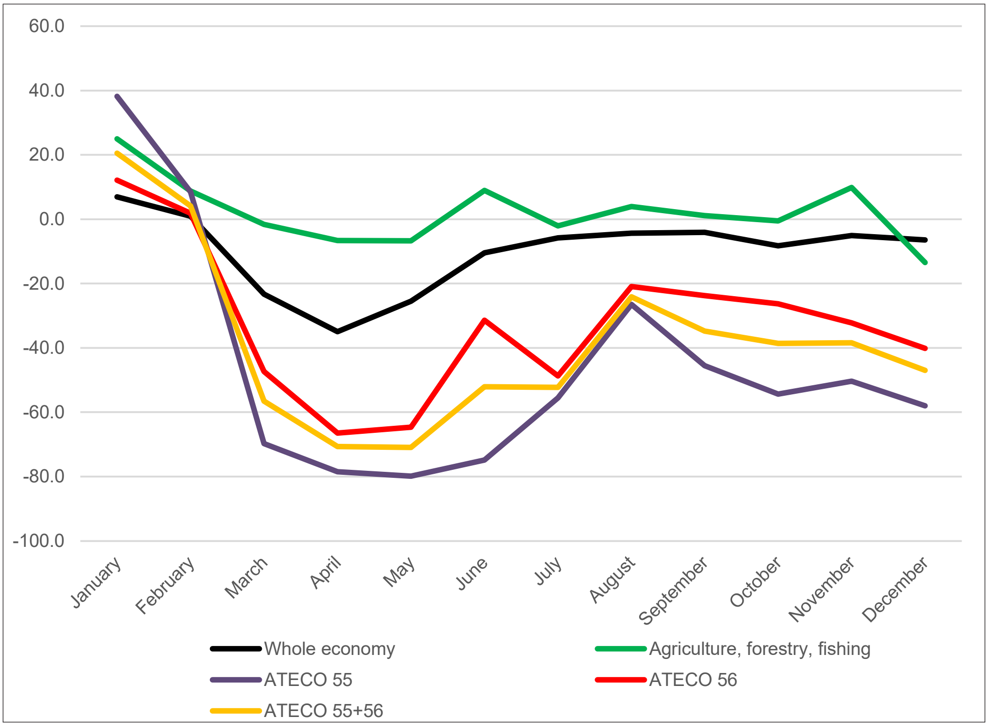
Since the beginning of 2020, Istat has been receiving from the Italian Tax and Revenue Authority (*Agenzia delle Entrate*), the data on electronic invoicing of Italian enterprises (excluding those participating in the flat-rate scheme). The variables, now available, are the number of invoices issued, the sum of incomes and the taxable amount. The data, with weekly and monthly periodicity, are available on a national scale after about 40 days from the end of the reference month. The available monthly data cover the whole 2020 and made it possible to estimate the monthly revenues of 2020 compared to 2019.

Although electronic invoicing data do not relate to the totality of transactions in the entire national economic system, they represent an objective measurement of the cyclical dynamics at a stage of complexity such as the lockdown period, and although to a smaller extent but still relevant, of the current period. Even though electronic invoicing data cannot be referred only to farmhouse activities, the available information allows timely monitoring of how individual economic sectors are reacting to the crisis or not, although with detail limited to the two-digit ATECO and therefore to the economic divisions.

The percentage change of taxable incomes between each month of 2020 and the corresponding month of 2019 has been calculated for the available period from January to December 2020.

The data were aggregated for the whole economy (sum of all economic activity divisions), for division 01+02+03 (agriculture, forestry and fisheries), 55 (accommodation activities), 56 (catering services) and for the set 55+56. A hypothesis underlying the estimates set out below is that the dynamics recorded under Divisions 55 and 56 may approximate the dynamics related to farmhouses, since the 2007 ATECO classification has codified with 55.20.52 "Farm-related housing activities" and "Farm-related catering activities" with 56.10.12. The main outcomes should be analysed considering that: 1) farmhouse activities can be carried out by entrepreneurial structures configured as agricultural holdings only; 2) the tourist and catering activities are those that characterise the main types of services offered by farms to customers.

Figure 2 - Percentage change in the taxable incomes recorded in the months of 2020 compared to the corresponding months of 2019



Source: Processing based on data by Istat and Agenzia delle Entrate

The monthly trends (Figure 2) show the surprising resilience of the primary sector, which in the lockdown period shows only slight declines compared to 2019. The Italian economic system shows the most marked difficulties, with the lowest peak of -34.9% reached in April 2020. In this framework, the accommodation and catering services activities are characterised by a quite worst behaviour, with strong turnover losses in the lockdown period (ranging from 40% to almost 80%), recovery during the Summertime (always characterised by losses with respect to 2019) and new worsening trends from September to December 2020.

Table 6 shows, for Divisions 56 (Catering) and Lodging plus Catering (55+56), the trend changes in the taxable incomes for the 12 months of 2020, which are the same as shown in Figure 2. If one assumes, as stated above, that the data in Division 56 are attributable to farmhouses with catering activities only and that those in Divisions 55+56 are attributable to farmhouses offering catering and accommodation, Table 6 shows the following outcomes:

- in the first 5 months of 2020, which include the lockdown phase, catering facilities have suffered a decline in revenues, compared to the same period of 2019, equal to 35.9%; this contraction (-39.9%) was more evident for facilities with catering and accommodation;
- in the only quarter characterised by lockdown, losses compared to 2019 were 59.5% for catering only and 66% for catering and accommodation;
- in 2020, the declines compared to 2019 were equal to 33.9% and 40.9% respectively.

Table 6 - Percentage change in the taxable incomes recorded in the months of 2020 compared to the corresponding months of 2019

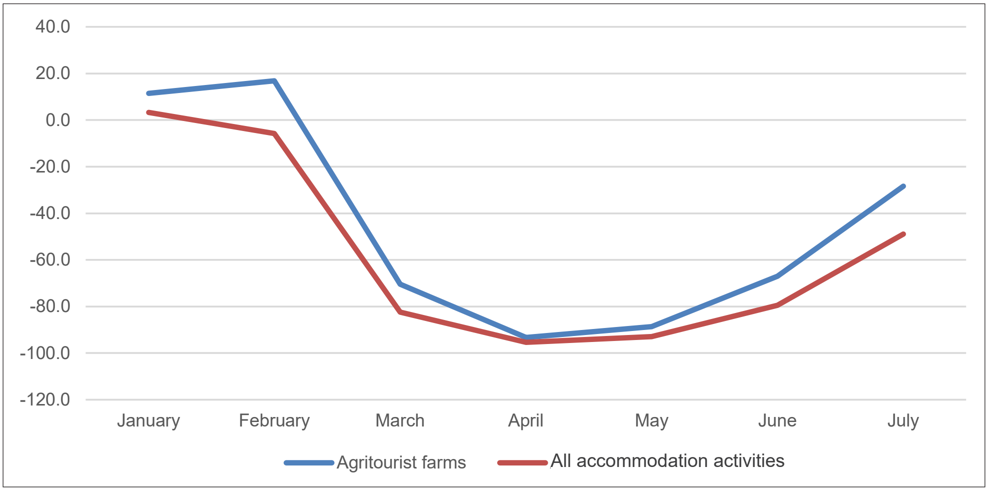
MONTH	% change with respect to the same month of 2019		
	ATECO 55: LODGING	ATECO 56: CATERING	ATECO 55+56: LODGING AND CATERING
January	38.2	12.1	20.5
February	8.5	1.9	4.2
March	-69.8	-47.3	-56.5
April	-78.5	-66.5	-70.7
May	-79.8	-64.7	-70.9
June	-74.9	-31.4	-52.1
July	-55.5	-48.7	-52.3
August	-26.5	-20.9	-24.1
September	-45.5	-23.7	-34.8
October	-54.3	-26.3	-38.6
November	-50.4	-32.2	-38.4
December	-58.0	-40.2	-47.0
From January to May	-46.6	-35.9	-39.9
From March to May (lockdown)	-76.0	-59.5	-66.0
Whole 2020 (average of % change)	-45.5	-32.3	-38.4
Whole 2020 (average % change)	-50.2	-33.9	-40.9

Source: Processing based on data by Istat and Agenzia delle Entrate data

As well as for descriptive purposes, the electronic invoicing data can be used to outline three possible scenarios related to the type of reaction that the Italian farmhouse sector could implement during and after the lockdown period. Even though electronic invoicing data are real data and not simulated data, they cannot be supposed to be used as they are for analysing the economic behaviour of farmhouses, because AFs are classified as part of the ATECO divisions 55 and 56 and no more detailed electronic invoicing sector data are available up to now.

The use of electronic invoicing data for defining different scenarios makes it necessary to introduce another basic information source, which is given by the available data derived from tourism statistics carried out by Istat. The survey on "*overnight stays of costumers in hotels and other accommodation establishments*" is carried out by Istat with the support of statistical regional offices. Data picked up monthly are detailed by kind of accommodation and are available for AFs as well. At the end of January 2021, the last available data referred to July 2020, so that the available monthly time series for 2020 covers the months from January to July. Percentage change of nights spent compared to the same month of 2019 is available as well. According to the survey monthly data (Figure 3), in the first 7 months of 2020 the number of nights spent in agritourist accommodation decreases by the 51% if compared to the first 7 months of 2019.

Figure 3 - Percentage change of night spent in AFs and in all kind of accommodation in the first seven months of 2020 compared to the corresponding months of 2019



Source: Processing based on Istat data - Survey on nights spent in tourist accommodations

Hereafter the description of three draft scenarios based on tourism statistics.

1. Pessimistic scenario: it assumes that the AFs cannot reduce the gaps with respect to 2019 during the months from August to December 2020. We suppose that the uncertainties after the lockdown and the second wave of COVID-19 occurred after Summer 2020 produced a worsening effect, which led to monthly increases of losses with respect to 2019 from August to December. We suppose that losses in December 2020 are equal to the average loss accounted in the first seven months of 2020 (-51%) and that the path from August to December 2020 is characterised by the monthly increase of losses with respect to 2019 equal to 4.5%. Consequently, the monthly losses would be: -28.4 in July (true data), -32.9% in August, -37.4% in September, -41.9% in October, -46.5% in November. According to this scenario, the loss in 2020 with respect to 2019 would be equal to 44.7%.
2. Intermediate scenario: the monthly loss registered in all the months from August to December 2020 is assumed equal to the loss recorded in July 2020. As a matter of fact, in July 2020 nights spent in AFs were the 28.4% lower than the nights spent in July 2019. Therefore, in the whole 2020 nights spent in AFs would be 40.4% less than in 2019. This scenario is moderately optimistic since it would imply a loss (40.4%) lower than the -51% registered during the first seven months of 2020.
3. Optimistic scenario: it assumes that the AFs can achieve a progressive reduction in losses from July 2020 until December 2020. We suppose that the loss equal to zero with respect to the same month of 2019 could be reached in December 2020, and that reduction of losses from July to December follows a steady monthly pattern. Starting from the -28.4% registered in July 2020, we suppose -22.7% in August, -17% in September, -11.3% in October, -5.7% in November, with a monthly loss reduction equal to 5.7%. According to this scenario, the loss in 2020 with respect to 2019 would be equal to 34.9%.

Obviously, these are theoretical scenarios, because:

- true data on nights spent are not available from August to December 2020;

- no separate scenarios for AFs with: a) catering only and with b) catering and lodging are available;
- no regional breakdown is available.

To obtain more specific scenarios concerning AFs with (a) catering only and with (b) catering and lodging, the three previous scenarios derived from tourism statistics have been integrated with the electronic invoicing trends resumed in the Table 6. The three yearly losses estimated for the whole agritourist sector according to the pessimistic scenario (-44.7%), the intermediate scenario (-40.4%) and the optimistic scenario (-34.9%) have been reparametrised to the losses of incomes accounted by electronic invoicing for the whole 2020 separately for AFs with catering only (-33.9%) and the AFs with lodging and catering (-40.9%). The results, resumed in the Table 7, are as follows:

- pessimistic scenario: AFs with catering only lose 40.5% of incomes, while AFs with lodging and catering lose 48.9%;
- intermediate scenario: AFs with catering only lose 36.6% of incomes, while AFs with lodging and catering lose 44.2%;
- optimistic scenario: AFs with catering only lose 31.6% of incomes, while AFs with lodging and catering lose 38.2%.

Table 7 - Estimated changes of AFs' incomes in 2020 compared to 2019, according to three scenarios, based on electronic invoicing data and tourism statistics

SCENARIOS	Night spent in agritourism farms	Agritourist farms with catering only	Agritourist farms with lodging and catering
		% change 2020/2019	% change 2020/2019
Pessimistic Scenario	-44.7	-40.5	-48.9
Intermediate Scenario	-40.4	-36.6	-44.2
Optimistic Scenario	-34.9	-31.6	-38.2

Source: Processing based on data by Istat and Agenzia delle Entrate

In order to obtain estimates for each Italian Region, the main problem was the lack of short-term indicators on available agritourism revenues at regional level. Therfore identifying different modalities and speed of reaction with regard to the pandemic depending on the territorial component was not possible.

Now, since there are not territorial indicators aimed at measuring the revenues of the farmhouses on a monthly basis, the data on the monthly tourist attendance in the Italian AFs[16] in 2019 have been used (Table 8). Therefore, it was possible to assign to each region (or autonomous province) a specific seasonality pattern of revenues, crossing the number of monthly overnight stays of Italian or foreign guests with the 2019 and 2020 monthly electronic invoicing data available for Italy as a whole.

Table 8 - Estimated losses by Region based on three scenarios (percent losses with respect to the 2019 yearly revenues)

REGIONS	Pessimistic scenario		Intermediate scenario		Optimistic scenario	
	Lodging and catering	Catering	Lodging and catering	Catering	Lodging and catering	Catering
Piemonte	-46.6	-38.8	-41.1	-34.2	-35.9	-29.9
Valle d'Aosta/ *Vallée d'Aoste*	-45.5	-37.2	-41.1	-33.6	-36.6	-30.2
Lombardia	-45.7	-38.4	-39.5	-33.1	-34.8	-29.1
Bolzano/*Bozen*	-48.0	-39.6	-43.9	-36.4	-38.0	-31.5
Trento	-44.8	-36.8	-40.6	-33.5	-36.0	-29.8
Veneto	-43.3	-35.8	-38.8	-32.2	-34.5	-28.7
Friuli-Venezia Giulia	-47.5	-39.7	-42.6	-35.7	-37.2	-31.1
Liguria	-49.5	-41.1	-45.3	-37.9	-39.1	-32.7
Emilia-Romagna	-47.9	-40.0	-43.0	-36.0	-37.7	-31.5
Toscana	-49.8	-41.5	-45.1	-37.7	-38.7	-32.3
Umbria	-48.8	-40.8	-41.8	-34.7	-36.2	-30.0
Marche	-52.4	-43.7	-48.7	-40.8	-41.8	-35.1
Lazio	-49.5	-41.5	-44.5	-37.5	-38.4	-32.3
Abruzzo	-50.5	-42.2	-45.6	-38.2	-39.4	-33.0
Molise	-49.0	-41.1	-43.4	-36.5	-37.7	-31.6
Campania	-50.5	-42.4	-44.5	-37.3	-38.4	-32.2
Puglia	-53.0	-44.4	-48.5	-40.8	-40.9	-34.3
Basilicata	-50.9	-42.3	-45.5	-37.8	-38.8	-32.2
Calabria	-53.1	-44.2	-48.0	-40.0	-39.9	-33.3
Sicilia	-48.1	-40.2	-42.3	-35.3	-36.1	-30.1
Sardegna	-54.5	-45.9	-49.6	-42.0	-41.0	-34.4
Italy	*-48.9*	*-40.5*	*-44.2*	*-36.6*	*-38.2*	*-31.6*

Source: Processing data by Istat (Nights spent by customers in accommodation facilities) and Agenzia delle Entrate

16 Istat, 2020*d* e 2020*e*.

A limitation of this approach is that in each region, given its specific seasonality about the accommodation service, it was necessary to assign a seasonal profile to the catering service like that of the accommodation activity. A further problem arises, as already stated, because of the lack of short-term territorial indicators that can provide information on the type of reaction that each region was able to put in place since the end of the lockdown and along all the 2020 months. According to Table 8, for example, in Piemonte, with the pessimistic scenario, in 2020 farms with accommodation and catering would lose 46.6% of revenues, while those with only catering would lose 38.8%. According to the intermediate scenario, losses would be 41.1% and 34.2% respectively.

Based on these estimated changes, the largest losses would be recorded for farms that supply highly seasonal catering or accommodation activity. In fact, the decline in the revenues of enterprises with catering only – less subject to seasonality – should be lower than the decline for those that also offer accommodation[17]. This interpretation also explains the territorial differences between the performance of farmhouses of the North-Centre of Italy and those of the South, on which the effects of seasonal demand weigh more significantly. In this context, the territorial reality that suffers of minor losses in all three contexts analysed is the autonomous province of Bolzano/*Bozen*.

To bridge the evaluations and the draft scenarios introduced and commented in this section with the microsimulation carried out in the next section, the concluding recommendation is as follows. The previous scenarios are “draft” because they are based on sources, as electronic invoicing and tourism statistics, which are not fully suitable for the economic activities analysed in this context. While electronic invoicing data are available for the whole economic divisions 55 and 56 (and not for the agritourist activities alone), tourism nights spent are available until July 2020 (and not for the whole 2020) and nights spent are only proxies of the true unknown AFs' incomes.

17 Istat, 2020*c*.

5. Microsimulation of the effects of COVID-19

The most recent trends in turnover of accommodation and restaurant services from the previous paragraph is incorporated in the following micro-founded analysis and represent a benchmark of the sectoral range of the change at regional level. In this way, we link macro analysis with micro results. It means that the values of the loss of turnover at micro level are calibrated to obtain the same percentage decreases for the farmhouses with prevalent activities in the ATECO 55 and 56 divisions (accommodation and restaurant).

The methodology used for assessing the impact of the pandemic on the turnover and of the closing risk at micro level is the microsimulation[18]. Traditionally, this model estimates the effects of specific policy changes (*e.g.* tax policy). In this case, we provide an estimate of the turnover reduction that depends on the increasing in regulatory constraints that affect the reduction of accommodation and restoration capacity. The first step involved the projection based on the sectoral and regional trends of National Account data[19], to adjust the VA level to the year 2019, which represents the base year of the microsimulation analysis. The equation relating to the reduction of revenues in the year 2020, for each scenario (s) is the following:

$$\Delta R^{20,s}_{r,c,i} = rp^{19}_{r,c} np^{19}_{r,c,i} \Delta\% np_s \tag{5.1}$$

where *r*=region, *c*=class of workers and *i*=agritourism.

The reduction of the turnover is calculated multiplying an indicator of average unit yield of available restaurant seats and beds in lodging ($rp^{19}_{r,c}$)[20] for the number of seats and beds ($np^{19}_{r,c,i}$), that is the accommodation capacity of the AFs, and for the assumed reduction $(\Delta\% np_s)$. This varies for each of the three scenarios[21]:

- (s = 1) the first more optimistic assumes a reduction in accommodation capacity of -25%, which corresponds to a closure of three months;

18 Caiumi and Di Biagio, 2016.

19 Istat 2020*a*. Benchmark values for 2019 consider a disaggregation level of 29 business sectors and 21 Regions (NUTS2).

20 The indicator is equal to the ratio between turnover and places (beds and restaurant seats) of the AFs with activities prevalent in the Ateco 55 and 56 divisions (accommodation and restaurant) by region and class of employees. The median value of this indicator has been attributed to all the farms of the belonging stratum (region r and class of workers c) to have a measure of the performance of the available places and of the losses in turnover.

21 The percentage reduction in the number of seats and beds is proportional to the months of closure during the year.

- (s = 2) the second also includes a reduction in places in order to comply with the distance rules and therefore a reduction of -37.5% of annual places is assumed;
- (s = 3) the third expects a more drastic reduction of -50% of the accommodation capacity.

Different scenarios refer to the different abilities of farms to readjust to the new distancing regulations. They also consider the different closures at the regional level. For the regions with major restrictions with “red” colour (Piemonte, Valle d’Aosta/*Vallée d’Aoste*, Lombardia, Bolzano/*Bozen* and Calabria), an additional month of closure is expected and half a month for the orange ones (Liguria, Toscana, Umbria, Abruzzo, Puglia, Basilicata and Sicilia).

The result of AFs with zero turnover is linked to the previous survival analysis made in the second paragraph. The theoretical number of market exit in three years was used as a benchmark to reset the turnover values of the units ordered according to the levels estimated in 2020 in the three scenarios at a province level. In cases where the number of AFs with zero turnover in the province is higher than the theoretical one, a minimum value of provincial turnover has been imputed to the farm. The complement to one of the survival rate by provincial population in 2018 gives a theoretical number of exit equal to 2,004. This value was re-proportioned to the number of farmhouses with zero turnover the province calculated in the three scenarios by province, respectively, equal to 2,313 (s=1), 3,904 (s=2) and 5,559 (s=3). The adjustment of the number of farms with zero turnover, which therefore are at risk of market exit, follows in this way the same mortality distribution by province of the previous survival analysis. At last, we obtain a total number of units without revenues from catering and accommodation equal to 2,257 in the first scenario, 3,850 in the second and 5,508 in the third one.

5.1 Final results

The number of farm units at risk of closure ranges from 9.6% in the most optimistic scenario (Table 9) to 23.3% in the worst scenario. From a dimensional point of view, the largest farms and those with 5-9 workers are less affected. Small AFs with less than one annual worker suffer more and a risk of zero turnover is estimated at 29.7% in the worst scenario. For larger ones, the risk is always under the average risk of closure. Among the regions, Lazio records the highest rates (Table 10), followed by Calabria and Molise, while Bolzano/*Bozen* records the lowest rates from 4.4% to 10.6%.

Table 9 - AFs at risk by class of workers, percentage values. Years 2019-2020

CLASS OF WORKERS	Exit risk rate		
	Scenario 1	Scenario 2	Scenario 3
1	11.6	20.1	29.7
2	7.9	13.5	18.6
3-4	6.8	10.7	14.0
5-9	5.1	8.2	9.7
10-19	6.7	8.5	11.1
20 and more	3.4	6.3	7.4
Total	*9.6*	*16.3*	*23.3*

Source: Processing based on both Istat and administrative data - Extended Farm Register (FR2)

Table 10 - AFs at risk by region, percentage values. Years 2019-2020

REGIONS	Exit risk rate		
	Scenario 1	Scenario 2	Scenario 3
Piemonte	5.0	8.6	12.4
Valle d'Aosta/*Vallée d'Aoste*	6.7	11.7	18.3
Lombardia	7.8	13.4	19.2
Bolzano/Bozen	4.4	7.4	10.6
Trento	8.6	14.7	21.0
Veneto	6.3	10.7	15.5
Friuli-Venezia Giulia	5.2	8.8	13.0
Liguria	8.7	15.1	21.6
Emilia-Romagna	7.1	12.3	17.7
Toscana	9.7	16.5	23.5
Umbria	9.4	16.0	22.9
Marche	5.2	8.8	12.6
Lazio	22.7	38.6	54.9
Abruzzo	11.0	18.8	27.3
Molise	18.8	32.8	46.9
Campania	12.6	21.6	30.8
Puglia	13.2	22.8	32.5
Basilicata	11.8	20.3	29.4
Calabria	20.5	35.3	50.4
Sicilia	16.3	27.7	39.8
Sardegna	16.1	27.5	39.2
Italy	*9.6*	*16.3*	*23.3*

Source: Processing based on both Istat and administrative data - Extended Farm Register (FR2)

Based on the decrease in turnover, the decrease in the VA of the whole business unit was calculated[22]:

$$\Delta VA_{r,c,i}^{20,s} = \frac{\Delta R_{r,c,i}^{20,s}}{R_{r,c,i}^{19}} VA_{r,c,i}^{19} \tag{5.2}$$

In 2019, the turnover reached 3.15 billion and in 2020 we expect a reduction that may range between 14.3% and 23.4%. Value added amounts to more than 1.6 billion and in the three scenarios this value was equal to 1.33, 1.24 and 1.16 billion, respectively. The change is equal to -16.9% (Table 11) in the best scenario, falls to -22.6% in the intermediate scenario and reaches -27.8% in the worst-case scenario. The strongest reduction is recorded for smaller farms (from -25.5% to -40.6%), while larger ones contain losses from -5.7% to -9.7% due to the greater diversification of activities, such as agricultural production, processing and marketing.

Table 11 - Turnover and value added by class of workers, million euros and percentage change. Years 2019-2020

CLASS OF WORKERS	Turnover				Value added			
	2019	2020 scenario 1	2020 scenario 2	2020 scenario 3	2019	2020 scenario 1	2020 scenario 2	2020 scenario 3
1	382.5	-22.0	-28.8	-35.1	211.8	-25.5	-33.4	-40.6
2	334.6	-20.9	-27.9	-34.4	204.7	-22.2	-29.9	-37.0
3-4	665.0	-20.6	-27.5	-33.5	402.0	-21.6	-29.0	-35.6
5-9	407.3	-15.6	-20.8	-25.8	218.5	-16.1	-21.8	-27.0
10-19	395.8	-13.0	-17.1	-20.8	176.2	-15.7	-20.6	-25.0
20 and more	961.8	-4.6	-6.3	-7.9	392.2	-5.7	-7.8	-9.7
Total	*3,147.0*	*-14.3*	*-19.1*	*-23.4*	*1,605.4*	*-16.9*	*-22.6*	*-27.8*

Source: Processing based on both Istat and administrative data - Extended Farm Register (FR2)

Toscana holds the economic record with 441 million of VA (Table 12) followed by Bolzano/*Bozen* and Lombardia. Among the regions of the South and Islands area (South and Islands), Puglia records about 49 million VA, with drops that could be between -25.3% and -39.9%. The regions of the South and Islands area recorded the greatest declines and the worst situation could happen for Basilicata, Campania and Sardegna, with decreases twice higher than the Italian average.

22 In this analysis, the assessment of the economic effects of the pandemic considers the business unit including auxiliary activities. Therefore, the changes in revenues are lower if compared with the trends of the sectors that include all the hospitality businesses. The change in the VA depends on the change in turnover and does not consider the efficient behaviour of the business unit in terms of reducing intermediate variable costs.

Considering the regions of the Centre, Umbria shows the greatest drops (between -22.3% and -35.3%), while Toscana shows smaller drops in terms of VA in the three scenarios. Among the Northern regions, except for Bolzano/Bozen, Liguria and Valle d'Aosta/*Vallée d'Aoste*, which could suffer the greater seasonality, the expected decreases are always lower than or equal to the national average.

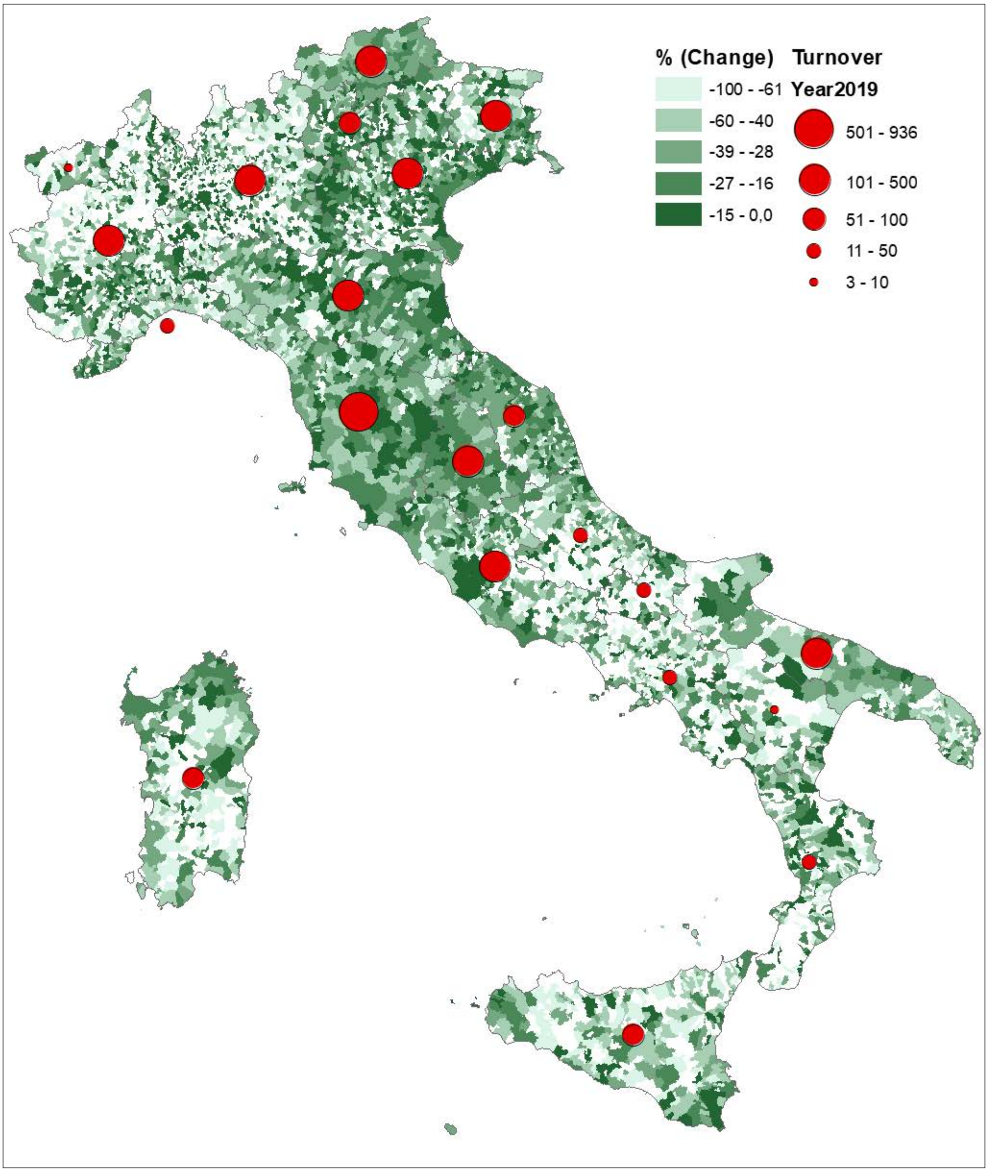
Table 12 - Turnover and value added by Region, million euros and percentage change. Years 2019-2020

REGIONS	Turnover				Value added			
	2019	2020 scenario 1	2020 scenario 2	2020 scenario 3	2019	2020 scenario 1	2020 scenario 2	2020 scenario 3
Piemonte	176.1	-14.3	-18.5	-22.5	95.2	-14.8	-19.6	-23.9
Valle d'Aosta/*Vallée d'Aoste*	2.8	-26.1	-34.6	-42.7	1.3	-41.5	-54.3	-67.1
Lombardia	416.0	-14.8	-18.9	-23.1	199.7	-16.8	-21.7	-26.7
Bolzano/*Bozen*	291.9	-22.0	-29.2	-35.7	215.7	-22.6	-30.3	-37.3
Trento	74.2	-10.5	-14.1	-17.4	37.9	-12.1	-16.7	-20.9
Veneto	254.9	-10.5	-14.1	-17.4	129.2	-12.0	-16.2	-20.2
Friuli-Venezia Giulia	101.1	-14.8	-20.2	-24.8	58.9	-14.0	-19.8	-24.6
Liguria	48.0	-17.3	-22.9	-27.2	26.9	-19.6	-26.3	-31.5
Emilia-Romagna	221.9	-13.0	-17.3	-21.3	110.3	-13.6	-18.4	-22.9
Toscana	935.6	-10.3	-13.9	-17.0	441.2	-13.8	-18.3	-22.3
Umbria	109.5	-15.1	-20.0	-25.0	44.5	-22.3	-28.8	-35.3
Marche	62.9	-18.3	-24.2	-28.9	28.7	-20.7	-28.1	-33.9
Lazio	112.0	-15.8	-22.0	-27.9	55.4	-19.1	-26.8	-33.3
Abruzzo	26.0	-21.4	-28.5	-34.7	12.8	-23.2	-31.3	-38.5
Molise	11.0	-6.5	-9.0	-11.7	2.5	-16.6	-22.6	-28.4
Campania	37.7	-23.5	-31.6	-38.5	16.7	-28.9	-38.6	-46.9
Puglia	104.8	-21.8	-28.6	-33.8	48.7	-25.3	-33.4	-39.9
Basilicata	9.1	-24.9	-32.6	-39.3	4.0	-31.8	-40.2	-47.4
Calabria	29.0	-19.8	-26.5	-33.8	12.8	-23.8	-31.2	-41.0
Sicilia	66.9	-19.5	-25.5	-31.2	32.7	-21.5	-28.4	-34.7
Sardegna	55.7	-21.4	-30.2	-38.8	30.3	-22.5	-32.1	-42.0
Italy	*3,147.0*	*-14.3*	*-19.1*	*-23.4*	*1,605.4*	*-16.9*	*-22.6*	*-27.8*

Source: Processing based on both Istat and administrative data - Extended Farm Register (FR2)

In the following map (Figure 4) we can represent the distribution of losses at the municipality level. The red circles represent the importance of the region in terms of turnover in the base year (2019). The most affected areas of Italy are those with the less intensive green colour, which identifies the AFs with greater loss of turnover (South and Islands), while the most intense colour shows the municipalities with a relatively better performance (North-Centre).

Figure 4 - Turnover, regional levels in million euros and percentage change by Municipality (Years 2019-2020 intermediate scenario)



Source: Processing based on both Istat and administrative data - Extended Farm Register (FR2)

6. Conclusions

The pandemic is producing diversified effects on the economic system. From the Istat survey "Situation and prospects of enterprises in the health emergency COVID-19", emerged a first empirical evidence on the types of the most affected enterprises, which are those operating in the construction and service sectors, in particular catering, as confirmed by sectoral trends from electronic invoicing data. About agriculture, the secondary activities of farms which include agritourism were mainly affected with a decrease of 18.9% (Istat, 2021).

Another important evidence came from sectoral analysis made by ISMEA, where 86% of a sample of farmhouses recorded a decrease in revenues in 2020 and for two thirds of these the decrease in turnover was more than 30%. Starting from these analyses, in this work we have estimated the impact of the health emergency on the economic performance of the AFs: a sector that based its success on increasing hospitality with a raise of units of 15.7% between 2011 and 2018 and, therefore, suffered more than other farms during 2020. This work provides a comprehensive picture of the territorial distribution of AFs and survival together with a prospective micro economic analysis.

Thanks to data integration, at micro level, of the agritourism survey with the FR2, we measured the economic dimension of the AFs sector at a very detailed level and developed an impact analysis due to the pandemic crisis. Among the 24 thousand AFs, 51% belongs to the subset of businesses with prevalent agricultural activity, 26% is associated with SBS enterprises (with prevalent activities in industry or services) and the remaining 23% concerns less structured farms. In 2018, the value-added reached 1.6 billion of euros. The largest share (25%) is concentrated in farms with 3-4 workers and in some regions of the Centre and the North (Toscana, Lombardia and Bolzano/*Bozen*). For these 25 thousand farms, it was possible to explore the economic performance in terms of labour productivity, profitability and export share with Centre-North and larger farms more competitive with respect to South ones.

Several legislative actions were launched in 2020 to support agriculture and related activities, depending on the economic activity carried out. The last intervention was that of 2021 (Legislative Decree 41 of 23/03/2021), which provides for the exemption from contributions for agricultural enterprises. Sectoral trends confirm the diversified impact on the whole economic system and catering activities show the main drops. Based on these estimated changes, the largest losses would be recorded for farms that supply highly seasonal catering or accommodation services. In fact, the decline in the revenues of enterprises with catering only, less subject to seasonality, should be lower than the decline for those that also offer accommodation.

The macro analysis of the pandemic, from electronic invoicing data, allowed us to calibrate micro-founded analysis of the effect of lockdown on each single unit and to reproduce macro trends through a bottom-up approach.

The microsimulation of economic data in 2020 includes three scenarios, so as to provide a range of variation that depends on different periods of closure or contraction of activities. We estimated a percentage of AFs with risk of closure between 9.6% and 23.3%. From a dimensional point of view, larger farms are less affected. Smaller farms risk more, with a rate of 29.7% in the worst-case scenario. Estimates for 2020 foresee a reduction in turnover between -14.3% to -23.4% (-19.1% in the mean scenario) and a reduction in value added between -16.9% and -27.8% (-22.6% in the mean scenario).

The strongest reduction is estimated for the small farms, while the large ones contain it below the 10%, due to their larger diversification of activities, such as agricultural production, processing and marketing. The regions of South and Islands area recorded the greatest drops together with Bolzano/*Bozen*, Liguria and Valle d’Aosta/*Vallée d’Aoste*, while Veneto, Trento and Toscana show lower drops. The final map offers a portrait of the distribution of losses at the municipal level and locate the fading areas, that is the territories most affected by the spread of the pandemic on the activities of AFs throughout the Italian territory.

Finally, the dimensional aspect plays a crucial role, indeed the areas with minor loss of turnover are the same area characterised by a larger average size where the capacity of diversification allows mitigating the loss in revenues for the tourism lack.

References

Amato, R. 2020. "Agriturismi: buco da un miliardo per il COVID. Subito gli aiuti". *la Repubblica* (9/11/2020).

Baccino, F. 2020. "Il fatturato degli agriturismi crolla a 600 milioni di euro nel 2020". *Terra e Vita* (20/07/2020).

Caiumi, A., and L. Di Biagio. 2016. "The ISTAT–MATIS corporate tax model". *Rivista di statistica ufficiale*, N. 2/2016. Roma: Istat.

Centro Studi Confagricoltura. 2020. *Agriturismo: nel 2020, riduzione del fatturato prossima a 1,5 miliardi*. Roma: Centro Studi Confagricoltura.

Istituto di Servizi per il Mercato Agricolo Alimentare – ISMEA. 2021. Emergenza COVD-19. *IV Rapporto sulla domanda e l'offerta dei prodotti alimentari nell'emergenza COVID-19*. Roma: ISMEA.

Istituto di Servizi per il Mercato Agricolo Alimentare – ISMEA. 2020*a*. "Agriturismi, un miliardo di perdite nel 2020". *Il Sole 24ore* (05/05/2020).

Istituto di Servizi per il Mercato Agricolo Alimentare – ISMEA, e Rete Rurale Nazionale 2014-2020. 2020*b*. *Agriturismo e multifunzionalità. Scenario e prospettive. Rapporto 2020*. Roma: ISMEA.

Istituto di Servizi per il Mercato Agricolo Alimentare – ISMEA, e Rete Rurale Nazionale 2014-2020. 2019. *Agriturismo e multifunzionalità. Scenario e prospettive. Rapporto 2019*. Roma: ISMEA.

Istituto Nazionale di Statistica - Istat. 2021. "Stima preliminare dei conti economici dell'agricoltura. Anno 2020" *Statistiche Report*. Roma: Istat. https://www.istat.it/it/archivio/252819.

Istituto Nazionale di Statistica - Istat. 2020*a*. "Conti Economici Territoriali. Anni 2017-2019". *Statistiche Report*. Roma: Istat. https://www.istat.it/it/archivio/251960.

Istituto Nazionale di Statistica - Istat. 2020*b*. "L'andamento dell'economia agricola. Anno 2019". *Statistiche Report*. Roma: Istat. https://www.istat.it/it/archivio/243183.

Istituto Nazionale di Statistica - Istat. 2020*c*. "Una stagione mancata: impatto del COVID-19 sul turismo". *Statistiche Today*. Roma: Istat. https://www.istat.it/it/archivio/242017.

Istituto Nazionale di Statistica - Istat. 2020*d*. "Fatturato dei servizi. III Trimestre 2020". *Statistiche Flash*. Roma: Istat. https://www.istat.it/it/archivio/250918.

Istituto Nazionale di Statistica - Istat. 2020*e*. *Indagine Istat sul movimento dei clienti negli esercizi ricettivi*. Marzo 2020.

Istituto Nazionale di Statistica - Istat. 2020*f*. "Situazione e prospettive delle imprese nell'emergenza sanitaria COVID-19". *Statistiche Report*. Roma: Istat. https://www.istat.it/it/archivio/244378.

Istituto Nazionale di Statistica - Istat. 2019*a*. "Conti economici delle imprese e dei gruppi di impresa. Anno 2017". *Statistiche Report*. Roma: Istat. https://www.istat.it/it/archivio/234313.

Istituto Nazionale di Statistica - Istat. 2019*b*. "Le aziende agrituristiche in Italia. Anno 2018". *Statistiche Report*. Roma: Istat. https://www.istat.it/it/archivio/235458.

Istituto Nazionale di Statistica - Istat. 2019*c*. "Struttura e caratteristiche delle unità economiche del settore agricolo. Anno 2017". *Statistiche Report*. Roma: Istat. https://www.istat.it/it/archivio/236288.

Kim, J.K., and J. Shao. 2014. *Statistical Methods for Handling Incomplete Data*. Boca Raton, FL, U.S.: Chapman and Hall/CRC Press.

Luzi, O., and R. Monducci. 2016. "The new statistical register "Frame SBS": overview and perspectives". *Rivista di statistica ufficiale*, N. 1/2016. Roma: Istat.

Magliocchi, M.G., F. Oropallo, e F. Truglia. 2019. "La diversificazione e le strategie delle imprese agricole". In *XL Conferenza scientifica annuale AISRE*. L'Aquila, Italy, 16th - 18th September 2019.

Mastronardi, L., e V. Giaccio. 2011. "Le performance delle aziende agrarie con e senza agriturismo: un confronto con i dati RICA". *Agriregionieuropa*, N. 26, *Approfondimenti*: 83-86.

Oropallo, F. 2021. "Farm Register esteso: integrazione e potenzialità". Currently being published in *Istat working papers*, 2021. Roma: Istat.

Palmi, P., and G.E. Lezzi. 2020. "How Authenticity and Tradition Shift into Sustainability and Innovation: Evidence from Italian Agritourism".

International Journal of Environmental Research and Public Health, Volume 17, N. 15.

Ranerelli, T. 2020. "Cia: Non c'è rilancio delle aree interne senza agriturismi". *Terra e Vita* (07/10/2020).

Truglia, F.G., F. Oropallo, R. Gismondi, e M.G. Magliocchi. 2020. "Analisi dell'impatto COVID-19 sugli Agriturismi". In *XLI Conferenza scientifica annuale AISRE*. Web Conference, 2nd - 4th September 2020.

United Nations World Tourism Organization - UNWTO. 2020. *International tourist numbers could fall 60-80% In 2020, UNWTO Reports*. Madrid, Spain: UNWTO.

The *Rivista di statistica ufficiale* publishes peer-reviewed articles dealing with cross-cutting topics: the measurement and understanding of social, demographic, economic, territorial and environmental subjects; the development of information systems and indicators for decision support; the methodological, technological and institutional issues related to the production process of statistical information, relevant to achieve official statistics purposes.

The *Rivista di statistica ufficiale* aims at promoting synergies and exchanges between and among researchers, stakeholders, policy-makers and other users who refer to official statistics at different levels, in order to improve data quality and enhance trust.

The *Rivista di statistica ufficiale* was born in 1992 as a series of monographs titled "*Quaderni di Ricerca Istat*". In 1999 the series was entrusted to an external publisher, changed its name in "*Quaderni di Ricerca - Rivista di Statistica Ufficiale*" and started being published on a four-monthly basis. The current name was assumed from the Issue N. 1/2006, when the Italian National Institute of Statistics – Istat returned to be its publisher.

La Rivista di statistica ufficiale pubblica articoli, valutati da esperti, che trattano argomenti trasversali: la misurazione e la comprensione di temi sociali, demografici, economici, territoriali e ambientali; lo sviluppo di sistemi informativi e di indicatori per il supporto alle decisioni; le questioni metodologiche, tecnologiche e istituzionali relative al processo di produzione dell'informazione statistica, rilevanti per raggiungere gli obiettivi della statistica ufficiale.

La Rivista di statistica ufficiale promuove sinergie e scambi tra ricercatori, stakeholder, policy-maker e altri utenti che fanno riferimento alle statistiche ufficiali a diversi livelli, al fine di migliorare la qualità dei dati e aumentare la fiducia.

La Rivista di statistica ufficiale nasce nel 1992 come serie di monografie dal titolo "Quaderni di Ricerca Istat". Nel 1999 la collana viene affidata a un editore esterno, cambia nome in "Quaderni di Ricerca - Rivista di Statistica Ufficiale" e diventa quadrimestrale. Il nome attuale è stato scelto a partire dal numero 1/2006, quando l'Istituto Nazionale di Statistica - Istat è tornato a esserne l'editore.

ISSN 1828-1982 € 15,00