



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Head Office: Università degli Studi di Padova

Department of Cardiac, Thoracic, Vascular Sciences and Public Health

Ph.D. COURSE IN:

Translational Specialistic Medicine 'G.B. Morgagni'

CURRICULUM: Biostatistics and Clinical Epidemiology

SERIES: XXXIV

**SOCIAL NETWORK-BASED APPROACHES IN EPIDEMIOLOGICAL OUTBREAKS
SURVEILLANCE**

Coordinator: Prof. Annalisa Angelini

Supervisor: Prof. Dario Gregori

Ph.D. student: Honoria Ocagli

Università degli Studi di Padova

Department of Cardiac, Thoracic, Vascular Sciences and Public Health

Abstract

Social network-based approaches in epidemiological outbreaks surveillance

By Honoria Ocagli

In public health, determining the population's size is an important issue, especially for those who are hidden or hard to reach. Prevalence estimates are required in three key areas: resources of the target population and public health surveillance and epidemiology.

Different techniques are available in the literature to estimate these populations, such as the enumeration method, capture-recapture technique, multiplier method, and the network scale-up method (NSUM). This last method was developed to count the death after an earthquake in Mexico and was then widely applied in the field of a hard-to-reach population.

This thesis is developed in five chapters. The first chapter is the introduction to the problem. The second chapter presented a systematic review retrieving studies proposing methodological developments and applications of the NSUM estimator. The third and fourth chapters proposed a modified parametrization of the Bayesian formulation of the Maltiel NSUM method accounting for a short-form questionnaire to reduce the non-response bias. The proposed estimator considers the network size as partially unknown; only one (randomly sampled) question is posed to the respondents among the known populations, typically used in an NSUM parametrization to estimate the network size.

The third chapter is a study protocol proposing the application of modified NSUM for estimating the prevalence of COVID-19 undocumented cases during the early waves of the pandemic. A simulation study is also presented in this chapter. The proposed estimator demonstrated a slightly higher bias and variability in the estimates compared to the original Maltiel's formulation. A great advantage is reducing the non-response bias and the drop-outs compared to a long-form questionnaire as typically

used for an NSUM estimation. The study protocol in this chapter was published in the International Journal of Environmental Research and Public Health (DOI: 10.3390/ijerph18115713).

The fourth chapter reports the results of the Hidden COVID-19 Cases Network Estimation (INCIDENT) study, as reported in the study protocol, estimating the prevalence of COVID-19 cases and undocumented cases in the Veneto, Piemonte, and Lombardy regions. The number of people that transferred from one region to another after the lockdown restriction and the number of cohabitants of subjects affected by COVID-19 were also considered hidden populations.

The fifth chapter reports the conclusion of this thesis. The method proposed in this work tries to overcome some limitations of the original NSUM. To evaluate the performance of this method, it will be helpful to compare a different survey to the original one.

Acknowledgments

In this P.h.D. journey, I have encountered many people to whom I must be grateful. My appreciation and thanks go to my P.h.D. supervisor, Prof. Dario Gregori, Full Professor of Statistics at the University of Padova, for he has trusted me since the beginning of this journey. I am grateful to Giulia Lorenzoni, who has supervised me. I want to thank Danila Azzolina, who is my co-author of this work and has encouraged and guided me through the draft of this thesis.

I also would like to thank all the people who have been involved in my P.h.D. journey; each of them has contributed to my results.

Special thanks go to my husband and my daughter.

Table of contents

Abstract	3
Acknowledgments	6
List of figures	9
List of Tables	10
Chapter 1	11
Introduction	11
Chapter 2	13
A systematic review of network scale-up use: from the traditional model to its modifications	13
2.1 Introduction	13
2.2 Methods	14
2.2.1 Information source and search strategy	15
2.2.2 Inclusion criteria	15
2.2.3 Study selection	15
2.2.4 Data extraction	15
2.2.5 Assessment of risk of bias	15
2.3 Results	16
2.3.1 Study characteristics	17
2.3.2 Hidden population	17
2.3.3 Estimation of the network size in a different survey	21
2.3.4 NSUM modifications	21
2.3.5 Agreement among the different NSUM	22
2.3.6 NSUM use in official reports	23
2.4 Discussion	23
2.5 Conclusion	24
Chapter 3	25
Using Social Networks to Estimate the Number of COVID-19 Cases: The Incident (Hidden COVID-19 Cases Network Estimation) Study Protocol	25
3.1 Summary	25
3.2 Introduction	26
3.3 Materials and Methods	27
3.3.1 Study Design	28
3.3.2 Procedures	28
3.3.3 Ethics	28
3.3.4 The Network Scale-Up Method in the Literature	28
3.3.5 NSUM Questionnaire	32
3.4 NSUM Assumption	33
3.4.1 NSUM Estimator	33
3.4.2 Bayesian NSUM Estimation	34
3.4.3 Extended Random Degree Model	35

3.4.4 Performance of the Modified Maltiel Estimators.....	36
3.4.5 Statistical Analysis	36
3.5 Results of the Simulation Study	37
3.5.1 Performances of the Modified Maltiel’s Estimators	37
3.6 Discussion.....	38
Chapter 4.....	40
Indirect estimation of COVID-19 asymptomatic cases using a shortened Network Scale-Up approach. An application to Italy in the First Wave	40
4.1 Summary.....	40
4.2 Introduction.....	41
4.3 Materials and methods.....	42
4.3.1 Study design	42
4.3.2 Model Estimation	42
4.3.3 Random-degree (RD) model	43
4.3.4 Barrier effect (BE) model.....	44
4.3.5 Transmission bias (TB) model	44
4.3.6 Maltiel model modifications	45
4.3.7 Data Preparation	45
4.3.8 Known subpopulation.....	46
4.3.9 Personal network size estimation	46
4.4 Results.....	46
4.4.1 Descriptive characteristics of the sample	46
4.4.2 Estimates of COVID-19 cases.....	55
4.4.3 Undocumented COVID-19 cases	58
4.4.4 Cohabitants of isolated COVID-19	58
4.4.5 Number of people that transfer after the Law act (DPCM).....	58
4.5 Discussion.....	59
4.6 Conclusions.....	61
Chapter 5.....	63
Conclusion	63
Bibliography.....	65
Supplementary materials.....	79

List of Figures

Figure 1 Prisma Flowchart diagram.....	16
Figure 2 Flowchart of study review	29
Figure 3 Estimated prevalence of the four target questions in the Veneto, Piemonte, and Lombardia regions for μ that goes from 3 to 8 (20 to 30000 network size) with the random degree model. Estimates are reported over 100 inhabitants.	52
Figure 4 Estimated prevalence of the four target questions in the Veneto, Piemonte, and Lombardia regions for μ that goes from 3 to 8 according to the transmission bias model. Estimates are reported over 100 inhabitants.	53
Figure 5 Estimated prevalence of the four target questions in the Veneto, Piemonte and Lombardia regions for μ that goes from 3 to 8 according to the barrier effect model. Estimates are reported over 100 inhabitants.	54
Figure 6 Average prevalence and trend of COVID-19 positive prevalence in the Italian regions considered in the study: Veneto (blu), Lombardia (red) and Piemonte (green). Reported prevalence refer to the proportion of COVID-19 cases in the population of each region in the period between 15 April 2020 and 6 May 2020.	55
Figure 7 Variation from official data for the random degree model according to region and different μ	57
Figure 8 Variation from official data for the Transmission bias model according to region and different μ	57
Figure 9 Variation from official data for the Barrier Effect model according to region and different μ	58

List of Tables

Table 1 Full lists of the retrieved studies.	18
Table 2 Studies that estimate the hidden population using the network scale-up method or its modification.	30
Table 3 List of known-size populations used in the questionnaire to estimate respondents' network sizes.	32
Table 4 Simulation results for the modified Maltiel's method performance.	37
Table 5 Descriptive statistics of the study sample according to the region. Continuous data are reported as median (I, III quartiles); categorical data are reported as a percentage and absolute frequencies. In the table are reported demographic characteristics, respondents' answers to the target population, and the known populations.	47
Table 6 Estimates of the four target questions in Veneto, Piemonte, and Lombardia regions between 15 April 2020 and 6 May 2020. The estimates are reported for network size, μ goes from 3 to 8, which means a network size ranging from 20 to 3000. For each estimate, the posterior median, lower and upper bound for the 95% Credible Interval are reported.	50
Table 7 Differences of the estimates from swab COVID-19 positive with default μ , the network size, for each region considered. The differences are reported in percentages.	56

Chapter 1

Introduction

Including all populations in studies in the public health field is an important issue. In developed countries, studies are often based on the response from a specific population, such as the middle class [1]. Also, in clinical trials has been noted that minorities are not always represented, as shown in the review of Heiat et al. [2]. The problem of reaching all the subset of the population is reported especially for disadvantageous groups. “Hard to reach” [3] are defined as those who are difficult to access. Instead, hidden populations are those who, for various reasons, do not want to be reached, e.g., homeless, criminals, HIV positive [4]. In this framework arise two problems, one is to include these subjects in health and medical research, and the second is to count them. The first hurdle can be solved through interventions that go from sampling to retention [1]. This thesis will focus on the second topic. In literature, as will be shown later, various methods are used to count these hidden populations, such as the capture-recapture technique, enumeration methods, and multiplier estimates [5]. These methods, along with the Network Scale-Up method, are suggested by the UNAIDS and WHO as the primary instruments to quantify key populations at risk for HIV [6]. These methods can be divided into two categories, as Abdul-Quader et al. [7] suggest: *i)* collect data directly from the key population at risk, *ii)* collect data from the general population. The first category includes the capture-recapture technique, enumeration, multiplier, and census. In the second category are included survey and NSUM.

The methods that directly access these hidden populations have substantial limitations, considering the difficulty of reaching them. Moreover, the estimates derived are uncertain and often produce different results from each other [7]. Also, NSUM features these disadvantages, but the fact that it does not require direct access to the hidden population is very promising. In this thesis, the first chapter presented a systematic literature review on studies that have applied the NSUM. The review was done in two steps: in the first were searched all the articles that used NSUM, so both application and methodological studies are included. The second step considered only those who applied the NSUM. Moreover, the first search was done in four databases and the second in only one.

The network scale-up method “is based on the assumption that personal networks are, on average, representative of the general population”[8]. So, the method can be divided into two parts: one for estimating the network size and the other to estimate the hidden population. Both dimensions are based on the concept of “knowing someone”. In surveys that use NSUM is essential to define the meaning of knowing someone. Bernard et al., in their study, define knowing someone as follows:

“you know them, and they know you and you have been in contact with them in the last two years” [5]. The systematic review provides an overview of the modifications of the NSUM to overcome some biases that arise because the assumptions are strong. The main biases are transmission bias, recall bias, and barrier effect [5,9]. To estimate the network size of the respondents in the surveys are always reported questions related to the dimension of the known population. All the methods use a significant number of questions since it has been reported that the model estimates better with more than 20 known populations [10]. Using such a great number of questions could reduce the response rate, especially when the survey has multiple purposes. The response rate is lower for long questionnaires [11]. This limits the usability of the NSUM in studies that inquire about other themes than the estimation of the hidden population. So, from this standpoint, the third and fourth chapters of this thesis reported the protocol and then the results of a new modified version of the NSUM estimator. Chapter three is presented the statistical assumption behind our new model. Specifically, we propose a modified parametrization of the Bayesian formulation of the Maltiel NSUM method accounting for a short-form questionnaire to reduce the non-response bias. To determine the network size, only one question was randomly selected.

Another novelty of the use of the NSUM in this thesis is related to its use in the undocumented cases of COVID-19. Specifically, this population has been considered hidden due to its characteristics in the first wave of the pandemic. The fourth chapter reports the results of applying the modified NSUM to our survey. Defining the number of undocumented cases of COVID-19, especially in the period considered, was a real challenge [12]. Undocumented cases, including asymptomatic and paucisymptomatic, have shown the same transmissibility as symptomatic ones [13]. During the first wave, the virus was relatively new, and the measures used by policymakers to contain the contagion were pervasive [14]. Hence, undocumented COVID-19 cases could be defined as a hard-to-reach population since there are difficult to recognize.

The proposed modified version of the NSUM tries to overcome some limitations of the original estimator. Applying this method to a population related to clinical research may introduce new aspects that go beyond public health purposes.

Chapter 2.

A systematic review of network scale-up use: from the traditional model to its modifications

2.1 Introduction

The estimate of prevalence is a crucial issue in public health studies. However, some groups of the population are “hidden”, “do not wish to be found or to be contacted” [4], or are “difficult to reach”. Researchers must endure physical and geographical efforts to involve these groups in a study [3]. These key populations are still difficult to estimate and require looking for data difficult to obtain or make assumptions specific for each key population. Collecting data that are representative of all of the population is the base for helping policymakers reduce the inequalities [1]. Including all the population groups can be handled at different stages of the research, like sampling, recruitment, methods to improve response rate, use of inclusive language, intervention phase, and retention, as the systematic review of Bonevski et al. [1] reports.

In key populations, direct methods of size estimation are often impractical or unreliable: hence the need for indirect methods of estimation. Those methods aim to analyse the observed data set or combine it with other information to estimate the key population and then estimate the desired prevalence. In literature, there are different indirect estimation methods, such as the capture-recapture technique, multiplier method, the wisdom of the crowd, and the network scale-up.

The capture-recapture technique is a method recommended by the World Health Organization/Joint United Nations Programme for defining the prevalence of HIV/AIDS [15,16]. This method, in the first stage, “captures” a group of the key population, “tags” them, and then “releases” back them. The second sample in a second moment is “recaptured” independently. Some of the recaptured individuals have been previously tagged. The proportion of tagged and recaptured members is used to estimate the key population of the whole population [17]. This method is based on two critical assumptions: the two samples should be independent, and there is no migration among the people.

The multiplier method also requires two sources of data: one is a direct count of the target population; the other is a representative sample. The size is then estimated by multiplying the number of those who have received the service by the proportion reporting the service [18].

The enumeration method creates a sample of the population, selects a part of it, and counts it [6]. The number derived is scaled up to the size of the sample defined. This method performs well, with some uncertainties for hidden populations [18]. The study by Mutagoma et al. [18] estimated the size of

female sex workers (FSW) in Rwanda using the capture-recapture technique and the enumeration method considering the overall population. The multiplier method was used instead in the city of Kigaly. In this case, the methods provided similar results and suggested combining different size estimation methods.

The limit of these methods is that they rely on the need to access directly to the population and to have two independent samples with the assumption to have the same probability of appearing in each list, this is not always possible [19].

Other methods are based on social link tracing. For example, the “wisdom of the crowd” is a method “based on the assumption that the central tendency in the response of a population on the number of population members approximates or is proportional to the actual number of members in that population” [20]. This method is based on the discovery of Galton [21] in 1906. In his statistical analysis on the guess on the actual weight of an ox he found out that the collective guess was better than that of the winner of the contest and than the guesses made by cattle experts.

The NSUM estimates the size of the key population without direct access to it, and in the survey, participants are not asked direct questions related to their behaviours. This method relies on the assumption that the social network of a single individual is representative of the general population [5]. The NSUM, compared to the other indirect estimation methods, is an easy tool to use when it is not possible to access the key population directly. Two recent reviews have been published related to the NSUM. The first is the one of McCormick [22], in which the method is considered through NSUM application. The second review is by Laga et al. [9], in which the NSUM is explored, as said by the authors, through the modeling perspective.

Instead, this article systematically reviews the literature to search for articles related to the network scale-up method and its application. This work aims to identify all the studies associated with the NSUM and summarize the findings.

2.2 Methods

Despite not being a systematic review focusing on evaluating the effect of the intervention or exposure, this review has been based on the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement [23] as a basis for reporting this systematic review.

2.2.1 Information source and search strategy

Databases search includes Pubmed (through Medline), Embase, CINHALL, and Scopus. The last search was conducted on the 6th of September 2021. The reference list of other systematic reviews and included studies were also screened (backward citations searching). The search string is the combination of the concepts “network scale-up” and “hidden/hard to reach” population. Keywords are combined with the Boolean operators AND or OR. Table S1 reported the full string search.

2.2.2 Inclusion criteria

In the review are included articles that consider the NSUM approaches from a methodological and applied perspective. Systematic reviews, editorial, and commentaries were excluded. Studies not published in the English language were excluded too.

2.2.3 Study selection

The whole phase of the screening was done independently by two reviewers on the Covidence platform [24]. Disagreements were solved through discussion and, when needed, consulting another author. Figure 1 shows the phases of the study selection according to Prisma 2020 [25].

2.2.4 Data extraction

Two reviewers independently assessed extracted data on a form prepared in Covidence. A pilot data extraction test was conducted at first for three articles and then finalized for whole articles retrieved. The collected information are the following: author, country, study design, number of participants, hidden population considered, sampling technique, age of participants, response rate, number of questions to define the network size, the method for defining “knowing someone”, the reporting of a comparative approach and the bias quantification.

2.2.5 Assessment of risk of bias

The articles considered are heterogeneous concerning their study design, so a systematic scoring for evaluating the risk of bias (rob was not used). In the systematic review of Bonevski et al. [1] regarding strategies to improve health and medical research with a hard-to-reach population, a graded system of evidence based on study design was applied. Even this method was not suitable for this systematic review because the study design of the totality of the studies is observational, with a poor level of evidence.

2.3 Results

The electronic searches yielded 1366 articles from the three different databases, 66 were considered in the final analysis as reported in the PRISMA flowchart (Figure 1).

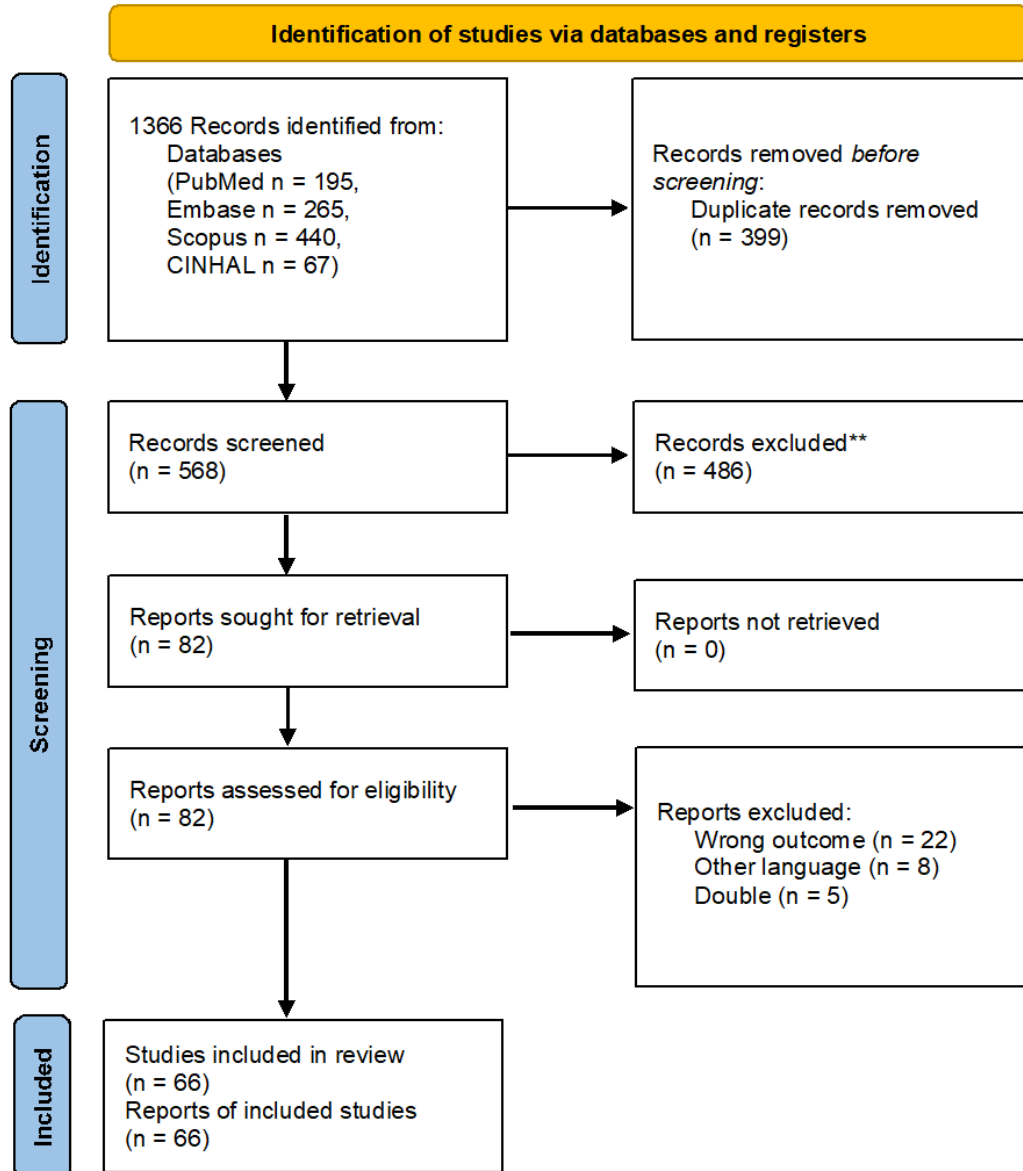


Figure 1 Prisma Flowchart diagram

Table 1 reports the characteristics of the selected studies. Most of the included articles were based on research conducted in Iran, 29 (46%), while 9 (14%) were conducted in the United States of America. The number of participants in the study ranged from 72 participants in the study of Snidero et al. [26] to 12960 in the study of Rastegari et al. [27]. This difference in the sample size is related to the fact that in the first case, the participants were medical specialists. In contrast, in the second case, the study was conducted nationwide in 31 provinces of Iran on the number of abortions.

The number of known populations varies from a minimum of 5 [28] and 6 [29,30] to a maximum of 48 in the studies of Jing et al. [31,32].

2.3.1 Study characteristics

The studies are mainly structured as interviews or surveys. According to what was reported by the authors, the response rate is higher in studies that use interviews than surveys. Survey's response rate range between 38.18% [33] to 97.7% [34], instead interviews response rate goes from 87% [35] to 99% [36].

Participants are usually enrolled through random sampling 15 (%), 11 (%) through multistage sampling, convenience sampling 9 (%), and the remaining in a mix of random cluster sampling, two-stage cluster sampling, and street-based random sampling.

Most of the studies have a cross-sectional study design. Some are methodological studies, and results are based on simulated data or data from other works. The data of the four telephone surveys of McCarty et al. [37] were used in the study of McCormick et al. [22,37,38], and Maltiel et al. [39] used the data on heavy drug users in Curitiba, Brazil [40,41].

2.3.2 Hidden population

The hidden population mostly considered among the selected studies are drug users, female sex workers, people at risk for HIV/AIDS, smoking habits, abortion, men who have sex with men, raped women, and extramarital sex. Other minor populations considered as hidden were people with cancer, choking injuries, religious status, earthquake death, people living with a disability, number of people unsatisfied, suicide deaths, and attempts of suicide. Table 1 reported the complete list of articles retrieved.

Table 1 Full lists of the retrieved studies.

Author year	Country	N participants	Hidden population	Sampling method	Response rate	N known population
Ahmadi-Gohari 2019	Iran	363	alcohol users, drug users	RS		Shokoohi et al. 2010
Bernard 1989	Mexico	400	quake victims	RS		6
Bernard 1991	Mexico		quake victims, raped woman	RS		6
BPU and CIF 2015	Georgia	1015	drug users: intravenous	CRS		24
Carletti 2017	Italy	299	number of people un-satisfied	CS	38.18	15
Ezoe 2012	Japan	1500	Men who have sex with men	CS	23	10
Feehan 2016	Rwanda	simulated 5000	Populations at risk for HIV/AIDS	TSC	97	22
Feehan 2016	Africa	4669	Female sex workers, Men who have sex with men, drug users: intravenous, male clients of female sex workers	STSC	97	22
Guo 2013	China	2957	Female sex workers, Men who have sex with men, male clients of female sex workers, drug users: intravenous, and others	MS	97.7	19
Habecker 2015	United States	550	People who have moved to Nebraska from another state in the U.S.	RS	27.5	18
Haghdoost 2015	Iran	3023	Cancer	CS	99	
Halimi 2020	Iran	461	drug users: Marijuana	MRS	92.2	6
Heydari 2019	Iran	2550	Methadone maintenance therapy users	MS	87	25
Jafari 2014	Iran	500	Female sex workers, drug users, male clients of female sex workers, men who have sex with men, alcohol users	PS		29
James 2013	UK and Southern Ireland	513	performance-enhancing drugs/substances	CS		
Jing 2014	China	319	Men who have sex with men	STSS	96.4	48
Jing 2018	China	7964	female sex worker	TSC	0.11	48
Kadushin 2006	United States	5892	drug users: heroin			6
Kanato 2015	Thailand	3790	Injecting drug users	SFSS		19
Kazemzadeh 2016	Iran	563	drug users, Friendship and close relations with the opposite sex, alcohol users	SS		Rastegari et al. 2013
Killworth 1998	United States	1524	Seroprevalence	RS		26
Killworth 1998	United States	1554	Seroprevalence, women raped, homelessness			29

Chapter 2

Maghsoudi 2014	Iran	839	drug users, female sex workers	CS		
Maghsoudi 2017	Iran	390	smoking: cigarette, hookah, drug users: tramadol, opium, extra/pre-marital sexual contact, alcohol users	CS		20
Maltiel 2015	Brazil	Salganik et al. 2011	HIV/AIDS positive	RS		29
McCormick 2010	United States	McCarty et al. 2001	HIV positive, women raped, homelessness	RS		29
McCormick 2012	United States	McCarty et al. 2001	HIV positive, women raped, homelessness			
McCormick 2015	United States	McCarty et al. 2001	HIV positive			
Mirzazadeh 2018	Iran	265	HIV risk behaviours	MS		
Mohebbi 2014	Iran	3023	living with a disability	RS	99	Rastegari et al. 2013
Moradinazar 2019	Iran	500	suicide deaths	RS	94.3	Rastegari et al. 2013
Motazedian 2020	Iran	86	network size			5
Narouee 2019	Iran	1000	drug users	RS		Rastegari et al. 2013
Nasiri 2019	Iran	200	drug users: tramadol	CS		
Nikfarjam 2016	Iran	7535	drug users: opium	RSS		Rastegari et al. 2013
Nikfarjam 2017	Iran	12293	alcohol users	MS		Rastegari et al. 2013
Rastegari 2013	Iran	7454	network size	MS		23
Rastegari 2014	Iran	12960	Abortion	MS		Rastegari et al. 2013
Sajjadi 2018	Iran	801	Students with high-risk behaviours	CS		
Salganik 2011	Brazil	500, 303, 6006	drug users	RS		
Salganik 2011	Brazil	294	drug users			
Scutelnicuic 2012	Kazakhstan		HIV risk behaviours			
Sharifi 2017	Iran	1337	female sex worker	RS		
Shati 2014	Iran	1029	Social network size	CS		13
Sheikhzadeh 2016	Iran	420	drug users, Alcohol users, Extra-marital sex	RSS	84	

Chapter 2

Sheikhzadeh 2016	Iran	500	alcohol users, extramarital sex, drug users: opium, intravenous	RS	84	5
Shokoohi 2010		500	network size	RS		6
Shokoohi 2012	Iran	500	drug users, males who have extra-marital sex with females, clients of female sex workers, alcohol users, men who had sex with men	APS		
Snidero 2007	Italy	72	choking injuries	CS		25
Snidero 2012	Italy	1081	choking injuries	RS		33
Sulaberidze 2016	Georgia	1015	Men who have sex with men	sTsc		
Sully 2020	Africa	6648	Abortion	TSC		
Teo 2019	Singapore	199	female sex workers, Men who have sex with men, drug users: intravenous, male clients of female sex workers			19
Vardanjani 2015	Iran	1995	Cancer	MCS	86.7	
Verdery			HIV positive			
Wang 2015	China	3907	Men who have sex with men	MS		
Yang 2017	United States		Christians and Buddhists	R		
Zahedi 2018	Iran	2157	smoking, drug users, alcohol users	MnRS	83.6	
Zahedi 2019	Iran	1730	extra/pre-marital sexual behaviors, drug users	CS	80.2	
Zamanian 2016		1275		MS		25
Zamanian 2016	Iran	1275	network size	MS		25
Zamanian 2018	Iran	1275	Abortion	MS		25
Zamanian 2019	Iran	1020	Abortion	MS	70.3	25
Zamanian 2019	Iran	2550	Abortion	RS		25
Zheng 2006	United States	McCarty et al. 2001	people in prison			29

Abbreviations: APS, Adaptive purposive sampling; CS, convenience sampling; MS, multistage sampling; MCS, multistage cluster sampling; MnRS, multistage non-random sampling; PS, purposive sampling; RS, random sampling; RSS, random stratified sampling; SFSS, stratified four stages sampling; SS, stratified sampling; STSC, Stratified two-stage cluster; TSC, two-stage cluster

2.3.3 Estimation of the network size in a different survey

The estimation of the network size in a separate study than the one used to estimate the hidden population has been performed in some studies. Shati [42] et al. estimated the social network size of Tehran inhabitants as 259.1 (CI95%: 242.2, 276); Motazedian [43], instead, estimated a social network size of 17 for children of Shiraz. The choice to calculate the network size has been considered mainly by a group of researchers from Iran. This approach could not necessarily be applied in all contexts. Even these authors reported different network sizes according to the diverse populations considered, with values that go from 234 people [20] in Kermanian women to 303 in Kermanian males.

Other authors, instead, estimated the whole network size at the country level; Rastegari et al. [44] the entire country defined at 308 people, Shokoohi [45] of the kermanian males estimating at 303 people, and Zamanian et al. [46] have estimated that kermanian women know about 234 people.

2.3.4 NSUM modifications

The most used revised version of the NSUM approach is proposed by Killworth et al. [47], called the maximum likelihood estimator (MLE). Habecker et al. [48] presented a modified version of the MLE to control for the increased variance using one or more known populations of small size. In this formulation, they choose a known population of similar size and adjust the mean of sums (MoS) estimator for survey characteristics. A generalized scale-up estimator (GNSUM) has been proposed by Feehan and Salganik [49], requiring additional data collected directly from the hard-to-reach population. The GNSUM has been implemented by Verdery et al. [50] in the venue-based generalized scale-up estimator (VBGNSUM), which requires a venue-based sampling. The Bayesian models' framework has also been used to control for biases; for example, the transmission bias added large overdispersion, so Zheng et al. [51] proposed the overdispersed model inquiring on the relationship between known subpopulations and the overdispersion. To account for recall bias, McCormick et al. [38] introduced latent non-random mixing, adding in another work the use of MCMC to estimate an unknown population [52]. Maltiel et al. [39] considered five models: the random degree model, the barrier effects model, the transmission bias model, the combined model, and the recall bias adjustment. Also, Teo et al. [53] account for transmission error and barrier effects in their two models that incorporate demographic characteristics as a coefficient. Through the years, has been an increase in the complexity of the models proposed. However, the NSUM in its simplest form is the most used,

as also reported in the work of Josephs et al. [54]. This is probably because guidelines report a guide on implementing a survey in key populations using the network scale-up approach based on the traditional NSUM estimator. Moreover, as McCormick et al. [38] suggested, if the survey is well-designed, then simple models can have a suitable level of accuracy compared to more complex ones.

To overcome transmission error, some authors have estimated the transmission rate of a subpopulation, defined as the visibility factor. So, the crude NSUM estimate is divided by the visibility factor. This factor estimates the proportion of respondent's aware of their behaviour [55]. Haghdoost et al. [56] reviewed the methods to assess VF in-network scale-up studies. The methods applied were the game of contacts, expert opinion, social respect, and the coming out rate.

2.3.5 Agreement among the different NSUM

The estimates obtained with different methods vary widely. In the study of Salganik et al. [41], the generalized NSUM estimator was compared with the multiplier methods without reaching a clear conclusion. The NSUM estimator produces a higher prevalence of 5-10 times compared to standard methods [41]. Even the various version of the NSUM provided estimates that are not similar. The study of Laga et al. [9] selected different models and applied them to the dataset of heavy drug users of McCarty [37]. In this simulation, even slight differences in the modeling could produce different estimates [9]. The accuracy of estimates is difficult to compare even among the NSUM. The limits of comparison are related to the fact that other methods may require additional information. For example, Laga et al. could not apply the model of e McCormick et al. [38] and McCormick and Zheng [52]. Those studies require data from another population. Also, the method of Verdery et al. [50] was not applied since it requires a different sampling.

Recently, to estimate the size of the hidden population, multiple methods to derive the best estimates have been proposed [7]. Okal et al. [20] estimate the size of people at risk for HIV infection using the WOTC method, multiplier method, literature review, and stakeholder consensus. In this review, we have also retrieved articles that, along with NSUM, use other indirect methods to have a more precise estimate or to overcome some biases. Jing et al. combined the randomized response technique (RRT) with NSUM to adjust for the response bias and then compared it to a multiplier method. In this case, the adjusted NSUM estimate was close to the estimate of the multiplier method and in line with official data [31].

2.3.6 NSUM use in official reports

The NSUM estimator was also used to produce reports by official organizations. The Rwanda biomedical center [57] technical report estimates the size of the key populations at higher risk of HIV infection, and one of the Bemoni Public Union (BPU) estimates the number of people who inject drugs in Georgia [58] are some examples.

2.4 Discussion

The NSUM, as evidenced in this literature review, has been widely used to determine the size of the hard-to-reach population for more than 30 years. The main advantage of this method relies on the fact that to obtain the estimates, there is no need to access the target population at the individual level. Moreover, the questionnaire can be structured in order not to ask directly to the respondents about their behaviours or opinions and, in a single survey is possible to estimate the size of multiple hidden populations.

Despite these advantages, the guidelines for estimating the size of populations most at risk of HIV by UNAIDS and WHO suggest using first the census, the capture-recapture technique, and the multiplier method rather than NSUM [6].

The NSUM limitations rely on assumptions typically violated in a real research setting. The first assumption is that the probability of knowing someone in a subpopulation is equal for all the responders (barrier effects bias); the NSUM moreover assumes the respondent is aware of the subpopulation belonging to each individual in the personal network (transmission bias); the third NSUM assumption relies on the fact that each responder can recall everyone in their network (recall bias) [9,37,47,59]. To overcome these biases, several variants of the NSUM have been proposed. These variants have been widely explained in the methodological review of Laga et al. [9].

There is no consensus on which one of the different modifications is the most suitable despite the presence of various estimation methods. Halimi et al. [29], for example, compared the NSUM and the proxy respondent method (PRM) and found that the frequency of PRM was closer to the real data, whereas the NSUM underestimated the phenomenon.

Determining the minimum number of questionnaires to collect to obtain reliable estimates in the hidden population investigated is a critical issue, as suggested in a recent work [54]. In our review, the size of the population considered in the studies that apply the NSUM varies from a few tens to

tens of thousands. Josephs et al. [54] proposed a sample size heuristic that can be applied in NSUM studies. They apply their method to published studies finding that the sample sizes of the studies considered are larger than the minimum retrieved using their sample size technique. Defining the sample size would help reduce the costs related to the interviews and survey duration.

2.5 Conclusion

This review offers a comprehensive overview of the network scale-up method regarding methodological aspects and its applications. The main issues that arise from this work are related to the several biases that still affect this method, given its strong assumptions. The attempt to overcome these biases has brought to the creation of models that had increased complexity in their structure, such as the one based on a Bayesian framework. However, it is still difficult to handle all the main biases together. Even the study that proposed to balance all the three main issues in the same model suggests accounting for recall of bias in a postprocessing phase [39].

Chapter 3.**Using Social Networks to Estimate the Number of COVID-19 Cases: The Incident (Hidden COVID-19 Cases Network Estimation) Study Protocol****3.1 Summary**

Recent literature has reported a considerable percentage of asymptomatic or paucisymptomatic cases in subjects with COVID-19 infection. This proportion can be difficult to quantify; therefore, it constitutes a hidden population. This study aims to develop a proof-of-concept method for estimating the number of undocumented infections of COVID-19. This is the protocol for the INCIDENT (Hidden COVID-19 Cases Network Estimation) study, an online, cross-sectional survey with snowball sampling based on the network scale-up method (NSUM). The original personal network size estimation method was based on a fixed-effects maximum likelihood estimator. We propose an extension of previous Bayesian estimation methods to estimate the unknown network size using the Markov chain Monte Carlo algorithm. On 6 May 2020, 1963 questionnaires were collected, 1703 were completed except for questions to define the network size (random question), and 1652 were completed in both demographics, target question, and network size question. The algorithm was initialized at the first iteration and applied to the whole dataset. Knowing the number of asymptomatic COVID-19 cases is extremely important for reducing the spread of the virus. Our approach reduces the number of questions posed. This allows us to speed up the completion of the questionnaire with a subsequent reduction in the nonresponse rate.

This chapter was published as:

Ocagli H, Azzolina D, Lorenzoni G, Gallipoli S, Martinato M, Acar AS, Berchialla P, Gregori D, On Behalf Of The Incident Study Group. Using Social Networks to Estimate the Number of COVID-19 Cases: The Incident (Hidden COVID-19 Cases Network Estimation) Study Protocol. *Int J Environ Res Public Health*. 2021 May 26;18(11):5713. doi: 10.3390/ijerph18115713. PMID: 34073448; PMCID:PMC8198250.

3.2 Introduction

Since December 2019, China and subsequently the whole world has been dealing with a pandemic due to a beta coronavirus related to the Middle East respiratory syndrome virus (MERS-CoV) and the severe acute respiratory syndrome virus (SARS-CoV2), named COVID-19 by the World Health Organization (WHO) [60]. The virus quickly spread globally [61,62]. In the Italian territory, the outbreak started with cases of pneumonia of unknown etiology at the end of January 2020.

Recent literature has highlighted a high percentage of undocumented cases among COVID-19-infected subjects. Such cases are mostly asymptomatic or paucisymptomatic, as their lack or scarcity of symptoms does not reach the attention of the healthcare system. Undocumented cases have been found to expose a higher proportion of the population due to the lack of quarantine measures [63] and to be hard to recognize, as asymptomatic or mildly symptomatic patients often do not seek medical attention due to a lack of symptoms [64]. While challenging, the prevalence estimation for asymptomatic and mildly symptomatic cases is fundamental given the highly contagious nature of the virus. Zou et al. [65] reported that the viral load in asymptomatic patients was similar to that in symptomatic carriers. Therefore, both asymptomatic and symptomatic patients may have the same transmissibility potential. Confirmed positive but asymptomatic people also need to be isolated to limit their contact with others. Consequently, accurate epidemiological monitoring of COVID-19 prevalence in asymptomatic people may further decrease viral contagion. Moreover, it will help properly distribute resources, and tailor the prevention program to the outbreak's containment [66].

Several studies have tried to reveal undocumented cases. For example, using a networked dynamic metapopulation model and a Bayesian inference in mobility data within China, Li et al. [62] estimated that 86% of all infections were undocumented (95% CI: 82–90%) before the 23 January 2020 travel restrictions. In this model, the authors considered both the spatial distribution of people and their mobility. This method was used mainly to describe epidemic outbreaks taking into account the connectivity among people. Mizumoto et al. [67], in their study conducted on the Diamond Princess cruise ship, showed that there was a considerable proportion of asymptomatic individuals among all infected cases, which was 17.9% (95% credible interval: 15.5–20.2%). Other estimates of undiagnosed patients with COVID were among the evacuated citizens. Nishiura et al. [64,66] estimated a proportion of 33.3% asymptomatic cases (95% CI: 8.3–58.3%) among Japanese citizens evacuated from Wuhan. Undocumented infections seem to facilitate the geographic spread of SARS-CoV-2 [62]. In the Veneto region, in the municipality of Vo', which was one of the initial outbreak sites in Italy, the choice to test the overall population helped to identify the proportion of

positive COVID-19, revealing 37.7% asymptomatic patients (95% CI 25.5–51.9%) [68]. Properly estimating the number of COVID-19 positive cases, even if asymptomatic, is important since person-to-person transmission can occur from asymptomatic COVID-19 cases to the community, as shown in previous studies [69,70].

Traditional (i.e., direct) methods to detect positive cases are based on the seroepidemiological testing procedures of potentially exposed or infected populations. Real-time PCR tests or other laboratory tests may identify asymptomatic infections [71]. However, these approaches are time-consuming and require considerable financial resources [72]. Other indirect sample estimation methods, instead, are not suitable since they feature limitations, such as the use of independent samples, direct access to the source of data, or data of each country, as in capture-recapture technique, multiplier method, synthetic estimation, and multivariate indicator methods, respectively [73,74].

Since undocumented infections are undefined, they constitute a hidden population. The NSUM, first proposed by Bernard et al. [75,76], is among the recommended methods to estimate the hidden population available in the literature. NSUM has been widely used to estimate the size of hard-to-reach populations due to the stigmatizing nature of its knowledge, such as HIV [77], injuries [78], men who have sex with men (MSM) [79], and others. It relies on the idea that the probability of knowing someone in a specific subpopulation is related to the relative size of that subpopulation, i.e., the proportion computed based on the population size of all individuals.

3.3 Materials and Methods

The Hidden COVID-19 Cases Network Estimation (INCIDENT) study aims to develop a proof-of-concept study for estimating the number of undocumented COVID-19 infections using a Bayesian approach to the traditional NSUM.

3.3.1 Study Design

This is a cross-sectional survey-design study to assess the prevalence of undocumented COVID-19 symptoms in Italy using an anonymous online questionnaire. The data were collected starting on 15 April 2020 and ending on the 1st of September 2020.

3.3.2 Procedures

To avoid unnecessary interactions, this study was structured only for electronic distribution, given the social distancing and the limitations imposed by the Italian government.

The questionnaire was created through LimeSurvey (LimeSurvey GmbH, Hamburg, Germany), a professional open-source online survey tool. The respondents had to be at least 16 years old and were required to sign an informed consent.

The study design mimicked snowball social network sampling but had a nonrandom entry point in the population. The snowball sampling method was chosen since it is widely applied and evaluated as particularly useful in studies that consider hidden populations as a target [80,81]. The researchers have defined the initial sample among their acquaintances, created a list of potential groups that work with the researchers, and sent the invitation to participate in the survey. Along with the invitation to participate in the study, they were asked to share it with their acquaintances. In each phase was ensured that study participants would not be identified. Each participant will continue to recruit others until the end of the survey. The questionnaire was advertised via social networks, mobile messaging systems, emails, and newspapers. There was no compensation for respondents.

3.3.3 Ethics

The INCIDENT study was approved by the Ethics Committee of the University of Torino, protocol No 458163, 10 October 2020.

3.3.4 The Network Scale-Up Method in the Literature

NSUM was first applied to estimate the death toll of the 1985 Mexico City earthquake [76] due to the missing reports regarding fatalities by the official registry. Since then, NSUM has been widely used to estimate the size of a subpopulation that consists of hard-to-identify individuals [82], such as individuals with high-risk behaviors that lead to stigmatization and discrimination, such as individuals living with HIV/AIDS [83–90], MSM [79,91], sex workers [82,92], drug addicts [92–97], or alcohol users [93,98]. NSUM has also been used to estimate the number of treatment failures [99],

people with disabilities [100], number of abortions [101,102], and suicide attempts [102]. Teo et al. [85] suggested using NSUM for estimating the hidden population to improve surveillance, prevention, and treatment after proper methodological adjustments.

Table 3 shows articles using NSUM to estimate hidden populations. The search was based on a review in PubMed (Figure 2).

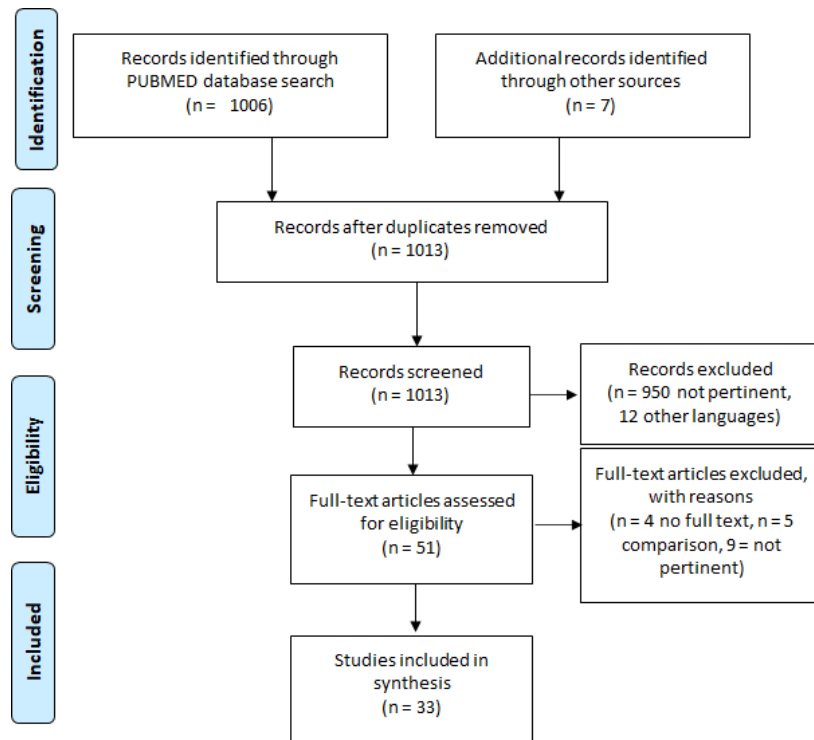


Figure 2 Flowchart of study review

In the Italian context, NSUM has been used to estimate the number of children with foreign body injuries [103] and to assess the perceived quality of care (PQC) in an oncological center. In the latter study, estimation of PQC was lower than in the traditional questionnaire; in some cases, the level of dissatisfaction was 20-fold higher [104]. Among other selected applications, Paniotto et al. [105] estimated the number of drug addicts, sex workers, and MSM in their study, showing that NSUM estimates were similar or lower compared to other estimation methods. Additionally, estimated populations of seropositive, homeless, and female victims of violence in Killworth et al. [83] were comparable to official data. One limitation of NSUM is that social and physical barriers, such as ethnicity, occupation, or location of residence, may influence the likelihood that respondents know people in hidden populations. This is known as the barrier effect.

On the other hand, individuals may not know everything about other people in their personal network. This instance, in which a contact does not share information with the respondent, is termed transmission bias. Other limitations in applying these methods include recall bias [88] and response

bias [106]. Several authors have tried to address such limitations, as shown in Table 2, McCormick et al. [107] adjusted NSUM for recall bias, various authors adjusted for barrier effects [56] and transmission errors [87–89,98,101,107,108], and Jing et al. [82] adjusted for response bias.

Table 2 Studies that estimate the hidden population using the network scale-up method or its modification.

Author	Year *	<i>N</i> Known Population	Hidden Population	<i>N</i> Respondents	Method	Adjustment	Place
Ahmadi [93]	2019	1	Drug/alcohol users before driving	363	NSUM		Iran
Bernard [76]	1991	6	Deaths in earthquake		NSUM		Mexico
Carletti [104]	2017	20	Oncological patients	299	NSUM		Italy
Ezoe [79]	2012	3	Men who have sex with men		NSUM	Transmission Error	Japan
Feehan [86]	2016	22	Populations at risk for HIV/AIDS	4669	Blended Scale-up		Rwanda
Guo [84]	2013	3	Populations at risk for HIV/AIDS	2957	NSUM		China
Habecker [109]	2015	18	Moved to Nebraska in US during last 2 years, do not approve of interracial dating, heroin users	618	Mean Of Sums NSUM		United States
Haghdooost [110]	2015	**	Population of breast, ovarian/cervical, prostate, and bladder cancers	3052	NSUM		Iran
Heydari [99]	2019	25	Treatment failure	2550	NSUM		Iran
Jafari [87]	2014	29	Populations at risk for HIV/AIDS	500	NSUM	Transmission Bias, Barrier Effects	Iran
Jing [82]	2018	48	Female sex worker		RRT, NSUM	Response Bias	China
Kadushin [97]	2006	3	Heroin users		NSUM		United States
Kazemzadeh [111]	2016	**	High-risk behaviors	563	CM, NSUM		Iran
Killworth [83]	1998	24	HIV prevalence, women who have been raped, the homeless	1554	NSUM		United States
Maghsoudi [92]	2014	20	Injection drug users, female sex workers	600	NSUM	Barrier Effect	Iran
Maltiel [88]	2015	29	Populations at risk for HIV/AIDS	500	Bayesian NSUM	Transmission Bias, Barrier Effects	Brazil

					Latent Non- Random Mixing Model NSUM	Transmission Bias, Barrier Effects, And Recall Bias.	
Mccormick [107]	2010	12	Personal network size	1370			Brazil
Mohebbi [100]	2014	**	People with disabilities	3052	NSUM		Iran
Moradinazar [112]	2019	**	Suicides and suicide attempts	500	NSUM		Iran
Narouee [94]	2019	**	Injection drug users	1000	NSUM	Barrier Effect	Iran
Narouee [113]	2020		Rural area	1000	MLE— NSUM		
Nikfarjam [98]	2017	**	Alcohol use	12,000	NSUM	Transmission Bias, Barrier Effects	Iran
Nikfarjam [95]	2016	**	Illicit drug users	7535	NSUM		Iran
Rastegari [101]	2014	**	Abortions	12,960	NSUM	Transmission Bias, Barrier Effects	Iran
Sajjadi [108]	2018	6	Students with high-risk behaviors	801	NSUM	Transmission Bias, Barrier Effects	Iran
Salganik [89]	2011	20	Populations at risk for HIV/AIDS		NSUM, GSU	Transmission Bias, Barrier Effects	Brazil
Shokoohi [114]	2010	6	Network	500	NSUM		Iran
Shokoohi [90]	2012	**	Populations at risk for HIV/AIDS	500	NSUM		Iran
Snidero [103]	2012	33	Foreign body injuries	1081	NSUM		Italy
Teo [85]	2019	24	Populations at risk for HIV/AIDS	199	Bayesian NSUM		Singapore
Vardanjani [115]	2015	**	Cancer	195	Generalized NSUM		Iran
Wang [91]	2015	22	Men who have sex with men	3097	NSUM		China
Zahedi [96]	2018	**	Drug users	2157	NSUM	Barrier Effect	Iran
Zamanian [46]	2016	25	Age-gender distribution of women	1275	NSUM		Iran
Zamanian [116]	2019	25	Abortion	1500	NSUM	Barrier Effect	Iran

N, Number; MLE, maximum likelihood estimation; NSUM = network scale-up method; RRT = randomized response technique; CM, crosswise model. * year of publication, ** Rastegari et al. [117].

3.3.5 NSUM Questionnaire

This questionnaire is structured into three sections: (1) four questions for the demographic characteristics of the respondent (gender, age, nationality, and region of residence), and (2) four target questions related to COVID-19 disease that were defined by consulting the available literature on COVID-19 [118], and (3) one question used for the estimation of social network size. The last question was randomly drawn from 15 known populations (see Supplementary for the full report of the questionnaire). The data source for the known population size is the Italian National Institute of Statistics (ISTAT) [119] (Table 3). So, each respondent will answer only one question to define its network size.

Table 3 List of known-size populations used in the questionnaire to estimate respondents' network sizes.

Sub-Population of Known Size	Population Size	Reference Year	Source
People who separated	99,611	2016	Demographic model
Foreign residents	5,255,503 ***	2019	Demographic model
Victims of car accidents with injuries	3334	2018	Demographic model
People who graduated	8530 **	2018	MIUR
People working part-time	3,689,153 ***	2019	Demographic model
Three-member families	4954	2019	AVQ
Cohabiting couples	14,110	2019	AVQ
People who married	195,778	2018	Demographic model
Children born	440,780	2018	Demographic model
People above 14 with smoking habits	10,122	2017	AVQ
People using the mass media (newspapers, magazines, TV, radio, etc.)	86,142 *	2017	AVQ
People who attend places of worship	14,264	2018	AVQ
People who walk to work	2750	2018	AVQ
People who go to school by bus	8743	2018	AVQ
Three-year-olds and above who used a PC and Internet	62,232	2017	AVQ

* Number watching TV 53953, radio 32189. ** <http://dati.ustat.miur.it/dataset/laureati/resource/43df861d-7345-481a-9803-2eb236aa022e> (accessed on 16 April 2020), (difference between 2018 and 2017, 326332-317802). *** cumulative data. MIUR, Ministry of Instruction, University and Research, <http://dati.ustat.miur.it/dataset/laureati> (accessed on 16 April 2020); AVQ, Aspects of daily life survey, <http://dati.istat.it/> (accessed on 16 April 2020).

All questions concerning the specific subpopulations were introduced with the sentence: “How many people do you know ...?”. In this study, we used the definition of knowing someone based on those provided by Bernard et al. [120] and already used in previous Italian studies [44,45].

3.4 NSUM Assumption

The NSUM estimation method, as explained by Bernard et al. [120] “rests on the assumption that people’s social networks are, on average, representative of the general population in which they live and move.” For example, suppose a responder knows 100 subjects on average and two of them are COVID-19 positive, we estimate a prevalence of 2/100 COVID-19-positive subjects considering the personal network size as a reference point. This estimated prevalence is combined with the known size of the general population to estimate the size of hard-to-reach populations, such as the COVID-19-infected population. The accuracy of the estimated size of the hidden population increases as the number of people who answer the question increases.

The NSUM, however, has some limitations. For example, people may not know all the characteristics of their network (i.e., a respondent may not know that a member of his or her network is affected by COVID-19.) This is called the transmission bias [107]. In addition, social and physical barriers, such as ethnicity, race, occupation, and location of residence, can cause variations in the probability that respondents know people in hidden populations; this is called the barrier effect [83]. Despite these biases, NSUM has two major advantages. First, this method does not ask the respondent for information on its characteristics. For example, stigmatized or hidden populations may be reluctant to disclose their status even in an anonymous survey [109]. Secondly, it is not necessary to directly interview the members of a hidden population. Still, the NSUM allows the use of considerably cheaper and easier-to-implement sampling techniques that use established sampling frames [97].

3.4.1 NSUM Estimator

The size of the general population is defined as T , while m_{ik} representing the number of subjects in the hidden population known by an individual i , in the subpopulation, k . c_i is the average size of the social network related to the individual i .

The scale-up estimator is based on the assumption that the number of subjects known to the respondent in the k – th subpopulation follows a binomial distribution [121] where

$$m_{ik} \sim \text{Binom}\left(c_i, \frac{T_k}{T}\right).$$

The scale-up estimator of the hidden population size is obtained by the following equation:

$$\hat{e}_k = T \frac{\sum_i m_{ik}}{\sum_i \hat{c}_i}$$

To estimate the size of the hidden population, we follow three steps:

(1) Estimate the average size of the personal network, c_i , by asking how many people the respondent personally knows about the k known populations (e.g., the number of people who were married in 2019). This number will then be divided by the number of people who got married in 2019 in Italy (T_k), where T_k is the total size of the subpopulation k .

$$\hat{c}_i = T \frac{\sum_{k=1}^{K-1} m_{ik}}{\sum_{k=1}^{K-1} T_k}$$

(2) Define the number of hidden COVID-19 cases present in each social network, for example, by asking the respondent how many people he/she knows with COVID-19.

(3) Calculate the COVID-19 population size obtained by multiplying the estimated proportions of the population in each subpopulation by the general Italian population. For example, if a respondent knows ten subjects with COVID-19 cases and has a personal network of 100 people and the total population is 1,000,000, the estimated number of hidden COVID-19 cases will be approximately calculated as $10/100 \times 1,000,000$.

3.4.2 Bayesian NSUM Estimation

Under maximum likelihood estimation, several known populations should be used to reduce the estimates' variance. This prolongs the time required to complete the questionnaire by increasing the likelihood of dropouts and nonresponses. Estimating the size of the network by considering the known population as partially unknown could be a solution to shorten the length of the survey. For this reason, the Bayesian estimation methods proposed by Maltiel et al. [88] will be extended to estimate the unknown network size using the Markov chain Monte Carlo (MCMC) algorithm.

Moreover, the original personal network size-estimation method proposed by Killworth and colleagues [121] was based on fixed-effects, maximum likelihood estimators in which the network size was considered a nonrandom component. Instead, Maltiel et al. [88] extended this approach in a Bayesian setting by treating personal network sizes as random variables. This allows us to generalize the model to account for the variation in respondents' propensity to know people in particular

subgroups (barrier effects), such as their tendency to know people like themselves, and their lack of awareness of recognizing their contacts' group memberships (transmission bias).

3.4.3 Extended Random Degree Model

The NSUM formulation proposed by Maltiel et al. [88] assumes that the estimate of an individual's network degree, c_i , improves if a respondent knows a considerable number of subjects in a subpopulation. Network estimation is embedded into a Bayesian hierarchical modeling framework where the lognormal distribution best fits the network estimates across multiple datasets [88].

$$\begin{aligned} m_{ik} &\sim \text{Binom}\left(c_i, \frac{T_k}{T}\right) \\ c_i &\sim \text{log Normal}(\mu, \sigma^2) \end{aligned}$$

m_{ik} values are the number of subjects that the i -th subject knew in the k -th subpopulation. Observed m_{ik} values are assumed to be a realization of a binomial random variable whose parameters are defined by the personal network degree (c_i) and the overall known proportion of subjects in the subpopulations $\left(\frac{T_k}{T}\right)$, where T_k is the size of the k -th known sub-population and T is the overall population size.

The parameters of the random degree will be estimated in a Bayesian manner using the uninformative priors $\pi(T_k)$ for the k -th subpopulation, as in the Maltiel et al. [88] work:

$$\begin{aligned} \pi(T_k) &\propto \frac{1}{T_k} 1_{T_k \leq T} \\ \mu &\sim U(3, 8) \\ \sigma &\sim U\left(\frac{1}{4}, 2\right) \end{aligned}$$

T_k priors (hidden population parameter) have been used in the literature for Bayesian estimation of population size with vague prior [88]. The μ and σ prior distributions were derived by Maltiel et al. [88], identifying the best fit to the network estimates across multiple datasets.

In our NSUM formulation, the number of subjects known by the respondent in the k -th subpopulation is unknown, except for the target question that identifies the hard-to-reach subpopulation and the question drawn of the known subpopulation extracted for each respondent from the list of the known subpopulation.

The model is reformulated by assuming m_{ik} partially unknown. The number of subjects that the i -th respondent knows in the k -th subpopulation for each MCMC iteration will be drawn from a binomial random variable (m_{ik}) except for the target question that identifies the hard-to-reach subpopulation and the known population.

3.4.4 Performance of the Modified Maltiel Estimators

The algorithm was initialized at the first iteration from the first resampled m_{ik} values defining the starting values using the Killworth method [121].

The performance of the proposed NSUM estimator was evaluated in a simulation study considering different study size scenarios (1000, 1500, 2000, 2500, 3000).

Answers were generated by assuming the sizes of known subpopulations as indicated in the McCarty NSUM study [121] for each scenario. Data were generated using the original `nsum.simulate()` function proposed by McCarty in the NSUM R package [122]. The data were generated 300 times. For each replication, Maltiel's NSUM model and our NSUM proposal were estimated with 500 iterations, discarding the first 50 (burn-in iterations).

The total Italian population in January 2020 was 60,244,639 [119]. The assumed prevalence of COVID-19-positive cases is 1.37% (approximately 800,000 cases, like the peak of 805,947 officially diagnosed positive cases on 22 November 2020).

To compute the modified version of Maltiel's short-form questionnaire for each responder, we randomly sampled one of the responses characterizing the known populations from the full generated database, leaving the other responses missing.

The 95% credible interval (CrI) was computed.

The performance of the estimators was computed by calculating the average CI over simulations with the average prevalence estimate and the bias for each study size.

Computations were performed using R 3.6.2 [123] software.

3.4.5 Statistical Analysis

Data will be presented in aggregate form, and it will not be possible to trace information or make comparisons on an individual level.

Continuous variables will be summarized using the median (quartiles I and III), and qualitative variables will be synthesized using percentages and absolute numbers. Comparisons between groups will be identified using the Wilcoxon–Kruskal–Wallis test for continuous variables and the Pearson chi-square test for categorical variables.

3.5 Results of the Simulation Study

A total of 1963 questionnaires were collected until 6 May 2020, 1703 were completed except for the question to define the network size (random question), and 1652 were completed in all three sections. The respondents were primarily female (1206, 61%), and the prevalent residents were the areas most affected by COVID-19, including Veneto, Lombardia, and Piemonte.

3.5.1 Performances of the Modified Maltiel's Estimators

The estimated prevalence remains approximately constant as the sample size increases for both Maltiel's method (close to 1.37%) and the modified version (close to 1.57%). The bias remains approximately -0.05% and 0.2% for the original and modified versions of the method, respectively. The length of the interval decreases as the sample size increases. In the simulation study, a sample size of 1000 to 3000 respondents was considered. A sample size of 2000 subjects guarantees a CrI length of 0.044% for the modified method (bias = 1.197) and 0.038 (bias = -0.051) for the original approach (Table 4).

Table 4 Simulation results for the modified Maltiel's method performance.

Study Size	Maltiel's Method			Modified Maltiel's Method			
	Benchmark Prevalence	Prevalence%	95% CI Length	Bias	Prevalence%	95% CI Length	Bias
1000	1.37	1.324	0.056	-0.056	1.561	0.074	0.191
1500	1.37	1.327	0.044	-0.053	1.57	0.056	0.200
2000	1.37	1.329	0.038	-0.051	1.567	0.044	0.197
2500	1.37	1.329	0.031	-0.051	1.566	0.038	0.196
3000	1.37	1.33	0.025	-0.05	1.574	0.032	0.204

For each study size, the average prevalence (%) and the 95% CI were reported for Maltiel's method and the modified version. The benchmark prevalence represents the true value used to generate the data. The bias (average estimated prevalence—benchmark) was reported for each method.

The modified version of the method demonstrates a slightly higher bias than the original method; however, prevalence estimates are similar between the two methods.

3.6 Discussion

The NSUM, as modified in this study, has numerous advantages. First, questions for the scale-up social network estimator can be easily integrated into the survey. As shown in our literature review, most of the time needed to complete the questionnaire is used to estimate the size of the social network [85,87,89,91,99,102–104,109]. Some authors have preferred to estimate the size of the active network population only once, as in a previous Iranian population study [117], to reduce the time for survey completion.

In our study, we posed questions only related to the hidden population [90,94–96,98,100,101,112,115]. Few studies have chosen to use questions related to known populations [17,20,25,26]. Ahmadi et al. [93] proposed only one known population, as in our work. However, with traditional NSUM, our questionnaire would have needed 20 known populations [5,10,47]. With our method, only one question is needed. This allows to speed up the completion of the questionnaire by reducing the dropout and nonresponse rate.

Moreover, the modified NSUM allows estimation of the personal network size independently from the survey on COVID-19. This is possible because data for the estimation of the size of the social network could be collected on a separate subpopulation from the general study.

Direct methods for estimating undocumented cases are resource and time-consuming due to the requirement of large-scale testing procedures. Viable alternatives that may overcome such limitations are the introduction of a statistically robust sampling design for estimating the parameters of the epidemics, such as the one proposed by Alleva et al. [124]. Other methodological statistical studies have tried to estimate the true number of COVID-19-infected people. Palatella et al. [63] attempted to indirectly estimate the number of cases based on PCR test alone and Noh and Danuser [125] based on daily government counts of confirmed cases. However, these estimates could underestimate the real proportion since the data are related to a specific population.

In literature, there are different indirect sample estimation methods, but those are not suitable for the purpose of our study. For example, the capture-recapture technique requires at least two independent and representative samples [73]. Multiplier and enumeration methods, on the other hand, require direct access to the target population [74].

3.7 Conclusions

NSUM showed its advantages when we utilized it for the same survey in different contexts, as in our study with mimicked snowball sampling. Therefore, data for estimating the social network size can be collected in a standardized way, both spatially and temporally. This would be extremely difficult if the methods instead required direct access to the hidden population as the target of the investigation. Despite the bias of the traditional NSUM, our modified version shows advantages, considering that only one question was used for defining the social network.

The NSUM can help trace the undocumented cases that could be known in the people's social networks but do not draw the attention of the healthcare system. Our method would be useful when government testing is not widespread enough to reach the overall population, such as with COVID-19.

Chapter 4.**Indirect estimation of COVID-19 asymptomatic cases using a shortened Network Scale-Up approach. An application to Italy in the First Wave****4.1 Summary**

Asymptomatic COVID-19 cases carriers play a relevant role in the widespread infection during the coronavirus disease (COVID-19) global pandemic. Estimating the prevalence of undocumented cases of COVID-19 has been a significant public health issue since the beginning of the pandemic. This study aims to estimate the proportion of undocumented cases of COVID-19 in three Italian regions using a modification of the NSUM method.

This is a cross-sectional survey design study with social networks sampling. The questionnaires were collected between 15 April 2020 and 6 May 2020. To estimate the prevalence of undocumented cases of COVID-19, the number of cases of COVID-19, quarantined people, and the number of people that transferred among regions after the first Law Act (Decreto del Presidente del Consiglio dei Ministri, DPCM 25th March 2020) from the Italian Government during the first wave of the pandemic we applied the models proposed by Maltiel: the random degree model, the barrier effects model, and the transmission bias model. The analysis has been performed by assuming several scenarios on the average network degree size on the log-Normal scale.

The respondents were 1484, with a median age of 39 years, 895 (40 %) were female, and 757 (38.56 %) were male. The random degree model, compared to the models that account for barrier effect and transmission bias, has a lower difference from official data on the number of COVID-19 positive in all the regions considered. These differences are 0.03, 0.02, and 2.2, respectively in Veneto, Piemonte, and Lombardia. Estimated undocumented cases are higher in Lombardia compared to Piemonte and Veneto, respectively 2.78%, 0.44%, and 0.24% of the population of each region, according to the random degree model.

Despite a gold standard method for detecting the size of undocumented cases, such as mass testing methods, the use of an indirect method could still help define the prevalence of a hard-to-reach phenomenon. This was true, especially in the first period of the pandemic, when the testing procedure was not widespread among the population.

4.2 Introduction

Asymptomatic carriers play a relevant role in the widespread infection during the global pandemic of coronavirus disease (COVID-19) [126]. They were found to have the same viral load of symptomatic carriers [127] and were initially supposed and then confirmed to have the same transmissibility rate as the latter [128]. Asymptomatic carriers of COVID-19 are those who do not present any symptoms after the development of the infection and, especially in the incubation period, go undetected. Due to the lack of symptoms do not know their infectious status [128]. A recent systematic review found that the proportion of asymptomatics was high and ranged from 14.3% to 84% in different populations [129]. Asymptomatic carriers in the literature were differentiated between those who are initially asymptomatic (presymptomatic) and those who remain asymptomatic [130]. Buitrago-Garcia et al., in their meta-analysis [130], reported that 20% (95% confidence interval [CI] 17%-25%), whereas the prediction interval, which is the uncertainty on a single observation estimated from the population, ranged from 3%-67% of infected stay asymptomatic also in the follow-up period. Asymptomatic and paucisymptomatic people often remain undocumented since they do not come to the attention of the surveillance systems. Estimating the prevalence of undocumented cases of COVID-19 has been a significant problem faced by public health since the beginning of the pandemic. Hence, identifying and consequently isolating asymptomatic people was the key to helping in controlling virus spread, as reported in the Italian experience [131].

Various studies have tried to estimate the proportion of undetected cases. The proportion of asymptomatic cases is unclear, and the statistics and infectiousness vary between different countries as reported, respectively, in the scoping review of McEvoy et al. [132] and the work of Chen et al. [133]. According to a networked dynamic metapopulation model, approximately 86% of undocumented infections within 375 Chinese cities. The model is divided into two classes: documented infected and undocumented infected individuals [63]. The same applies to Li et al. before the travel restrictions applied in Wuhan dropped to 35% between 24 January 2020 and 8 February 2020 [134]. In the municipality of Vo, where the first deaths of COVID-19 in Italy were reported, asymptomatic during the survey period was 42.5% (95% CI: 31.5-54.6%) of the confirmed cases [36].

A systematic review in the English context has shown that testing according to symptoms is not enough as a population-wide control strategy against the diffusion of the virus [135].

In this work, we propose modifying a commonly used indirect estimation method by assuming undetected cases of COVID-19 as a hidden population. The method proposed in this study is based on Bernard's Network Scale-Up Method (NSUM) [5] and its subsequent modification in a Bayesian framework [39]. In this approach, the size of a target population is estimated based on the number of people known by each respondent. The method is mainly used in studies that aim to estimate the 'hidden' or 'hard to reach' population, which is generally not part of official data due to geographical or social situation ("hard-to-reach") [3] or lack of will to be found ("hidden") [4]. Considering COVID-19 asymptomatic people as a hidden population, especially in the first wave of pandemics, NSUM has the advantage of not requiring direct access to the target population and official data.

This study aims to estimate the proportion of undocumented cases of COVID-19 in three Italian regions, Veneto, Piemonte, and Emilia Romagna, using a modification of the NSUM method. Secondary outcomes of the study are the estimation of the number of cases of COVID-19, quarantined people, and transfers among regions after the first Law Act from the Italian Government during the first wave of the pandemic.

4.3 Materials and methods

4.3.1 Study design

This is a cross-sectional survey design study with social networks sampling. The questionnaires were collected between 15 April 2020 and 6 May 2020. The study design, procedures, questionnaire, and assumptions of the NSUM and its estimator are extensively described in the current study protocol [136]. S2 reports the full questionnaire.

4.3.2 Model Estimation

Let e be the estimated size of the hidden population, T the size of the general population, m_i the number of people in the hidden population known by a person i is the individual, \hat{e}_i is the size of the social network for the respondent i . The NSUM estimator, as proposed by Bernard, could be defined by the following equation. The original NSUM estimator was based on a maximum likelihood estimation that m_{ik} assumed m_{ik} as a non-random component [10].

$$\hat{e} = T \frac{\sum_i m_i}{\sum_i \hat{c}_i},$$

This formulation could be defined as follows, where \hat{e}_k is the estimated size of the hidden population k and m_{ik} is the number of people in the hidden population k known by an individual i .

$$\hat{e}_k = T \frac{\sum_i m_{ik}}{\sum_i \hat{c}_i},$$

The hidden k populations, k , considered in this study are undocumented cases of COVID-19, cases of COVID-19, quarantined people, and the number of transfers among regions.

To estimate the prevalence of the hidden population of interest, the random degree model, the barrier effects model, and the transmission bias model proposed by Maltiel [39] were applied.

4.3.3 Random-degree (RD) model

The Bayesian RD model treated the m_{ik} parameter as a binomial random variable with the following parameters as suggested in the work of Killworth et al. [47] in a Bayesian framework [39] to regularize the number of people known by the respondent estimates, also called their degree or personal network size, in the random degree model.

$$m_{ik} \sim B \text{inom} \left(c_i, \frac{T_k}{T} \right)$$

m_{ik} was iteratively resampled 500 times, discarding the first 50 (burn-in iterations), during the MCMC estimation procedure from a binomial random variable whose parameters are characterized by the degree of personal network size degree (c_i) and the overall known proportion of subjects in the subpopulation $\left(\frac{T_k}{T} \right)$.

A log-normal distribution was chosen by Maltiel et al. [39], since it is the best fit to estimate the scale-up estimates I , whose parameters are determined by the authors investigating multiple datasets.

$$c_i \sim \text{logNorm}(\mu, \sigma^2)$$

Random degree model parameters have been estimated in a Bayesian approach by considering the prior $\pi(T_k)$ for the k -th subpopulation, as in Maltiel et al. [88] work:

$$\pi(T_{ik}) \propto \frac{1}{T_{ik}} 1_{T_k} \leq T$$

$$\begin{aligned}\mu &\sim U(3,8) \\ \sigma &\sim U\left(\frac{1}{4}, 2\right)\end{aligned}$$

The μ and σ prior hyperparameters were derived by Maltiel et al. [39] by identifying the best fit to the network estimates across multiple datasets.

Our previous work [136] reported the performances of our proposed model in a simulation study. The estimated prevalences were constant for both models (close to 1.37 in Maltiel's method and 1.57% in our model).

4.3.4 Barrier effect (BE) model.

The BE model assumes that the survey respondents may have a different propensity to meet people from groups other than their own [39]. In this theoretical framework, the probability that a survey respondent knows a subject in the subpopulation is assumed as a random component modeled by a Beta random variables s_{ik} . In this case, the model parameterization is as follows:

$$\begin{aligned}m_{ik} &\sim \text{Binom}(c_i, s_{ik}) \\ c_i &\sim \text{logNormal}(\mu, \sigma^2) \\ s_{ik} &\sim \text{Beta}(l_k, r_k)\end{aligned}$$

The l_k value represents the mean of the Beta random variables s_{ik} and r_k the dispersion.

4.3.5 Transmission bias (TB) model

The TB model assumes that a person may be unaware of or avoid recognizing their membership or acquaintance in the target population [39]. τ_k the transmission bias is considered as the proportion of people known by the respondents in the k population reported and is distributed in the model as a random variable Beta. The model multiplies this transmission bias effect by the binomial parameter denoted by the fraction of subjects known by the responder i in the k subpopulation ($\frac{T_{ik}}{T}$).

$$\begin{aligned}m_{ik} &\sim \text{Binom}\left(c_i, \tau_k \frac{T_{ik}}{T}\right) \\ c_i &\sim \text{logNormal}(\mu, \sigma^2). \\ \tau_k &\sim \text{Beta}(\eta_K, \nu_K)\end{aligned}$$

As reported by Maltiel et al. [39]. The τ_k value is equal to 1 for the $K-1$ known sub-population $K-1$, while it is modeled as a random beta component with shape and scale parameters η_K, ν_K for the k -th unknown subpopulation. It is assumed that the respondent could be aware of the

grouping membership of their network concerning the known subpopulation, while, for the unknown population, often characterized by a hidden or stigmatizing connotation, the respondent could not have full awareness; this perception bias τ_k is translated into the translation bias component for the unknown subpopulation k .

4.3.6 Maltiel model modifications

The modifications proposed by our approach to the original formulation of NSUM are the following: i) we considered more than one hidden population (k), ii) m is resampled from a binomial random variable and is stratified according to gender, age, and region, iii) and the \hat{c}_i component is partially unknown. By using this approach, the survey size reduces down to only nine items. On the contrary, having applied the traditional NSUM approach, even in a Bayesian framework, the minimum number of questions would have been 28, 20 related to known population and 8 of the questionnaire, as recommended by Bernard et al. [5,10,47].

4.3.6.1 Sensitivity to the prior choices

Analysis has been performed by assuming several scenarios on the average degree size of the network on the log-normal scale.

- 1) Default Prior was defined vaguely as indicated in the Maltiel et al. [39] approach by considering a uniform $\mu \sim U(3,8)$ prior.
- 2) Other weakly informative priors have been defined by setting a normal random variable defined as $\mu \sim Normal(v, 100)$. The v values range among the values suggested by Maltiel [39] from 3 to 8.

The priors defined for the transmission bias and barrier effect models are kept as in the original Maltiel formulation [39], equal and uninformative.

4.3.7 Data Preparation

The data considered in the models are those of the Veneto, Piemonte, and Lombardia regions. Other regions were excluded due to the small size of collected data. Lombardia was included, although a few questionnaires were collected, given its significance in the initial outbreak of the COVID-19 pandemic in Italy.

4.3.8 *Known subpopulation*

The known subpopulations were derived by census [119,137] and are reported in Table 3.

4.3.9 *Personal network size estimation*

The size of the personal network for each subject has been estimated by associating the known subpopulation with the respondent according to the individual demographic characteristics. For example, the network size for a female less than 65 years of age in the Veneto region has been defined as $T_{ik}T_{ik}$, the number of females living in Veneto having an age less than 65 years and not the overall Italian population.

4.4 Results

4.4.1 *Descriptive characteristics of the sample*

Table 5 reports the statistics of all the questions. The number of answers to the known populations is inferior to the overall number of responses since, for each respondent, the known population question was only one and was randomly selected among 15 populations.

The questionnaire was shared on social media platforms, e.g., Facebook, WhatsApp, and mailing lists. The respondents were overall 1963, for the analysis were considered 1484 questionnaires due to missing data in one or more questions. Respondents have an overall median age of 39 years, 895 (60 %) were female, and 589 (40 %) were male. The median age in the tree region is 39 years in Veneto, 31 in Lombardia, and 41 in Piemonte. The respondents were mainly from the Veneto (711, 36%) and Piemonte (657, 33%) regions. Those regions, along with Lombardia, were the first regions where the virus spread. The other Italian regions were involved later in the pandemic. The complete response rate was 76% (1484 complete answers, 1963 total respondents). The dropout rate (people who logged in to the survey platform but refused to provide a significant set of answers) was 24%.

Table 5 Descriptive statistics of the study sample according to the region. Continuous data are reported as median (I, III quartiles); categorical data are reported as a percentage and absolute frequencies. In the table are reported demographic characteristics, respondents' answers to the target population, and the known populations.

Variable	N	Veneto (N=711)	Lombardia (N=116)	Piemonte (N=657)	Overall (N=1484)
Demographic characteristics					
Age	1484	28.00/39.00/52.00	25.00/31.00/45.25	29.00/41.00/52.00	28.00/39.00/52.00
Gender: Female	1484	58% (411)	56% (65)	64% (419)	60% (895)
Target questions					
N paucisintomatic COVID-19	1434	0.00/0.00/2.00	0.00/1.00/3.25	0.00/1.00/3.00	0.00/1.00/2.00
N swab COVID-19 positive	1413	0/1/2	0/1/3	0/1/3	0/1/3
N cohabitant of isolated COVID-19	1387	0/0/1	0/1/3	0/1/2	0/0/2
N transfer after DPCM*	1373	0/0/1	0/0/2	0/0/1	0/0/1
Known population					
People who separated	76	0.00/0.00/1.00	0.75/1.50/2.50	0.00/1.00/2.00	0.00/0.50/2.00
Foreign residents	92	0.00/ 3.00/10.25	1.00/ 2.00/ 5.00	2.00/ 6.00/17.50	1.00/ 4.00/12.75
Victims of a car accident with injuries	94	0.75/1.00/3.00	0.50/2.00/4.50	0.50/2.00/4.00	0.25/1.50/4.00
People who graduated	84	0.00/1.00/3.00	0.75/2.00/6.25	0.50/2.00/5.00	0.00/2.00/5.00
People working part-time	91	1.00/ 3.00/ 7.25	2.00/ 3.50/16.25	1.00/ 3.00/ 5.50	1.00/ 3.00/ 7.00
3-member families	80	3.00/10.00/20.00	4.75/ 8.00/12.50	5.00/15.00/22.50	4.00/10.00/20.00
Cohabiting couples	87	4.00/10.00/20.00	11.50/20.00/20.00	6.25/11.00/37.50	5.00/10.00/30.00
Children born	103	0/1/2	0/1/1	0/1/2	0/1/2
People who attend places of worship	100	3/ 15/100	5/ 10/150	02/10/2020	03/10/1950
People with smoking habits	88	10.00/23.50/50.00	10.25/20.00/42.50	9.25/15.00/20.00	10.00/18.50/40.00
People who walk to work	85	0.0/0.0/3.0	0.0/0.0/0.0	1.0/2.0/9.5	0.0/1.0/5.0
People who go to school by bus	80	0.00/ 2.00/ 7.00	0.75/ 8.00/17.75	0.00/ 2.00/10.00	0.00/ 2.00/10.00

Chapter 4

People who married	92	0.00/2.00/2.00	0.00/1.00/2.75	0.00/1.00/2.00	0.00/1.00/2.00
People using the mass media	92	30/ 50/150	35/100/125	20/ 50/125	25/ 50/150
People who used a PC and Internet	78	0.0/ 1.0/ 6.0	2.0/ 3.0/ 4.5	1.0/ 3.5/10.0	0.0/ 3.0/10.0

*Decreto del Presidente del Consiglio dei Ministri, DPCM 25th March 2020

Table 5 also reports the statistics of the four target questions and known populations considered. In all the three regions, respondents reported knowing one person tested positive for COVID-19, one paucisymptomatic for COVID-19 in Lombardia and Piemonte, and zero in Veneto.

Estimates of hidden populations

The number of cases of COVID-19, of quarantined people, and the number of transfers among regions' prevalence were calculated according to three models considered by Maltiel et al. [39], random degree model, transmission bias model, and barrier effect model. For each estimate is reported the 95% credible interval (CrI) to capture the uncertainty in the location of the estimated value by assuming a uniform prior (3,8) on the network size and normal prior with the mean previously defined by Maltiel et al. [39], and variance equal to 100. The estimates were computed separately for the three regions due to their different approach to facing the pandemic in the first wave. The estimates, both in tables and in graphics, were computed with different dimensions of network size, μ ranging from 20 to 3,000, which is consistent with previous research on social networks and the NSUM [39]. Table 6 reports the estimates of the four target questions with different μ for each region considering the random degree model. The estimates reported in Table 6 show that they are similar considering different network sizes except for default μ and μ with larger (network size of 3000) and lower network sizes (network size of 20). The table could be interpreted as follow: in Lombardia, there is a 95% probability that the estimated value, 196599 cases of COVID-19, lies within 176432 and 222017 given the observed data (Table 6). The credible interval of Lombardia is larger compared to the other regions, which are respectively 11542 (95%CrI 10902-11600) in Veneto and 18335 (95%CrI 18027.71-18825.08) in Piemonte. This could be related to the fact that fewer questionnaires were collected for Lombardia.

Table 6 Estimates of the four target questions in Veneto, Piemonte, and Lombardia regions between 15 April 2020 and 6 May 2020. The estimates are reported for network size, μ goes from 3 to 8, which means a network size ranging from 20 to 3000. For each estimate, the posterior median, lower and upper bound for the 95% Credible Interval are reported.

		Veneto			Piemonte			Lombardia		
μ		Posterior median	Lower Bound	Upper Bound	Posterior median	Lower Bound	Upper Bound	Posterior median	Lower Bound	Upper Bound
Cohabitant of isolated COVID-19	3	22173	21154	22914	5951	5450	6344	117958	105767	142251
	Default	13192	13186	14384	19092	18423	19941	176149	165347	195840
	4	31884	31727	32851	12684	12561	15553	194862	176035	211572
	5	32411	31258	34537	12991	11781	13955	197409	186862	218783
	6	31895	30490	32336	12648	11924	14156	189128	167371	225001
	7	32331	30846	32331	12500	12285	13620	199886	181782	230668
	8	32291	31681	32986	12950	124556	13126	200453	181049	234153
Undocumented COVID-19 cases	3	31725	31405	32895	12654	12294	13297	194651	169770	232496
	Default	11451	10837	12468	18695	17974	19339	277132	252555	279693
	4	11580	10601	12288	18025	17152	19259	274273	260158	288069
	5	11577	11181	12387	18255	17710	18720	265877	255915	285173
	6	11385	11133	11833	18154	17833	18705	277824	265165	304146
	7	11621	10903	11835	17377	17377	19167	261914	255746	287323
	8	11577	10712	11961	17971	17267	20333	272026	244755	286921
Swab covid positive	3	32401	31390	32416	13029	12509	13785	196599	176432	222017
	Default	11542	10902	11600	18335	18028	18825	267067	246694	284537
	4	13547	13435	14163	19215	18676	19727	180796	173446	195074
	5	13705	13272	14130	19138	18927	19609	178609	156149	207736
	6	13878	13159	15267	19265	18463	21202	177297	165553	193672
	7	13972	13556	14385	19537	18871	19726	180969	164401	200029
	8	13728	12545	14706	19617	18909	19929	177847	162373	206540
N of transfer after DPCM*	3	21690	21036	21818	5841	5322	6205	118520	105443	148311
	Default	13479	13136	17304	18834	18701	19995	181193	149829	197897
	4	21580	20633	22015	5896	5485	6376	127780	104806	148168

Chapter 4

5	21526	20939	21994	5927	5377	6136	124418	106015	146035
6	21525	21068	22139	5759	5382	6117	117491	104350	137243
7	21185	20958	22300	5911	5195	6252	117096	105176	130271
8	20802	20604	22332	5786	5553	6240	119112	98688	131728

*Decreto del Presidente del Consiglio dei Ministri, DPCM 25th March 2020

Figure 3, Figure 4. and, Figure 5 show the estimates for the four hidden populations according to the region considered for the random degree model, transmission bias model, and barrier effect model, respectively.

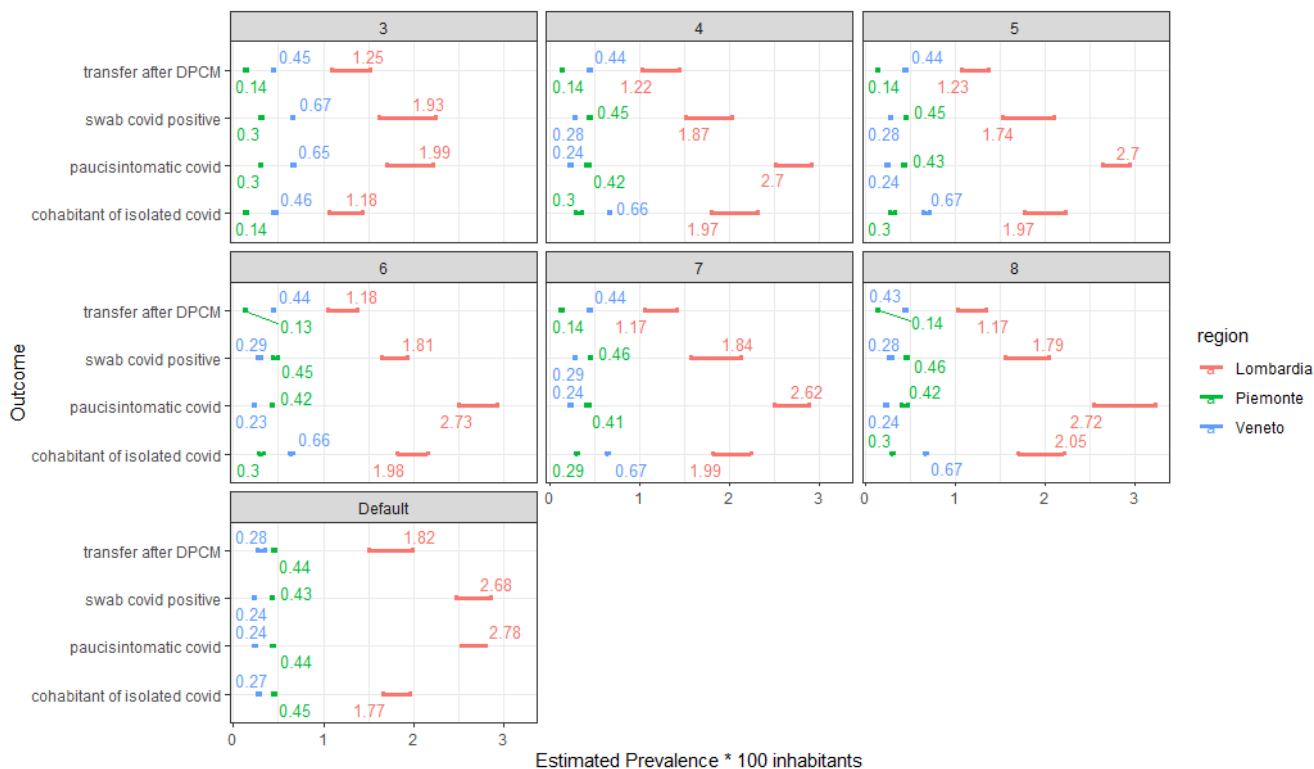


Figure 3 Estimated prevalence of the four target questions in the Veneto, Piemonte, and Lombardia regions for μ that goes from 3 to 8 (20 to 30000 network size) with the random degree model. Estimates are reported over 100 inhabitants.

Chapter 4

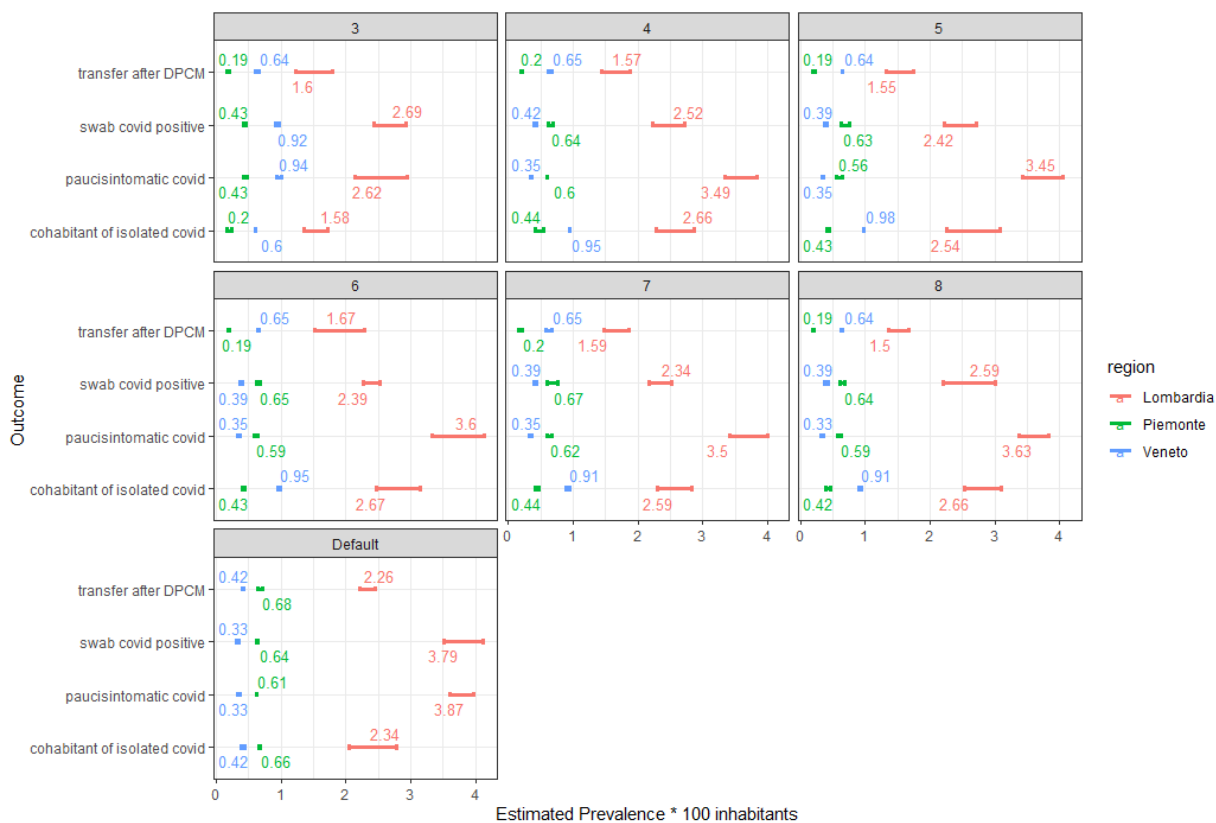


Figure 4 Estimated prevalence of the four target questions in the Veneto, Piemonte, and Lombardia regions for μ that goes from 3 to 8 according to the transmission bias model. Estimates are reported over 100 inhabitants.

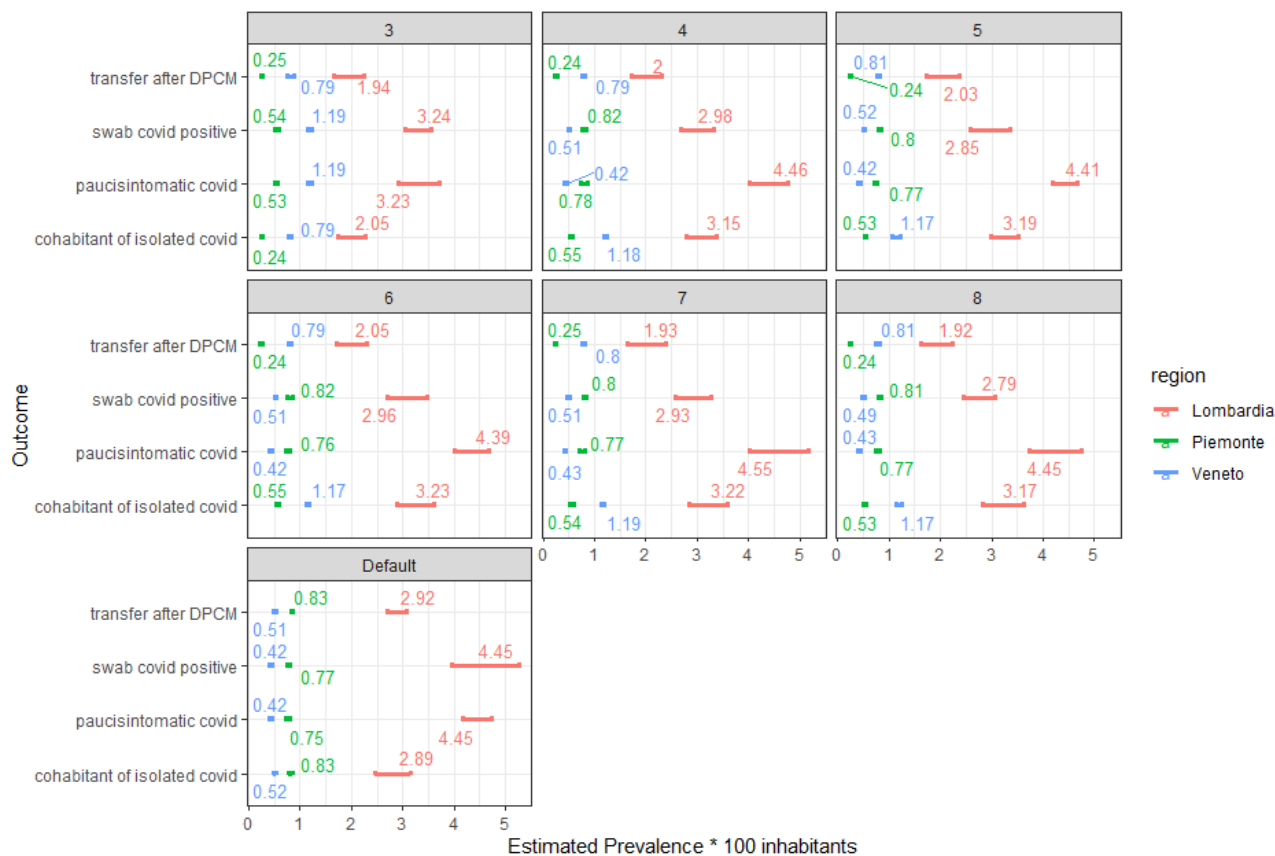


Figure 5 Estimated prevalence of the four target questions in the Veneto, Piemonte and Lombardia regions for μ that goes from 3 to 8 according to the barrier effect model. Estimates are reported over 100 inhabitants.

4.4.2 Estimates of COVID-19 cases

Among the hidden population considered in this survey, we have also included the number of COVID-19 cases. This choice will help us to evaluate how our estimates are close to official data.

Estimates of COVID-19 positive are higher in the Lombardia region for all the three models considered 2.68%, 3.79%, and 4.45%, respectively, with the default μ . Figure 6 reported the average and the trends of the prevalence of COVID-19 for each region according to official data. Estimates of COVID-19 positive are higher in Lombardia compared to Veneto and Piemonte.

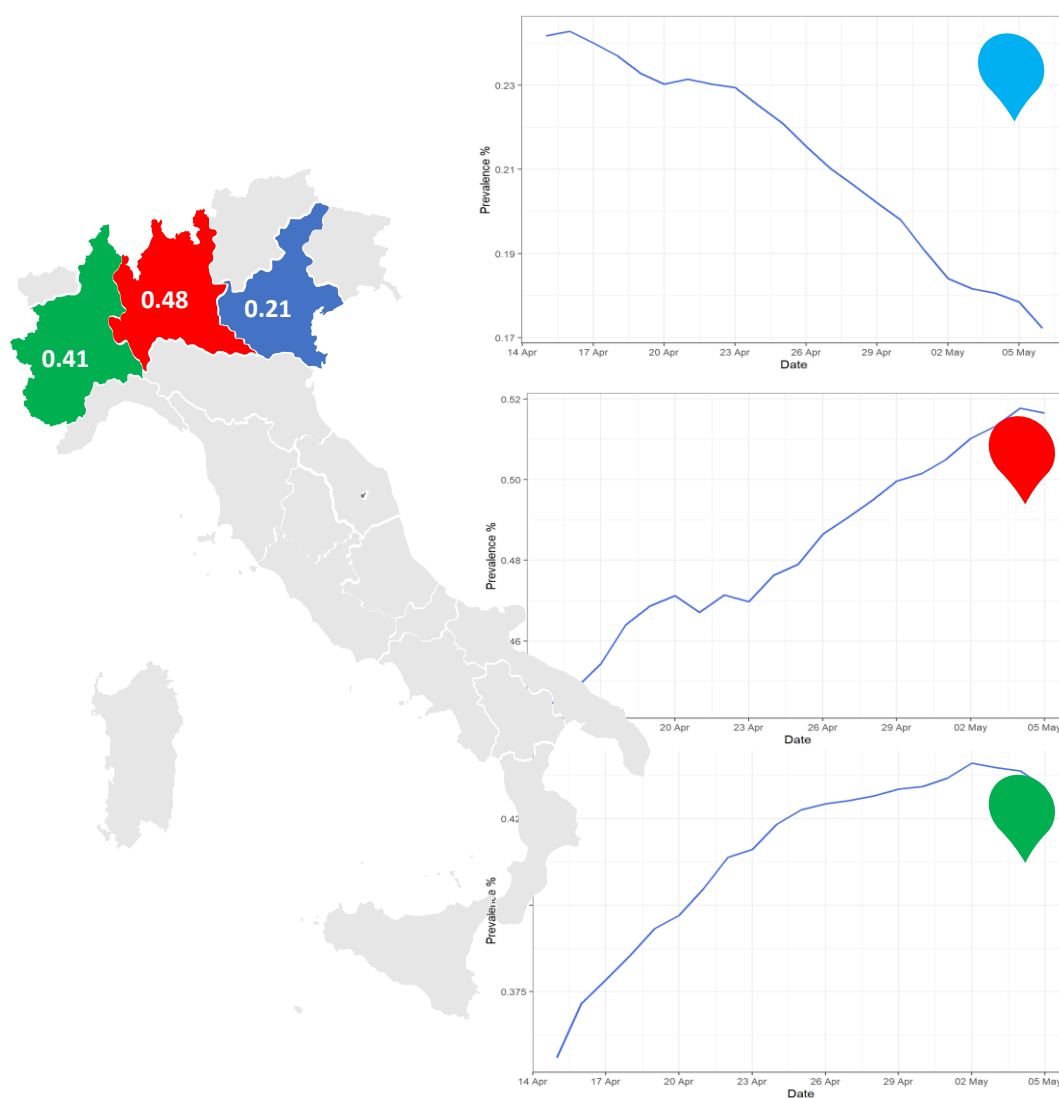


Figure 6 Average prevalence and trend of COVID-19 positive prevalence in the Italian regions considered in the study: Veneto (blu), Lombardia (red) and Piemonte (green). Reported prevalence refer to the proportion of COVID-19 cases in the population of each region in the period between 15 April 2020 and 6 May 2020.

The estimates for Veneto and Piemonte are closer to official data considering the random degree model (Figure 3) compared to those of Lombardia. The estimated prevalence of COVID-19 cases in Lombardia is 2.68% vs 0.41% of official data. In Veneto of 0.24% vs 0.21% and in Piemonte of 0.43% vs 0.41%.

The random degree model, compared to the models that account for barrier effect and transmission bias, seems to have lower differences from the official data in all the regions considered (Table 7). In Veneto, for example, the differences in percentages between the estimated number of COVID-19 cases and official data are 0.03%, 0.21%, and 0.12% in the random degree, barrier effect, and transmission bias model, respectively, considering default μ .

Table 7 Differences of the estimates from swab COVID-19 positive with default μ , the network size, for each region considered. The differences are reported in percentages.

	Question	Veneto	Piemonte	Lombardia
Random degree	swab COVID positive	0.03	0.02	2.2
	undocumented COVID-19 cases	0.03	0.03	2.3
Barrier effect	swab COVID positive	0.21	0.36	3.97
	undocumented COVID-19 cases	0.21	0.34	3.97
Transmission bias	swab COVID positive	0.12	0.23	3.31
	undocumented COVID-19 cases	0.12	0.2	3.39

In Figures 7, 8, and 9, the differences between estimated COVID-19 and undocumented cases from official data are reported for each dimension of network size μ according to the random degree, transmission bias, and barrier effect models, respectively. The random degree model estimates are closer to official data, especially for Veneto and Piemonte. The differences with default μ are 0.03 for Veneto, 0.02 for Piemonte, and 2.2 for Lombardia. In the transmission model, the differences are 0.12, 0.23, and 3.31, in the barrier effect 0.21, 0.36, and 3.97. The estimates for the Lombardia region are the most distant from official data in all three models.

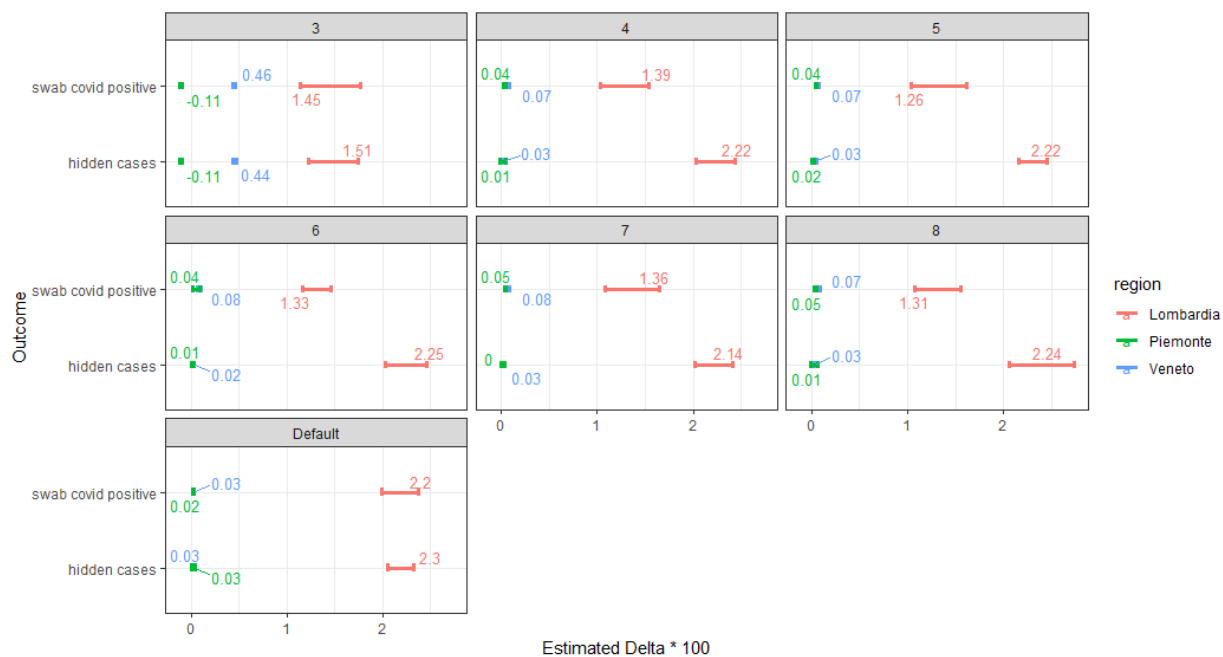


Figure 7 Variation from official data for the random degree model according to region and different μ .

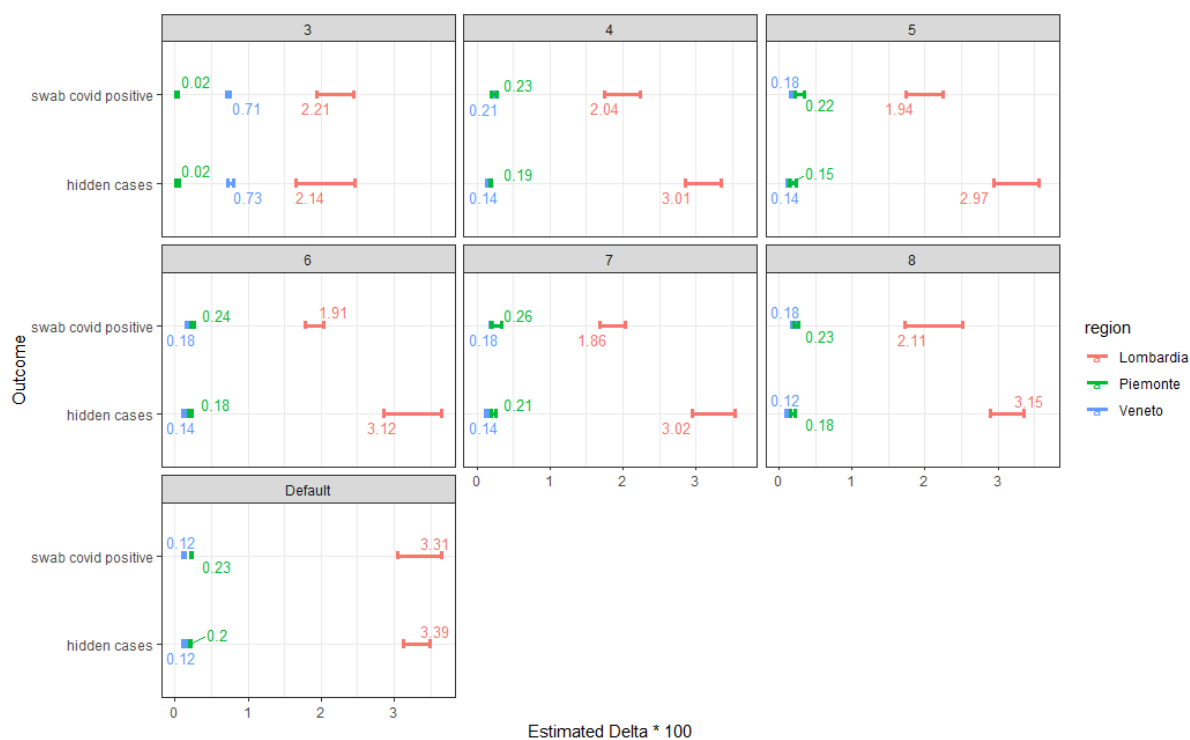


Figure 8 Variation from official data for the Transmission bias model according to region and different μ .

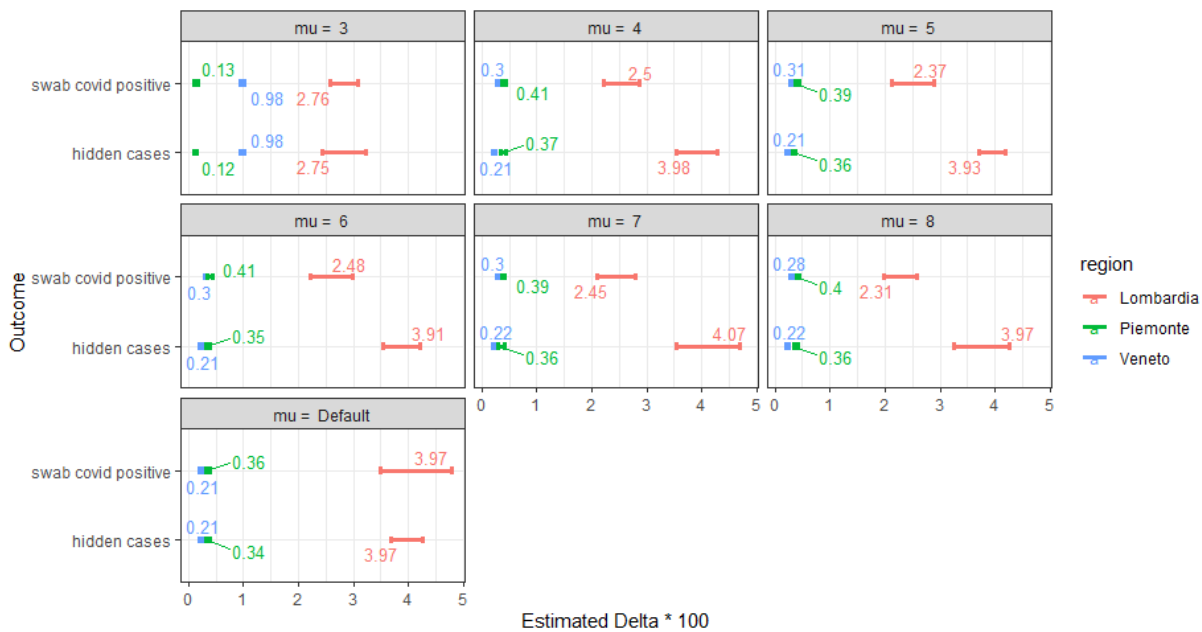


Figure 9 Variation from official data for the Barrier Effect model according to region and different μ .

4.4.3 Undocumented COVID-19 cases

Estimated undocumented cases are higher in Lombardia compared to Piemonte and Veneto, respectively, 2.78%, 0.44%, and 0.24% of the population of each region for default μ (Figure 3) according to the random degree model. In all the models considered, the estimates for Lombardia are higher. The estimates of undocumented cases are higher for the transmission bias model, respectively 3.87%, 0.61%, and 0.33%. In the barrier effect model are 4.45%, 0.75% and 0.43%. The estimates are indicated respectively for Lombardia, Piemonte, and Veneto, considering the default μ .

4.4.4 Cohabitants of isolated COVID-19

For what concerns the proportion of cohabitants of COVID-19 positive, our results report a higher proportion in Lombardia (1.77%) and Piemonte (0.45%) compared to Veneto (0.27%) according to the random degree model with default μ (Figure 3). Even in this hidden population, the estimates increased in the other two models (Figure 4 and Figure 5)

4.4.5 Number of people that transfer after the Law act (DPCM)

The estimated prevalence for the number of persons that have moved from one region to another is lower in Veneto (0.28%) compared to Piemonte, 0.44%, and Lombardia, 1.82% for default μ in the

random degree model (Figure 5). It means that an estimate of 13479 people moved from Veneto to another region (95% CrI 13136-17304), 18833.69 from Piemonte (95% CrI 18701.16-19994.94), and 181193 from Lombardia (95% CrI 149829-197897) (Table 7).

4.5 Discussion

Indirect estimations methods are helpful when direct access to the target population is impracticable. Our method benefits the advantages of the traditional NSUM: the target population is not accessed, not even a part of it; data needed to calculate the numerator and denominator of the equation can be collected in different moments; the estimate is proportioned to the total population, not to a part or benchmark of it. The proposed NSUM has shown in the simulation study a slightly higher bias than the one proposed by Maltiel [39]. The bias was computed as the difference between the average estimated prevalence and the benchmark. However, the bias levels are tolerable, 0.2 versus approximately -0.05, considering a study size of 1500 people (Table 4). The estimates are constant for both models considering the sample size that goes from 1000 to 3000 and are close to 1.5 and close to 1.37)[136]. So, the low loss in precision is tolerable compared to the benefits derived by the shortness of the questionnaire with a consequent reduction of drop-outs and response bias. Using only one question can increase the desire to conclude the survey, especially considering that for some people answering the questions to determine the network size may be unclear. Using an online survey has also reduced the costs and times required to manage it compared to a telephone survey. The costs faced by the researchers are those related to the creation, distribution of the survey, and costs of maintaining the platform. Other indirect estimate methods require direct access to the target population, like the capture-recapture techniques [15] and multiplier methods [138]. Moreover, there is no need for fieldwork as in enumeration methods, synthetic estimation, and multivariate indicator methods [7]. Testing for positivity has greater costs compared to a survey.

The estimates related to the COVID-19 cases are used as a comparator to evaluate our model precision. This is because it is not possible to apply our method to traditional NSUM datasets, even in the Bayesian formulation, since they require more than one anchoring question. The estimates for COVID-19 cases are closer to official data, especially for Veneto and Piemonte (Table 7). The random degree model especially shows a higher precision closeness to official prevalence, respectively a difference of 0.03% in Veneto and 0.02% in Piemonte. Lombardia estimates are instead partly distant from official data for each model considered. The differences vary from 2.2% in the random degree model to 3.39% in the transmission bias model.

Our model overestimates the undocumented cases of COVID-19, probably because the category of undocumented cases includes both asymptomatic and presymptomatic, as reported in Liu et al. work [12]. The level of overestimation was evaluated for the number of swab covid positive and undocumented cases as a variation from official data. The random degree model is the one that overestimates less compared to the barrier effect, and transmission bias models. Considering the default, the undocumented cases are higher in Piemonte and Lombardia compared to Veneto (0.44% and 2.78% vs. 0.24%, respectively). The same is true for COVID-19 cases. Estimates with our approach are higher in Piemonte and Lombardia compared to Veneto (0.43% and 2.68% vs 0.24%, respectively). These differences may be explained by the different strategies applied to face the pandemic. The Altems report of the Cattolica University of Roma compared the models of assistance in various Italian regions to respond to COVID-19 [139]. Veneto and partly Piemonte had a model based on the territory. Instead, Lombardia is mainly based on hospitalization. This means that there is a difference in the policy of mass testing in the three regions considered. Veneto started an early mass testing policy and quarantine measure. Lombardia and Piemonte started later the measure of containments.

The survey results represent a snapshot of the prevalence in the first wave of the pandemic. This survey could be repeated at different moments to evaluate the change in the estimates, especially the undocumented and COVID-19 cases. This would be of particular interest since the hidden population considered in this study is not the one usually evaluated in this kind of work, e.g., drug users, disadvantaged groups, and ethnic/racial groups. Especially the population of undocumented cases of COVID-19 was difficult to define in that period, the pandemic was just begun, and people did not have a clear definition of COVID-19 or asymptomatic/paucisymptomatic people. The virus and its consequences were still unknown and unpredictable in that phase.

Our estimates show a higher proportion of cohabitants with COVID-19 positive person in Lombardia, 1.77% compared to 0.45% of Piemonte and 0.27% in Veneto. This may be explained by the higher prevalence of COVID-19 cases (average prevalence for the period of 0.48% vs. 0.41% and 0.21% of Piemonte, Lombardia, and Veneto).

Lombardia reported the highest number of transfers after the first lockdown. The estimates are 181193 persons (95% CrI 149829-197897), followed by Piemonte with estimates of 18833.69 people (95% CrI 18701.16-19994.94), and Veneto 13479 (95% CrI 13136-17304). The transfers from north to south are probably higher in Lombardia since both the Governmental measures of 23 of February

[140] and the one of 8 of March [141] have put in lockdown mainly Lombardia and other provinces of Piemonte and some of Veneto.

Limitations

Despite the advantages, our model still has some issues that need to be reported. At first, using only one question to estimate the network degree added more variability since different network sizes are estimated by answering different questions, so that the final estimate may depend on the question randomly selected. Some of the known populations in this survey may be more familiar in a specific group. For example, the number of people who go to school by bus is a population that is mainly known by people that are parents. However, with the sample increase, the effect on the final estimate of the random question may be reduced. Our models results are difficult to compare to the traditional NSUM since we use only one question to compute the network size. Our estimates may suffer from uncertainty considering that persons might not identify with the key population, especially in the period considered the first phase of COVID-19. In the end, our estimates suffer from less precision in the region with fewer questionnaires answered, so a high number of respondents would probably increase the accuracy of our estimates. This probably could be solved with more effort in survey distribution and properly choosing the questions to define the network size.

4.6 Conclusions

Despite the presence of a gold standard method, such as testing methods that can detect either SARS-CoV-2 or biomarkers of SARS-CoV-2 or antibodies for detecting the size of undocumented cases, the use of an indirect method could still help in defining the real estimates of them. This was true, especially in the first period of the pandemic, a period in which the tests were not as diffuse as shown in our study. Different regions in the same nation had different testing policies. Testing all the population, even the asymptomatic ones, requires a great amount of money and resources. Moreover, using multiple methods to estimate hidden populations, such as the undocumented cases of COVID 19, could reduce the potential bias that can arise with a single method. Recent studies on hidden populations suggest using different methods to estimate the same people to have the most precise estimate [7,20,138].

Our approach may have the potential to make NSUM more readily implementable by reducing the number of questions to pose. This could potentially increase the response rate and consequently the estimate's precision. The choice of accounting for the barrier effect and transmission bias does not

show that it increases the estimate's precision. So, in this case, we suggest using the random degree model alone.

The results reported in this work are promising. Still, Maltiel's model modification proposed in this work must be tested in other studies to evaluate its sensitivity and with larger cohorts.

Chapter 5.

Conclusion

The NSUM has been used broadly in public health as it emerge from the review presented in this work. This estimator has been mainly used to estimate hard-to-reach populations which are the focus of public health. This work proposed to broaden the definition of hard-to-reach populations to a clinical setting such as COVID-19 cases. Undocumented cases of COVID-19 could be considered hard-to-reach in the specific period considered by our survey. In that period, the real magnitude of undocumented cases was unclear, even considering official data. So, a possible application of the NSUM could be a “pilot estimate” of the population of interest. This is viable thanks to the easiness of use of our short form and the reduced costs that are mainly based on the time required to distribute the survey. This “pilot estimate” could be helpful to define the magnitude of the phenomenon considered and to prepare for a more detailed study.

To evaluate the strength of the estimator proposed, it would be helpful to compare directly with the original estimator. This has been done in the simulation study proposed in the third chapter. However, comparing the two models was done using a simulated dataset. The ideal scenario would be to create another survey to compare the two models. One group will receive the short-form questionnaire, and the other group the long form.

Bibliography

- 1 Bonevski B, Randell M, Paul C, *et al.* Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. *BMC Med Res Methodol* 2014;**14**:42. doi:10.1186/1471-2288-14-42
- 2 Heiat A, Gross CP, Krumholz HM. Representation of the elderly, women, and minorities in heart failure clinical trials. *Arch Intern Med* 2002;**162**:1682–8. doi:10.1001/archinte.162.15.1682
- 3 Sydor A. Conducting research into hidden or hard-to-reach populations. *Nurse Res* 2013;**20**:33–7. doi:10.7748/nr2013.01.20.3.33.c9495
- 4 Lambert EY. *The collection and interpretation of data from hidden populations*. US Department of Health and Human Services, Public Health Service, Alcohol ... 1990. <https://archives.drugabuse.gov/sites/default/files/monograph98.pdf>
- 5 Bernard HR, Hallett T, Iovita A, *et al.* Counting hard-to-count populations: the network scale-up method for public health. *Sex Transm Infect* 2010;**86**:ii11–5. doi:10.1136/sti.2010.044446
- 6 UNAIDS/WHO. Estimating the size of populations at risk for HIV. Geneva: 2003. https://data.unaids.org/publications/external-documents/estimatingpopsizes_en.pdf (accessed 1 Jun 2022).
- 7 Abdul-Quader AS, Baughman AL, Hladik W. Estimating the size of key populations: current status and future possibilities. *Curr Opin HIV AIDS* 2014;**9**:107–14. doi:10.1097/COH.0000000000000041
- 8 Bernard HR, Johnsen EC, Killworth PD, *et al.* Estimating the Size of an Average Personal Network and of an Event Subpopulation. *The small world*;:159–75.
- 9 Laga I, Bao L, Niu X. Thirty years of the network scale-up method. *Journal of the American Statistical Association* 2021;**116**:1548–59.
- 10 Killworth PD, Johnsen EC, McCarty C, *et al.* A social network approach to estimating seroprevalence in the United States. *Social Networks* 1998;**20**:23–50. doi:10.1016/S0378-8733(96)00305-X

Bibliography

- 11 Rolstad S, Adler J, Rydén A. Response Burden and Questionnaire Length: Is Shorter Better? A Review and Meta-analysis. *Value in Health* 2011;**14**:1101–8. doi:10.1016/j.jval.2011.06.003
- 12 Liu P, Beeler P, Chakrabarty RK. Dynamic interplay between social distancing duration and intensity in reducing COVID-19 US hospitalizations: A “law of diminishing returns”. *Chaos* 2020;**30**:071102. doi:10.1063/5.0013871
- 13 Djilali S, Benahmadi L, Tridane A, *et al.* Modeling the Impact of Unreported Cases of the COVID-19 in the North African Countries. *Biology* 2020;**9**:373. doi:10.3390/biology9110373
- 14 Smith DRM, Opatowski L. COVID-19 containment measures and incidence of invasive bacterial disease. *The Lancet Digital Health* 2021;**3**:e331–2. doi:10.1016/S2589-7500(21)00085-6
- 15 Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev* 1995;**17**:243–64. doi:10.1093/oxfordjournals.epirev.a036192
- 16 Berchenko Y, Frost SDW. Capture-recapture methods and respondent-driven sampling: their potential and limitations. *Sex Transm Infect* 2011;**87**:267–8. doi:10.1136/sti.2011.049171
- 17 Mutagoma M, Kayitesi C, Gwiza A, *et al.* Estimation of the size of the female sex worker population in Rwanda using three different methods. *Int J STD AIDS* 2015;**26**:810–4. doi:10.1177/0956462414555931
- 18 Mutagoma M, Kayitesi C, Gwiza A, *et al.* Estimation of the size of the female sex worker population in Rwanda using three different methods. *Int J STD AIDS* 2015;**26**:810–4. doi:10.1177/0956462414555931
- 19 Yip PS, Bruno G, Tajima N, *et al.* International Working Group for Disease Monitoring and Forecasting. Capture—recapture and multiple-record systems estimation: history and theoretical development. *Am J Epidemiol* 1995;**142**:1047–58.
- 20 Okal J, Geibel S, Muraguri N, *et al.* Estimates of the size of key populations at risk for HIV infection: men who have sex with men, female sex workers and injecting drug users in Nairobi, Kenya. *Sex Transm Infect* 2013;**89**:366–71. doi:10.1136/sextrans-2013-051071
- 21 Galton F. Vox Populi. *Nature* 1907;**75**:450–1. doi:10.1038/075450a0

Bibliography

- 22 McCormick TH. The network scale-up method. *The Oxford Handbook of Social Networks* 2020;:153.
- 23 Moher D, Liberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;**151**:264–9, W64. doi:10.7326/0003-4819-151-4-200908180-00135
- 24 Covidence. Covidence - Better systematic review management. <https://www.covidence.org/> (accessed 3 Jun 2022).
- 25 Page MJ, McKenzie JE, Bossuyt PM, *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;**372**:n71. doi:10.1136/bmj.n71
- 26 Snidero S, Morra B, Corradetti R, *et al.* The network scale-up method: A simulation study in case of overlapping sub-populations. *Social Networks* 2007;**29**:527–38.
- 27 Rastegari A, Baneshi MR, Haji-maghsoudi S, *et al.* Estimating the Annual Incidence of Abortions in Iran Applying a Network Scale-up Approach. *Iran Red Crescent Med J* 2014;**16**. doi:10.5812/ircmj.15765
- 28 Sheikhzadeh K, Baneshi MR, Afshari M, *et al.* Comparing direct, network scale-up, and proxy respondent methods in estimating risky behaviors among collegians. *Journal of Substance Use* 2016;**21**:9–13. doi:10.3109/14659891.2014.942401
- 29 Halimi L, Babamiri M, Hamidi Y, *et al.* Indirect estimation of student marijuana consumers population in Hamadan using PRM and NSU methods. *Medical Journal of the Islamic Republic of Iran* 2020;**34**:1–7. doi:10.34171/mjiri.34.160
- 30 Kadushin C, Killworth PD, Bernard HR, *et al.* Scale-up methods as applied to estimates of heroin use. *Journal of Drug Issues* 2006;**36**:417–40.
- 31 Jing L, Lu Q, Cui Y, *et al.* Combining the randomized response technique and the network scale-up method to estimate the female sex worker population size: an exploratory study. *Public Health* 2018;**160**:81–6. doi:10.1016/j.puhe.2018.04.002
- 32 Jing L, Qu C, Yu H, *et al.* Estimating the sizes of populations at high risk for HIV: a comparison study. *PloS one* 2014;**9**.

- 33 Carletti G, Soriani N, Mattiazzi M, *et al.* A Social Network Approach to the Estimation of Perceived Quality of Health Care. *Open Nurs J* 2017;**11**:219–31. doi:10.2174/1874434601711010219
- 34 Guo W, Bao S, Lin W, *et al.* Estimating the size of HIV key affected populations in Chongqing, China, using the network scale-up method. *PloS one* 2013;**8**.
- 35 Heydari Z, Baneshi MR, Sharifi H, *et al.* Evaluation of the treatment failure ratio in individuals receiving methadone maintenance therapy via the network scale up method. *International Journal of Drug Policy* 2019;**73**:36–41. doi:10.1016/j.drugpo.2019.07.022
- 36 Haghdoost AA, Baneshi MR, Haji-Maghsoudi S, *et al.* Application of a Network Scale-up Method to Estimate the Size of Population of Breast, Ovarian/Cervical, Prostate and Bladder Cancers. *Asian Pac J Cancer Prev* 2015;**16**:3273–7. doi:10.7314/apjcp.2015.16.8.3273
- 37 McCarty C, Killworth PD, Bernard HR, *et al.* Comparing Two Methods for Estimating Network Size. *Human Organization* 2001;**60**:28–39. doi:10.17730/humo.60.1.efx5t9gjtgmga73y
- 38 McCormick TH, Salganik MJ, Zheng T. How many people do you know?: Efficiently estimating personal network size. *J Am Stat Assoc* 2010;**105**:59–70. doi:10.1198/jasa.2009.ap08518
- 39 Maltiel R, Raftery AE, McCormick TH, *et al.* Estimating Population Size Using the Network Scale Up Method. *Ann Appl Stat* 2015;**9**:1247–77. doi:10.1214/15-AOAS827
- 40 Salganik MJ, Mello MB, Abdo AH, *et al.* The Game of Contacts: Estimating the Social Visibility of Groups. *Soc Networks* 2011;**33**:70–8. doi:10.1016/j.socnet.2010.10.006
- 41 Salganik MJ, Fazito D, Bertoni N, *et al.* Assessing network scale-up estimates for groups most at risk of HIV/AIDS: evidence from a multiple-method study of heavy drug users in Curitiba, Brazil. *Am J Epidemiol* 2011;**174**:1190–6. doi:10.1093/aje/kwr246
- 42 Shati M, Haghdoost A, Majdzadeh R, *et al.* Social network size estimation and determinants in Tehran Province residents. *Iranian Journal of Public Health* 2014;**43**:1079–90.
- 43 Motazedian N, Sayadi M, Oboodi S, *et al.* Estimation of the active social network of street children in Shiraz (a city of Iran). *International Journal of High Risk Behaviors and Addiction* 2020;**9**. doi:10.5812/IJHRBA.94467

Bibliography

- 44 Rastegari A, Haji-Maghsoudi S, Haghdoost A, *et al.* The estimation of active social network size of the Iranian population. *Glob J Health Sci* 2013;**5**:217–27. doi:10.5539/gjhs.v5n4p217
- 45 Shokoohi M, Baneshi MR, Haghdoost AA. Estimation of the Active Network Size of Kermanian Males. *Addict Health* 2010;**2**:81–8.
- 46 Zamanian M, Baneshi MR, Haghdoost AA, *et al.* Estimating the Size and Age-gender Distribution of Women’s Active Social Networks. *Addict Health* 2016;**8**:170–8.
- 47 Killworth PD, McCarty C, Bernard HR, *et al.* Estimation of seroprevalence, rape, and homelessness in the United States using a social network approach. *Eval Rev* 1998;**22**:289–308. doi:10.1177/0193841X9802200205
- 48 Habecker P, Dombrowski K, Khan B. Improving the Network Scale-Up Estimator: Incorporating Means of Sums, Recursive Back Estimation, and Sampling Weights. *PLoS One* 2015;**10**. doi:10.1371/journal.pone.0143406
- 49 Feehan DM, Salganik MJ. Generalizing the Network Scale-Up Method: A New Estimator for the Size of Hidden Populations. *Sociol Methodol* 2016;**46**:153–86. doi:10.1177/0081175016665425
- 50 Verdery AM, Weir S, Reynolds Z, *et al.* Estimating Hidden Population Sizes with Venue-based Sampling. *Epidemiology* 2019;**30**:901–10. doi:10.1097/EDE.0000000000001059
- 51 Zheng T, Salganik MJ, Gelman A. How many people do you know in prison? Using overdispersion in count data to estimate social structure in networks. *Journal of the American Statistical Association* 2006;**101**:409–23.
- 52 McCormick TH, Zheng T. Latent demographic profile estimation in hard-to-reach groups. *The Annals of Applied Statistics* 2012;**6**:1795–813. doi:10.1214/12-AOAS569
- 53 Teo AKJ, Prem K, Chen MIC, *et al.* Estimating the size of key populations for HIV in Singapore using the network scale-up method. *Sex Transm Infect* 2019;**95**:602–7. doi:10.1136/sextrans-2018-053747
- 54 Josephs N, Feehan DM, Crawford FW. A sample size heuristic for network scale-up studies. arXiv 2021. doi:10.48550/arXiv.2111.09684

- 55 Maghsoudi A, Baneshi MR, Neydavoodi M, *et al.* Network Scale-Up Correction Factors for Population Size Estimation of People Who Inject Drugs and Female Sex Workers in Iran. *PLoS One* 2014;**9**. doi:10.1371/journal.pone.0110917
- 56 Haghdoost A, Ahmadi Gohari M, Mirzazadeh A, *et al.* A review of methods to estimate the visibility factor for bias correction in network scale-up studies. *Epidemiol Health* 2018;**40**. doi:10.4178/epih.e2018041
- 57 Rwanda Biomedical Center/Institute of HIV/AIDS DP and CD-R, SPH/Rwanda S of PH-, UNAIDS, *et al.* Estimating the Size of Populations through a Household Survey (ESPHS) Rwanda 2011. 2012. <https://dhsprogram.com/publications/publication-fr261-other-final-reports.cfm> (accessed 15 Jul 2020).
- 58 BPU, CIF. Population size estimation of people who inject drugs in Georgia 2016. 2016. <http://curatiofoundation.org/wp-content/uploads/2018/02/PWID-PSE-Report-2017-ENG.pdf> (accessed 2 Jun 2022).
- 59 Johnsen EC, Killworth PD, Robinson S. Estimating the size of an average personal network and of an event subpopulation. *The small world* 1989;:159–75.
- 60 WHO Statement Regarding Cluster of Pneumonia Cases in Wuhan, China. <https://www.who.int/china/news/detail/09-01-2020-who-statement-regarding-cluster-of-pneumonia-cases-in-wuhan-china> (accessed 16 Apr 2020).
- 61 Cascella M, Rajnik M, Cuomo A, *et al.* Features, Evaluation and Treatment Coronavirus (COVID-19). In: *StatPearls*. Treasure Island (FL): : StatPearls Publishing 2020. <http://www.ncbi.nlm.nih.gov/books/NBK554776/> (accessed 16 Apr 2020).
- 62 Li R, Pei S, Chen B, *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* Published Online First: 16 March 2020. doi:10.1126/science.abb3221
- 63 Palatella L, Vanni F, Lambert D. A phenomenological estimate of the true scale of CoViD-19 from primary data. *Chaos, Solitons & Fractals* 2021;**146**:110854. doi:10.1016/j.chaos.2021.110854

- 64 Nishiura H, Kobayashi T, Yang Y, *et al.* The Rate of Underascertainment of Novel Coronavirus (2019-nCoV) Infection: Estimation Using Japanese Passengers Data on Evacuation Flights. *Journal of Clinical Medicine* 2020;**9**:419. doi:10.3390/jcm9020419
- 65 Zou L, Ruan F, Huang M, *et al.* SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *New England Journal of Medicine* Published Online First: 19 February 2020. doi:10.1056/NEJMc2001737
- 66 Nishiura H, Kobayashi T, Suzuki A, *et al.* Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *International Journal of Infectious Diseases* 2020;:S1201971220301399. doi:10.1016/j.ijid.2020.03.020
- 67 Mizumoto K, Kagaya K, Zarebski A, *et al.* Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance* 2020;**25**:2000180. doi:10.2807/1560-7917.ES.2020.25.10.2000180
- 68 Lavezzo E, Franchin E, Ciavarella C, *et al.* Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature* Published Online First: 30 June 2020. doi:10.1038/s41586-020-2488-1
- 69 Bai Y, Yao L, Wei T, *et al.* Presumed Asymptomatic Carrier Transmission of COVID-19. *JAMA* 2020;**323**:1406–7. doi:10.1001/jama.2020.2565
- 70 Wei WE. Presymptomatic Transmission of SARS-CoV-2 — Singapore, January 23–March 16, 2020. *MMWR Morb Mortal Wkly Rep* 2020;**69**. doi:10.15585/mmwr.mm6914e1
- 71 Laboratory testing for 2019 novel coronavirus (2019-nCoV) in suspected human cases. <https://www.who.int/publications-detail-redirect/10665-331501> (accessed 2 Jul 2020).
- 72 Evans AS. *Viral Infections of Humans: Epidemiology and Control*. Springer Science & Business Media 2013.
- 73 Capture-recapture and multiple-record systems estimation II: Applications in human diseases. International Working Group for Disease Monitoring and Forecasting. *Am J Epidemiol* 1995;**142**:1059–68.

- 74 World Health Organization. *Guidelines on estimating the size of populations most at risk to HIV*. Geneva: World Health Organization 2010. https://apps.who.int/iris/bitstream/handle/10665/44611/9789241501668_eng.pdf (accessed 28 Apr 2021).
- 75 Johnsen EC, Killworth PD, Robinson S. Estimating the size of an average personal network and of an event subpopulation. *The small world* 1989;:159–75.
- 76 Bernard HR, Johnsen EC, Killworth PD, *et al.* Estimating the size of an average personal network and of an event subpopulation: Some empirical results. *Social science research* 1991;20:109–21.
- 77 Verdery AM, Weir S, Reynolds Z, *et al.* Estimating Hidden Population Sizes with Venue-based Sampling. *Epidemiology* 2019;30:901–10. doi:10.1097/EDE.0000000000001059
- 78 Snidero S, Morra B, Corradetti R, *et al.* Use of the scale-up methods in injury prevention research: An empirical assessment to the case of choking in children. *Social Networks* 2007;29:527–38.
- 79 Ezoe S, Morooka T, Noda T, *et al.* Population size estimation of men who have sex with men through the network scale-up method in Japan. *PLoS ONE* 2012;7:e31184. doi:10.1371/journal.pone.0031184
- 80 Sadler GR, Lee H-C, Seung-Hwan Lim R, *et al.* Recruiting hard-to-reach United States population sub-groups via adaptations of snowball sampling strategy. *Nurs Health Sci* 2010;12:369–74. doi:10.1111/j.1442-2018.2010.00541.x
- 81 Biernacki P, Waldorf D. Snowball Sampling: Problems and Techniques of Chain Referral Sampling. *Sociological Methods & Research* Published Online First: 29 June 2016. doi:10.1177/004912418101000205
- 82 Jing L, Lu Q, Cui Y, *et al.* Combining the randomized response technique and the network scale-up method to estimate the female sex worker population size: an exploratory study. *Public Health* 2018;160:81–6. doi:10.1016/j.puhe.2018.04.002
- 83 Killworth PD, McCarty C, Bernard HR, *et al.* Estimation of seroprevalence, rape, and homelessness in the United States using a social network approach. *Eval Rev* 1998;22:289–308. doi:10.1177/0193841X9802200205

Bibliography

- 84 Guo W, Bao S, Lin W, *et al.* Estimating the size of HIV key affected populations in Chongqing, China, using the network scale-up method. *PloS one* 2013;**8**.
- 85 Teo AKJ, Prem K, Chen MIC, *et al.* Estimating the size of key populations for HIV in Singapore using the network scale-up method. *Sex Transm Infect* 2019;**95**:602–7. doi:10.1136/sextrans-2018-053747
- 86 Feehan DM, Umubyeyi A, Mahy M, *et al.* Quantity Versus Quality: A Survey Experiment to Improve the Network Scale-up Method. *Am J Epidemiol* 2016;**183**:747–57. doi:10.1093/aje/kwv287
- 87 Jafari Khounigh Ali A, Haghdoost AA, SalariLak S, *et al.* Size estimation of most-at-risk groups of HIV/AIDS using network scale-up in Tabriz, Iran. *Journal of Clinical Research & Governance* 2014;**3**:21–6.
- 88 Maltiel R, Raftery AE, McCormick TH, *et al.* Estimating Population Size Using the Network Scale Up Method. *Ann Appl Stat* 2015;**9**:1247–77. doi:10.1214/15-AOAS827
- 89 Salganik MJ, Fazito D, Mello MB. Estimating the number of heavy drug users in Curitiba, Brazil using multiple methods. 2010.
- 90 Shokoohi M, Baneshi MR, Haghdoost A-A. Size Estimation of Groups at High Risk of HIV/AIDS using Network Scale Up in Kerman, Iran. *Int J Prev Med* 2012;**3**:471–6.
- 91 Wang J, Yang Y, Zhao W, *et al.* Application of network scale up method in the estimation of population size for men who have sex with men in Shanghai, China. *PloS one* 2015;**10**.
- 92 Maghsoudi A, Baneshi MR, Neydavoodi M, *et al.* Network Scale-Up Correction Factors for Population Size Estimation of People Who Inject Drugs and Female Sex Workers in Iran. *PLoS One* 2014;**9**. doi:10.1371/journal.pone.0110917
- 93 Ahmadi-Gohari M, Zolala F, Iranpour A, *et al.* Twelve-hour before Driving Prevalence of Alcohol and Drug Use among Heavy Vehicle Drivers in South East of Iran Using Network Scale Up. *Addict Health* 2019;**11**:256–61. doi:10.22122/ahj.v11i4.252
- 94 Narouee S, Shati M, Nasehi M, *et al.* The size estimation of injection drug users (IDUs) using the network scale-up method (NSUM) in Iranshahr, Iran. *Med J Islam Repub Iran* 2019;**33**:158. doi:10.34171/mjiri.33.158

- 95 Nikfarjam A, Shokoohi M, Shahesmaeili A, *et al.* National population size estimation of illicit drug users through the network scale-up method in 2013 in Iran. *Int J Drug Policy* 2016;**31**:147–52. doi:10.1016/j.drugpo.2016.01.013
- 96 Zahedi R, Noroozi A, Hajebi A, *et al.* Self-Reported and Network Scale-Up Estimates of Substance Use Prevalence among University Students in Kerman, Iran. *J Res Health Sci* 2018;**18**:e00413.
- 97 Kadushin C, Killworth PD, Bernard HR, *et al.* Scale-up methods as applied to estimates of heroin use. *Journal of Drug Issues* 2006;**36**:417–40.
- 98 Nikfarjam A, Hajimaghsoudi S, Rastegari A, *et al.* The Frequency of Alcohol Use in Iranian Urban Population: The Results of a National Network Scale Up Survey. *Int J Health Policy Manag* 2016;**6**:97–102. doi:10.15171/ijhpm.2016.103
- 99 Heydari Z, Baneshi MR, Sharifi H, *et al.* Evaluation of the treatment failure ratio in individuals receiving methadone maintenance therapy via the network scale up method. *International Journal of Drug Policy* 2019;**73**:36–41. doi:10.1016/j.drugpo.2019.07.022
- 100 Mohebbi E, Baneshi MR, Haji-maghsoudi S, *et al.* The Application of Network Scale Up Method on Estimating the Prevalence of Some Disabilities in the Southeast of Iran. *Journal of Research in Health Sciences* 2014;**14**:272-275–275.
- 101 Rastegari A, Baneshi MR, Haji-maghsoudi S, *et al.* Estimating the Annual Incidence of Abortions in Iran Applying a Network Scale-up Approach. *Iran Red Crescent Med J* 2014;**16**. doi:10.5812/ircmj.15765
- 102 Zamanian M, Zolala F, Haghdoost AA, *et al.* Methodological considerations in using the Network Scale Up (NSU) for the estimation of risky behaviors of particular age-gender groups: An example in the case of intentional abortion. *PLoS One* 2019;**14**. doi:10.1371/journal.pone.0217481
- 103 Snidero S, Soriani N, Baldi I, *et al.* Scale-up Approach in CATI Surveys for Estimating the Number of Foreign Body Injuries in the Aero-digestive Tract in Children. *Int J Environ Res Public Health* 2012;**9**:4056–67. doi:10.3390/ijerph9114056

- 104 Carletti G, Soriani N, Mattiazzi M, *et al.* A Social Network Approach to the Estimation of Perceived Quality of Health Care. *Open Nurs J* 2017;**11**:219–31. doi:10.2174/1874434601711010219
- 105 Paniotto V, Petrenko T, Kupriyanov O, *et al.* Estimating the size of populations with high risk for HIV using the network scale-up method. *Ukraine: Kiev International Institute of Sociology* 2009.
- 106 World Health Organization - 2010 - Guidelines on estimating the size of populations m.pdf. http://data.unaids.org/pub/manual/2010/guidelines_popnestimationsize_en.pdf (accessed 21 May 2020).
- 107 McCormick TH, Salganik MJ, Zheng T. How many people do you know?: Efficiently estimating personal network size. *J Am Stat Assoc* 2010;**105**:59–70. doi:10.1198/jasa.2009.ap08518
- 108 Sajjadi H, Jorjoran Shushtari Z, Shati M, *et al.* An indirect estimation of the population size of students with high-risk behaviors in select universities of medical sciences: A network scale-up study. *PLoS One* 2018;**13**. doi:10.1371/journal.pone.0195364
- 109 Habecker P, Dombrowski K, Khan B. Improving the Network Scale-Up Estimator: Incorporating Means of Sums, Recursive Back Estimation, and Sampling Weights. *PLoS One* 2015;**10**. doi:10.1371/journal.pone.0143406
- 110 Haghdoost AA, Baneshi MR, Haji-Maghsoodi S, *et al.* Application of a Network Scale-up Method to Estimate the Size of Population of Breast, Ovarian/Cervical, Prostate and Bladder Cancers. *Asian Pac J Cancer Prev* 2015;**16**:3273–7. doi:10.7314/apjcp.2015.16.8.3273
- 111 Kazemzadeh Y, Shokoohi M, Baneshi MR, *et al.* The Frequency of High-Risk Behaviors Among Iranian College Students Using Indirect Methods: Network Scale-Up and Crosswise Model. *Int J High Risk Behav Addict* 2016;**5**. doi:10.5812/ijhrba.25130
- 112 Moradinazar M, Najafi F, Baneshi MR, *et al.* Size Estimation of Under-Reported Suicides and Suicide Attempts Using Network Scale up Method. *Bull Emerg Trauma* 2019;**7**:99–104. doi:10.29252/beat-070202.
- 113 Narouee S, Shatti M, Didevar M, *et al.* Estimating social network size using network scale-up method (NSUM) in Iranshahr, Sistan and Baluchestan Province, Iran. *Med J Islam Repub Iran* 2020;**34**:35. doi:10.34171/mjiri.34.35

- 114 Shokoohi M, Baneshi MR, Haghdoost AA. Estimation of the Active Network Size of Kermanian Males. *Addict Health* 2010;**2**:81–8.
- 115 Vardanjani HM, Baneshi MR, Haghdoost A. Total and Partial Prevalence of Cancer Across Kerman Province, Iran, in 2014, Using an Adapted Generalized Network Scale-Up Method. *Asian Pac J Cancer Prev* 2015;**16**:5493–8. doi:10.7314/apjcp.2015.16.13.5493
- 116 Zamanian M, Zolala F, Haghdoost AA, *et al.* Estimating The Annual Abortion Rate in Kerman, Iran: Comparison of Direct, Network Scale-Up, and Single Sample Count Methods. *Int J Fertil Steril* 2019;**13**:209–14. doi:10.22074/ijfs.2019.5721
- 117 Rastegari A, Haji-Maghsoudi S, Haghdoost A, *et al.* The estimation of active social network size of the Iranian population. *Glob J Health Sci* 2013;**5**:217–27. doi:10.5539/gjhs.v5n4p217
- 118 Gulati A, Pomeranz C, Qamar Z, *et al.* A Comprehensive Review of Manifestations of Novel Coronaviruses in the Context of Deadly COVID-19 Global Pandemic. *Am J Med Sci* Published Online First: 11 May 2020. doi:10.1016/j.amjms.2020.05.006
- 119 Istat.it. <https://www.istat.it/> (accessed 16 Apr 2020).
- 120 Bernard HR, Johnsen EC, Killworth PD, *et al.* Comparing four different methods for measuring personal social networks. *Social Networks* 1990;**12**:179–215. doi:10.1016/0378-8733(90)90005-T
- 121 Killworth PD, Johnsen EC, McCarty C, *et al.* A social network approach to estimating seroprevalence in the United States. *Social Networks* 1998;**20**:23–50. doi:10.1016/S0378-8733(96)00305-X
- 122 Maltiel R, Baraff AJ. *NSUM: Network Scale Up Method*. 2015. <https://CRAN.R-project.org/package=NSUM>
- 123 Team RC. *R: A language and environment for statistical computing*. 2013.
- 124 Alleva Giorgio, Arbia Giuseppe, Falorsi Pietro Demetrio, *et al.* A sample approach to the estimation of the critical parameters of the SARS-CoV-2 epidemics: an operational design with a focus on the Italian health system. doi:10.13140/RG.2.2.25507.40481
- 125 Noh J, Danuser G. Estimation of the fraction of COVID-19 infected people in U.S. states and countries worldwide. *PLoS One* 2021;**16**. doi:10.1371/journal.pone.0246772

- 126 World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> (accessed 5 Aug 2021).
- 127 Zou L, Ruan F, Huang M, *et al.* SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *New England Journal of Medicine* Published Online First: 19 February 2020. doi:10.1056/NEJMc2001737
- 128 Yin G, Jin H. Comparison of Transmissibility of Coronavirus Between Symptomatic and Asymptomatic Patients: Reanalysis of the Ningbo COVID-19 Data. *JMIR Public Health Surveill* 2020;**6**:e19464. doi:10.2196/19464
- 129 Yanes-Lane M, Winters N, Fregonese F, *et al.* Proportion of asymptomatic infection among COVID-19 positive persons and their transmission potential: A systematic review and meta-analysis. *PLOS ONE* 2020;**15**:e0241536. doi:10.1371/journal.pone.0241536
- 130 Buitrago-Garcia D, Egli-Gany D, Counotte MJ, *et al.* Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLOS Medicine* 2020;**17**:e1003346. doi:10.1371/journal.pmed.1003346
- 131 Day M. Covid-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village. *BMJ* 2020;**368**. doi:10.1136/bmj.m1165
- 132 McEvoy D, McAloon C, Collins A, *et al.* Relative infectiousness of asymptomatic SARS-CoV-2 infected persons compared with symptomatic individuals: a rapid scoping review. *BMJ Open* 2021;**11**:e042354. doi:10.1136/bmjopen-2020-042354
- 133 Chen Z, Wang B, Mao S, *et al.* Assessment of global asymptomatic SARS-CoV-2 infection and management practices from China. *Int J Biol Sci* 2021;**17**:1119–24. doi:10.7150/ijbs.59374
- 134 Li R, Pei S, Chen B, *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* Published Online First: 16 March 2020. doi:10.1126/science.abb3221
- 135 Mbwogge M. Mass Testing With Contact Tracing Compared to Test and Trace for the Effective Suppression of COVID-19 in the United Kingdom: Systematic Review. *JMIRx Med* 2021;**2**:e27254. doi:10.2196/27254

- 136 Ocagli H, Azzolina D, Lorenzoni G, *et al.* Using Social Networks to Estimate the Number of COVID-19 Cases: The Incident (Hidden COVID-19 Cases Network Estimation) Study Protocol. *Int J Environ Res Public Health* 2021;**18**:5713. doi:10.3390/ijerph18115713
- 137 Laureati - Laureati per anno - Open Data dell'istruzione superiore. <http://dati.ustat.miur.it/dataset/laureati/resource/43df861d-7345-481a-9803-2eb236aa022e> (accessed 31 May 2022).
- 138 Sharifi H, Karamouzian M, Baneshi MR, *et al.* Population size estimation of female sex workers in Iran: Synthesis of methods and results. *PLoS One* 2017;**12**:e0182755. doi:10.1371/journal.pone.0182755
- 139 Americo Cicchetti, Gianfranco Damiani, Maria Lucia Specchia, Michele Basile, Rossella Di, Bidino, Eugenio Di Brino, Maria Giovanna Di Paolo, Andrea Di Pilla, Fabrizio Massimo Ferrara,, Luca Giorgio, Maria Teresa Riccardi, Filippo Rumi, Angelo Tattoli. Analisi dei modelli organizzativi di risposta al Covid-19 Focus su Lombardia, Veneto, Emilia-Romagna, Piemonte e Lazio. Università Cattolica del Sacro Cuore 2020. <https://www.quotidianosanita.it/allegati/allegato2612154.pdf> (accessed 1 Jun 2022).
- 140 Presidente del consiglio dei ministri. Decreto del Presidente del Consiglio dei Ministri 23 febbraio 2020. 2020. <https://www.gazzettaufficiale.it/eli/id/2020/02/23/20A01228/sg> (accessed 16 Apr 2020).
- 141 Presidente del Consiglio dei Ministri. Decreto del Presidente del Consiglio dei Ministri 08 marzo 2020. 2020. <https://www.gazzettaufficiale.it/eli/id/2020/03/08/20A01522/sg> (accessed 16 Apr 2020).

Supplementary materials

Table S1 Full search string

Network Scale-Up Estimator"[Text Word] OR "Network scale-up method"[Text Word] OR "Network scale-up methods"[Text Word] OR "NSUM"[Text Word] OR "NSU"[Text Word] OR "Network scale-up"[Text Word] OR "network scale up method" [Text Word])

S2 INCIDENT questionnaire

INCIDENT

Descrizione dello studio

Capire quante persone sono potenzialmente positive al COVID-19 è di fondamentale importanza in un momento di emergenza come questo per definire le iniziative di contenimento dell'epidemia.

L'Università degli Studi di Padova insieme all'Università degli Studi di Torino e la Prochild Onlus hanno ideato lo studio INCIDENT (hIddeN CovID-19 casEs Network esTimation), con il supporto tecnico gratuito offerto da Zeta Research.

INCIDENT si propone pertanto di quantificare il numero di persone con infezioni da COVID-19 non documentate da tampone positivo attraverso il presente questionario volontario, anonimo e gratuito.

Il questionario è composto di 10 domande, di cui 4 socio-demografiche, 6 riguardanti il COVID-19 e l'attuale situazione di emergenza, e l'ultima è una domanda casuale per stimare la dimensione della rete di conoscenze.

I dati raccolti saranno trattati ad esclusivi fini di ricerca e divulgazione scientifica.

Definizione di conoscenza

In questo studio per conoscenza intendiamo qualcuno con cui hai reciproca conoscenza, di vista o di nome, oppure qualcuno con cui hai avuto un contatto (di persona, via telefono o per corrispondenza) negli ultimi due anni e che possa ripetersi oggi^{1,2}.

DOMANDE

- 1) Quanti anni hai?
- 2) Sesso
- 3) Paese d'origine
- 4) In che provincia risiedi attualmente?

Domande sulla sintomatologia

5) Quante persone conosci che hanno avuto almeno un sintomo come febbre ($>37.5^{\circ}\text{C}$), tosse secca, mal di gola, congestione nasale, malessere, mal di testa, dolore muscolare nelle ultime due settimane?

Domande sull'esposizione

6) Quante persone conosci che sono risultate positive al tampone per il COVID-19?

Domande sulle misure di sicurezza

7) Quante persone conosci che condividono l'abitazione con soggetti in isolamento a causa di manifestazione di sintomi simil-influenzali?

Domande relative agli spostamenti

8) Quante persone conosci che si sono spostate da una regione all'altra per rimanervi stabilmente dopo l'emanazione del decreto #iorestoacasa (D.P.C.M. 08.03.2020)?

Domanda random per definire il social network

9) Domanda per social network (random)

REFERENCES

1. Bernard HR, Hallett T, Iovita A, et al. Counting hard-to-count populations: the network scale-up method for public health. *Sex Transm Infect.* 2010;86(Suppl_2):ii11-ii15. doi:10.1136/sti.2010.044446
2. Bernard HR, Johnsen EC, Killworth PD, Robinson S. Estimating the size of an average personal network and of an event subpopulation: Some empirical results. *Soc Sci Res.* 1991;20(2):109-121.