



Analysis and Theoretical Exploration of Low Response Rate and Socioeconomic Bias in Survey Non-Response in Milan, Italy

David Consolazio^{1,2}, David Benassi¹, Simone Sarti³, Marco Terraneo¹, Corrado Celata⁴, Antonio Giampiero Russo²

¹Department of Sociology and Social Research, University of Milano-Bicocca, Milan, Italy

²Epidemiology Unit, Agency for Health Protection of the Metropolitan City of Milan, Milan, Italy

³Department of Social and Political Sciences, University of Milan, Milan, Italy

⁴Specific Prevention Unit, Agency for Health Protection of the Metropolitan City of Milan, Milan, Italy

Email: david.consolazio@unimib.it

How to cite this paper: Consolazio, D., Benassi, D., Sarti, S., Terraneo, M., Celata, C. and Russo, A.G. (2024) Analysis and Theoretical Exploration of Low Response Rate and Socioeconomic Bias in Survey Non-Response in Milan, Italy. *Open Access Library Journal*, 11: e11850.

<https://doi.org/10.4236/oalib.1111850>

Received: June 20, 2024

Accepted: July 23, 2024

Published: July 26, 2024

Copyright © 2024 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This article presents the findings of a web survey aimed at collecting new neighborhood-level information on individual and contextual health-related risk factors in Milan, Italy. The study utilized the *Social Determinants of Health* and *Urban Health* approaches to design data-driven health prevention and promotion interventions tailored to the local community. However, the survey experienced a significantly low response rate (6.2%) with a skewed representation of middle-aged, well-educated, white-collar individuals located in affluent city areas. As a result, the collected data was deemed unusable for public health and research purposes. The article discusses the rationale, structure, and development of the project and evaluates the magnitude of non-response and non-response bias due to socioeconomic characteristics. The issues of low survey participation and socioeconomic inequalities in response rates are explored in connection with sociological theories and existing empirical evidence from the literature.

Subject Areas

Sociology

Keywords

Survey Non-Response, Non-Response Bias, Web-Survey, Social Determinants of Health, Urban Health

1. Introduction

In December 2022, a Research Group (RG) led by the Epidemiology Unit of the Agency for Health Protection of the Metropolitan City of Milan (ATS of Milan), in collaboration with three University Departments from Milanese Universities (University of Milano-Bicocca, University of Milan, and Polytechnic of Milan), conducted a survey targeting a random sample of the Milanese population. The sample size consisted of 40,000 individuals, stratified based on their neighborhood of residence. The objective of the survey was to gather information on health-related behaviors and risk factors (e.g. diet regime, sedentary behaviours and physical activity, alcohol use, smoking habits, gambling, and neighbourhood stressors) that were previously lacking at the municipal level or below. This data would have facilitated the development and implementation of health prevention and promotion interventions, both at the individual and the contextual level. The survey design was guided by the framework of *Social Determinants of Health* [1] and *Urban Health* [2], incorporating both individual and contextual-level risk factors. The aim was to empirically examine the territorial variations in proximate risk factors (e.g., unhealthy diet, smoking, alcohol consumption, physical inactivity) as well as distal socioeconomic factors (e.g., education, occupation), urban morphology, and the quality of the living environment (e.g., air and noise pollution, road traffic, perceived safety), underscoring the importance of examining both individual and contextual risk factors to address health disparities. Our study aimed to bridge the gap in local health data in Milan by collecting detailed neighborhood-level information, thereby enabling more targeted and effective public health interventions. This approach is supported by extensive literature that highlights the significance of localized health data in understanding and mitigating health inequities [3]. However, for various reasons, the response rate to the survey was much lower than anticipated, at 6.2%. Moreover, there was a significant imbalance in responses between higher and lower socioeconomic strata, rendering the collected data unusable for public health and research purposes. Despite the failure in data collection, the obtained information proved valuable in exploring methodological and conceptual issues related to survey participation, power, and knowledge. Furthermore, it sheds light on the perpetuation of social and health inequalities at both individual and neighbourhoods levels.

In this paper, we present the objectives and structure of the project, and introduce the theoretical framework that guided the data collection. We also discuss the current limitations in data availability within the Italian context. We then illustrate the sampling procedure, questionnaire design, and the practical steps involved in survey delivery. Next, we compare the survey population to the reference population to assess the extent of non-response and its associated socioeconomic patterns. We also provide theoretical considerations regarding the potential mechanisms leading to low survey participation and socioeconomic inequalities in response rates.

2. The Project

The project received funding from Fondazione Cariplo, a charitable foundation that supports public, private, and third-sector organizations in implementing socially beneficial projects and research. The Epidemiology Unit of the ATS of Milan established a Research Group (RG) in collaboration with three departments from Milanese universities in response to the call “Data Science for Science and Society. The units involved were the Department of Sociology and Social Research of the University of Milano-Bicocca, the Department of Political, Economic, and Social Sciences of the University of Milan, and the Department of Architecture, Built Environment and Construction Engineering of the Polytechnic of Milan; each unit of the RG oversaw specific tasks leading to the development of a *Computer-Assisted Web-Interviewing* (CAWI) questionnaire to be filled by a sample of Milanese citizens, invited by means of a postcard delivered at the residence address and containing a QR code to access the online platform. The project aimed at collecting information about the territorial distribution of health-related risk factors across the city neighbourhoods to gain valuable information to design locally tailored public health interventions. Linking the information gathered with ecological data on the distribution of non-communicable diseases deriving from the administrative healthcare databases of the ATS of Milan, and with built environmental open data from the municipality of Milan, it would have been possible to estimate the association between health-related behaviours, neighbourhood characteristics, and health outcomes.

2.1. Theoretical Framework

A substantial body of literature has consistently demonstrated the existence of a social gradient in health, indicating a strong link between socioeconomic status (SES) and health outcomes [1]. Individuals from lower socioeconomic groups tend to experience higher rates of morbidity, mortality, and poorer physical and mental well-being compared to those from higher socioeconomic classes [4]. The distribution of the population within a city is not random but reflects the distribution of socioeconomic resources within the population itself [5]. Consequently, social stratification is evident in the geographic distribution of residents, with individuals from disadvantaged backgrounds often residing in areas of lower economic value, situated further away from valuable services and amenities [6]. However, grouping within the same areas of individuals at greater health risk is not the only cause for the identification of urban clusters in worse health conditions, as living environment features might have a positive or negative on their residents' health [7] [8]. This extends beyond physical environmental characteristics, such as air pollution, and encompasses features of the urban environment like walkability, access to green spaces, provision of public transportation, and the quality of the food environment, that can shape individuals' exposure to the risk of developing various chronic diseases through a range

of mechanisms [7] [8]. Recognizing the importance of population health from an urban perspective, approaches have emerged that specifically aim to understand the phenomenon and develop strategies for health protection and promotion that go beyond individual-focused interventions, focusing on place characteristics and the way they shape risk exposure, such as the *Urban Health* approach [2].

2.2. State of the Art in Italy

Effective implementation of local planning for urban regeneration, with a focus on meeting the health needs of the population, requires accurate knowledge of those needs. This emphasizes the importance of having detailed and reliable information on health status, risk factors, and characteristics of the living environment. Such information enables an accurate assessment of the phenomenon, taking into account the territorial diversity within the urban context. Unfortunately, in Italy, including the case of Milan, the availability of data to support such an approach is limited. Nationally, sample surveys do not collect information at the necessary spatial detail, and there is a lack of secondary data archives from administrative sources such as local health units and municipalities. The available sources generally provide data on health-risky behaviours only at the provincial or district level, with accessing municipal-level information being unfeasible in most cases. On the other hand, administrative healthcare databases contain substantial individual health information that can be geographically referenced to various local units. These databases are currently used to create indicators for assessing health outcomes and the performance of health facilities. However, they do not contain information on risk factors and socioeconomic characteristics. To address this data gap, the prevalence of behaviours and lifestyles is often inferred from the prevalence of diseases correlated to them, using biomedical indicators as proxies for risk factors. However, this approach has important limitations. A disease can be associated with multiple risk factors, making it difficult to determine which one is the main determinant of the examined condition. For example, a high prevalence of respiratory diseases in a neighbourhood may be attributed to smoking, air quality, or exposure to toxic agents in the workplace, which could also be located elsewhere. Similarly, various risk factors can contribute to the onset of multiple diseases. For instance, an unhealthy diet and physical inactivity can increase the risk of both cardiovascular disease and diabetes. The lack of a clear and direct correspondence between risk factors and diseases suggests that inferring the prevalence of a specific risk factor in a particular area solely from the prevalence of correlated diseases should be done cautiously, especially when more unbiased alternatives are unavailable or its correlated diseases only in the absence of less biased alternatives.

2.3. Sampling Procedure

To ensure representativeness of the sample at both the municipal and neigh-

bourhood levels, a cluster sampling approach was employed. This involved selecting a random population sample within each neighbourhood, while also ensuring representation in terms of age, sex, and citizenship. The administrative subdivision of the municipality of Milan into 88 NILs (*Nuclei di Identità Locale, Local Identity Cores*) was used as the territorial units. NILs can be defined as historically or project-defined neighbourhoods, each with specific characteristics that distinguish them from one another. The selection of individuals invited to participate in the survey was based on information available in the civil registry. To achieve a desired level of precision, assuming an average population of 11,517 individuals aged between 18 and 75 in each area, a two-sided confidence interval at 95% with a margin of error of 5%, it was necessary to sample an average of 372 individuals per NIL. This resulted in a total sample size of 32,736 subjects.

More in detail, the average neighbourhood sample size was calculated using Cochran's formulas [9]:

$$n = \frac{z^2 * \hat{p}(1 - \hat{p})}{\varepsilon^2}$$

$$n' = \frac{n}{1 + \frac{z^2 * \hat{p}(1 - \hat{p})}{\varepsilon^2 * N}}$$

where n is the sample size to be calculated under the assumption of an unlimited population, to be used in the following step to calculate the sample size for a finite population, which is the value of interest to be calculated, n' ; z is the z -score, which value for a 5% margin of error was set to 1.96; \hat{p} is the population proportion set to 0.5; ε is the margin of error set at 0.05; and N is the population size, namely the average population per neighbourhood, corresponding to 11,517 units. Substituting the values in the formulas:

$$n = \frac{1.96^2 * 0.50(1 - 0.50)}{0.05^2} = 384$$

$$n' = \frac{384}{1 + \frac{1.96^2 * 0.50(1 - 0.50)}{0.05^2 * 11517}} = 372$$

The final sample size (32,736) was determined by multiplying the calculated value by the number of NILs. We performed oversampling leading to a final extraction of 40,000 subjects to increase the likelihood of getting a representative sample of the target population. The extraction process was carried out in the SAS environment using the PROC SURVEYSELECT procedure. The code was programmed to randomly select 40,000 cases, ensuring that the selection was proportionate to the number of residents in each NIL (using the proportional allocation method), which led to a minimum selection of 4 subjects in the NIL “Giardini Porta Venezia”, which is an urban park with 43 residents, to a maximum selection of 1745 subjects in the NIL “Buenos Aires—Porta Venezia—Porta Monforte”, the most inhabited NIL with 62,371 residents.

2.4. Questionnaire Design

The questionnaire was developed using previously validated items that are commonly used in national and international individual and household surveys. To minimize participant dropout and enhance response rates, the questionnaire was designed to be concise and efficient, consisting of a maximum of 26 questions, organized into four sections:

1) *Personal data*: age, sex, citizenship, educational level, occupational condition, occupational class.

2) *Neighbourhood environment*: perceived characteristics of the neighbourhood of living (public transport, traffic, air, noise and odour pollution, litter, crime, street lighting), social cohesion in the context of residence.

3) *Health and lifestyle*: self-perceived health, height and weight, diet regime, sedentary behaviours and physical activity, alcohol use, smoking habits, gambling.

4) *Access to services*: forgo medical examination or drug purchase in case of need for any reason.

2.5. Survey Setting and Delivery

The development of the online platform for survey administration and data storage, as well as the production and delivery of invitation postcards, were outsourced to an external agency affiliated with the Italian postal service provider. The postcards contained a brief description of the project's objectives and invited recipients to scan the QR code or copy the provided link to access a short introductory video, which featured a well-known Italian TV presenter to increase the response rate. At the end of the video, participants were encouraged to complete the questionnaire using any device they had available, such as a smartphone, tablet, or computer. The entire process of answering the questionnaire typically took between 5 to 10 minutes. A press conference was organized to give the activity higher visibility and communication was disseminated through the official social media channel of the ATS of Milan. The data collection phase commenced on December 5, 2022, and concluded on January 26, 2023. It is worth noting that despite the data collection method fell under the CAWI category, the engagement was made through the traditional mail method, given the possibility to reach the sampled individuals exclusively through their residential address. All the data collected remained anonymous, and no identifying values could link the information to the participants. Each QR code was associated with one of the 88 NILs to avoid the need for respondents to provide their residential addresses, as self-reported neighbourhood identification can be subjective and subject to bias [10] [11].

3. Results

Out of the 40,000 invitations sent, only 2472 individuals completed the questionnaire, resulting in a response rate of 6.2%. Around 90% of those who

scanned the QR code accessed and completed the questionnaire, indicating that the low response rate was primarily due to individuals disregarding the postcard and not scanning the code, rather than a high dropout rate. The distribution of key variables indicates the presence of a noticeable self-selection bias, with a majority of responses coming from individuals in the medium-high/high social class. As shown in **Table 1**, the sample exhibited a relatively balanced distribution by sex (54.8% female, 44.3% male). However, there was a clear underrepresentation of non-Italian residents, comprising only 2.6% of survey respondents compared to 17.8% in the target population. This deviation was expected since all the materials provided were in Italian, with English subtitles for the introductory video.

Table 1. Comparative statistics of respondent population (n = 2472) vs sampled population (n = 40,000) for sex and citizenship.

	Survey respondents		Survey sample	
	<i>n</i>	%	<i>n</i>	%
Sex				
<i>Female</i>	1354	54.8	20,863	52.2
<i>Male</i>	1095	44.3	19,137	47.8
<i>Other/Prefer not to say</i>	23	0.9	-	-
Citizenship				
<i>Italian</i>	2407	97.4	32,897	82.2
<i>Foreigner</i>	65	2.6	7103	17.8

Considering age, the respondents were underrepresented in the sampled population until the age of 50 and after the age of 66 - 70, being conversely overrepresented between the ages of 51 and 65 (**Figure 1**). The median age among respondents (51 years) was slightly higher compared to the sampled population (49).

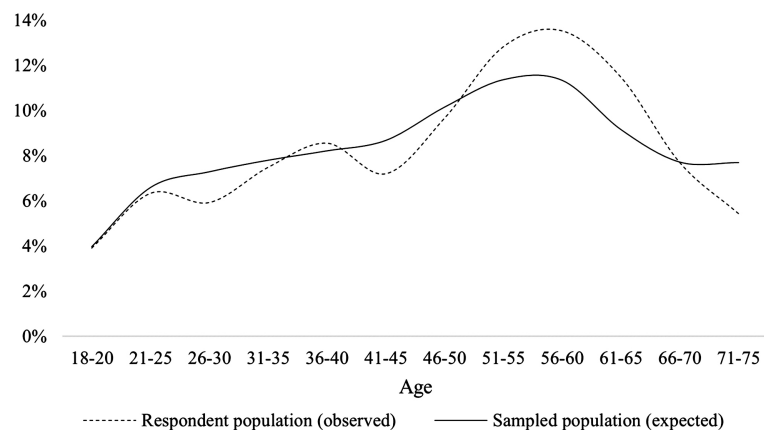


Figure 1. Comparative statistics of standardized distributions of respondent population (n = 2472) and sampled population (n = 40,000) by age groups.

As for the remaining personal information, we were unable to directly compare the respondents with the sample population due to the lack of education and occupation data in the civil registry. Therefore, we conducted a comparison with official statistics using both census and survey data. Regarding educational achievement, while there was alignment in the presence of individuals with upper-secondary education (35.6%), there was a clear imbalance in relation to other categories. Low-educated respondents were considerably underrepresented (5.7% compared to 34.1%), while graduates were overrepresented (58.2% compared to 30.3%). In terms of occupational status, there was an overrepresentation of employed individuals (64.0% compared to 52.3%), along with an underrepresentation of the unemployed (2.8% compared to 5.8%), domestic workers (4.4% compared to 7.2%), and retirees (16.5% compared to 22.0%). However, the percentage of students was similar between the survey and the census population (7.2% compared to 7.4%). Occupational position was compared with data from the Istat Labour Force Survey. Among the self-employed, entrepreneurs were significantly overrepresented (4.2% compared to 1.9%), while sole traders (low-skilled business owners) were clearly underrepresented (2.1% compared to 8.5%). Looking at employees, manual workers were much less represented in the survey (5.7%) compared to Istat data (22.3%), while high-skilled and managerial positions such as employees with managerial responsibilities (31.7% compared to 11.1%) and executive, manager, or employed professionals (13.0% compared to 4.9%) were overrepresented. In summary, the majority of respondents were engaged in white-collar jobs, particularly highly specialized and/or autonomous positions. The overall picture suggests a sample that is biased towards a relatively middle-aged Italian population of white-collar workers with high educational achievements and a good socioeconomic position (See [Table 2](#)).

Table 2. Comparative statistics between respondent population ($n = 2472$) and census population (education $n = 1,253,402$; occupation $n = 1,229,560$) for educational attainment and occupational status, and between respondent population ($n = 1989$) and 2018 Istat Labour Force Survey ($n = 407,069$) for occupational position.

	Survey respondents		Official data	
	<i>n</i>	%	<i>n</i>	%
Education				
<i>No education</i>	3	0.1	46,435	3.7
<i>Primary education</i>	2	0.1	119,866	9.6
<i>Lower-secondary education</i>	137	5.5	260,940	20.8
<i>Upper-secondary education</i>	880	35.6	446,137	35.6
<i>Tertiary education</i>	1438	58.2	380,024	30.3
<i>Other/Prefer not to say</i>	12	0.5	-	-
Occupational status				
<i>Employed</i>	1582	64.0	642,990	52.3

Continued

<i>Unemployed</i>	70	2.8	70,850	5.8
<i>Domestic worker</i>	110	4.4	88,720	7.2
<i>Student</i>	179	7.2	90,620	7.4
<i>In retirement</i>	407	16.5	270,771	22.0
<i>Other/ Prefer not to say</i>	124	5.0	65,609	5.3
Occupational position				
<i>Self-employed</i>				
<i>Sole trader</i>	42	2.1	54,460	8.5
<i>Freelancer</i>	255	12.8	85,063	13.3
<i>Entrepreneur</i>	84	4.2	12,258	1.9
<i>Employees</i>				
<i>Manual worker</i>	112	5.7	143,025	22.3
<i>Office clerk</i>	527	26.5	234,424	36.6
<i>Employee with managerial responsibilities</i>	630	31.7	71,128	11.1
<i>Executive, manager or employed professional</i>	259	13.0	25,613	4.0
<i>Unknown/Prefer not to say</i>	80	4.0	15,288	2.4

Figure 2 illustrates the distribution of response rates across NILs, aiming to capture variations and compare them with the territorial socioeconomic heterogeneity. The response rate is calculated by dividing the number of respondents by the number of people sampled in each NIL. To minimize distortions caused by small sample sizes leading to high rates, adjustments were made. This involved multiplying the raw response rate by the ratio of respondents in a NIL to the overall number of respondents. The NIL-level deprivation index was computed using data from the latest Italian census, specifically using Rosano and colleagues' index [12]. This index combines five indicators: the percentage of individuals (aged 15 - 60) with at most primary education, the percentage of unemployed individuals (aged 15 - 60), the percentage of households that rent their homes, the percentage of single-parent families with underage children, and the average housing crowding (number of inhabitants per 100 m²) in each NIL. Reminding the overall low response, the map also reveals territorial heterogeneity, with lower response rates observed in the most peripheral NILs and higher response rates in semi-central neighbourhoods located just outside the historical centre. Considering the well-documented territorial dualism characterizing Milan, with wealthier groups residing in the centre and more disadvantaged individuals living in the outskirts, this disparity in responses aligns with (SES). There is a partial overlap between areas with higher response rates and the least deprived areas, and vice versa.

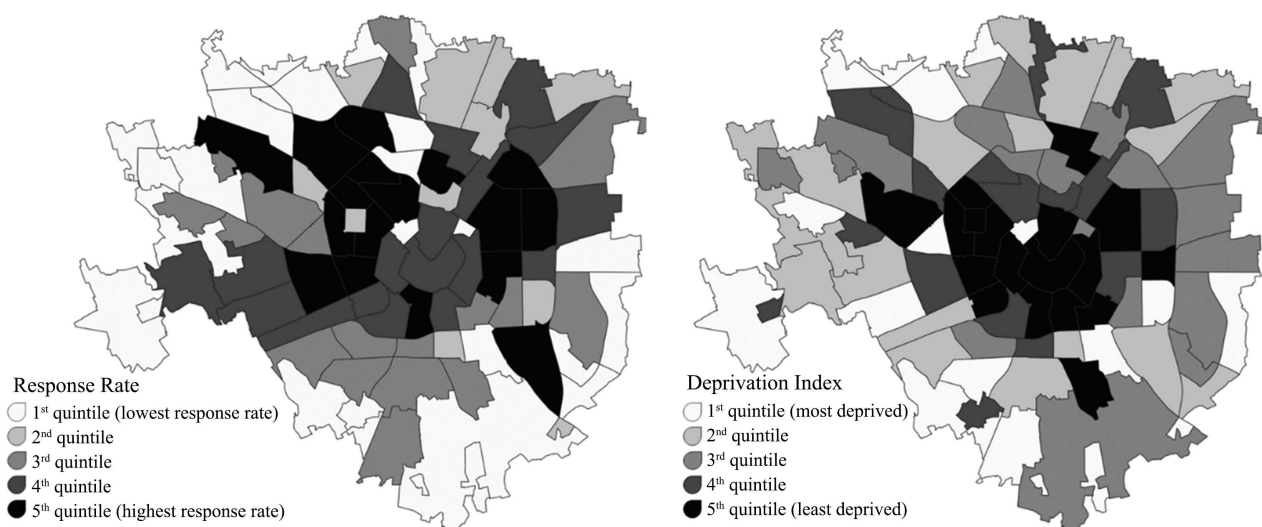


Figure 2. Adjusted response rate and deprivation index in each territorial unit (NILs), quintile distribution.

4. Discussion

Several critical aspects emerged from our study on survey participation and its implications for health research. Each of the following sections addresses distinct challenges and influences affecting data collection and interpretation. We delve into the impact of survey modes on response rates, the presence of socioeconomic bias in participant demographics, the role of power relations and civic engagement in shaping community involvement, the influence of institutional mistrust on survey participation, and strategies for mitigating SES bias in survey research. These topics are essential for understanding the complexities of survey research within the context of health interventions and population health [13].

4.1. Survey Mode and Response Rate: Impact on Data Collection

Although not anticipating significant engagement from the local population, the survey had an exceptionally low response rate, rendering the collected data useless for the intended study purposes. One possible reason for the challenge in achieving an adequate sample size could be attributed to the chosen survey mode for data collection. Research has shown that face-to-face surveys yield the highest response rates, followed by telephone surveys, and then mail and web surveys [14]. Among mail and web surveys, a meta-analysis of 39 studies found that mail surveys generally have higher response rates compared to web surveys [15]. However, web-based surveys are commonly employed as an alternative to traditional paper surveys due to their design flexibility, cost-effectiveness, and reduced data entry time [16]. In our case, initial engagement was conducted via mail, and respondents were directed to a web page using a QR code/link provided with the postcard, resulting in a mixed-mode survey. Access barriers associated with both methods, such as lack of interest in the received communication material or technological barriers to using digital devices, may have influenced the response rate. Additionally, the lack of a follow-up reminder due to privacy

concerns (as no personal data were collected for contacting non-respondents) likely impacted the response rate. Follow-up reminders have been proven effective in increasing responses to postal and electronic questionnaires [17]. Budget constraints also prevented the use of incentives, which can help reduce non-response [18] [19].

4.2. SES Bias in Survey Participation: Implications for Data Representativeness

Beyond the limited number of responses, an important issue arises concerning the (SES) bias in response rates. Ignoring this bias, even with acceptable response rates, can lead to biased findings and further exacerbate social and health inequalities [20]. In our data, most answers came from middle-aged, well-educated, white-collar individuals residing in affluent neighbourhoods. SES bias in survey non-response is a well-documented problem in web-based surveys [20]-[24]. While web surveys offer advantages over traditional methods, they can still face challenges, as participants who respond to surveys tend to be more motivated, technologically savvy, prone to civic participation, and actively engaged in the topic being studied, leading to self-selection bias [25]. Demographics are known to exert an influence on survey initiation and completion. Though gender did not influence participation, age appeared to have influenced the answering process. Internet accessibility and usage might be a mediator in this relationship, as research has shown that in Italy internet usage varies consistently across age groups, with 92% of subjects aged 16 - 24 using the internet, compared to 69% of adults aged 55 - 65 [26]. Similar variations were reported also in relation to educational level (tertiary education: 92%; upper secondary education: 85%; lower secondary/primary education: 60% [27]. A first source of difference in response rates could be represented by *electronic literacy* and its associated *grey digital divide* [28] [29]. This is consistent with our drop in responses after the age of 65, which might be due to lower internet usage, and it is also reflected in the nearly total absence of respondents with at most primary education, realistically the oldest population group. Additionally, studies have explored survey participation disparities based on race/ethnicity though these are not known to be directly related to internet usage [20]. Concerning our case, we already drew attention to how the extremely low participation of non-Italian residents was unavoidable and expected, due to the impossibility for non-Italian speakers to select the desired native language in the web platform developed.

4.3. Power Relations and Civic Engagement: Influences on Community Participation in Health Research

Beyond socio-demographic and technological barriers, other factors such as power relations, knowledge, social cohesion, and civic engagement may have played a significant role in influencing participation in the study. The primary objective of the research was to gather information to inform the development of health prevention and promotion interventions, with a particular emphasis on

understanding the contextual dimension. For example, recognizing that a particular neighbourhood has a high prevalence of metabolic disorders, partly attributed to the existence of an obesogenic food environment, highlights the critical importance of implementing targeted interventions to enhance the local environment by promoting physical activity and encouraging healthier dietary choices, irrespective of individual motivation or willingness. In a way, participating in the survey can serve as a unique opportunity to contribute to the improvement of one's residential context by providing bottom-up accounts of the most significant issues impacting the local environment. This claim is supported by numerous studies that demonstrate the impact of community engagement on urban planning and health outcomes. For instance, research shows that community input is crucial in identifying and addressing local health needs, leading to more effective and sustainable interventions [29] [30]. By involving residents in the data collection process, our study wished to empower communities and ensure that public health strategies are aligned with their specific needs and preferences. It is worth noting that interventions targeting urban development often concentrate in the more affluent areas of the city to enhance their appeal, resulting in the attraction of wealthier groups, tourism promotion, and increased revenues for a select few. Consequently, this may lead to the displacement of vulnerable populations to less desirable [31]-[33]. Accordingly, neighbourhood-level interventions that are not based on empirical evidence may contribute to improving certain local areas, but at the cost of widening socioeconomic and health inequalities across the entire city [34]. On the contrary, data-driven interventions based on issues highlighted directly by the residents themselves have the potential to "give voice" to less powerful and influential individuals who are typically excluded from decision-making processes that shape their living environment. Instead of blaming individuals with lower SES for not participating and missing an opportunity to improve their conditions, it is important to acknowledge the varying capabilities of individuals from different social groups in mobilizing strategic resources to their advantage. Indeed, among the mechanisms linking social cohesion to individual health, Kawachi [35] emphasized the ability to undertake collective action (e.g., the ability of a community to organize to protest the closure of a local hospital or the use of zoning restrictions to prevent the incursion of fast-food outlets), commonly referred to as *collective efficacy* [36]. This perspective aligns with the local application of *power resource theory* [37], which suggests that the provision of social welfare is influenced by power differentials between social classes, so that in countries where working-class individuals are more politically and collectively organized, socioeconomic inequalities tend to be lower.

4.4. Mistrust in Institutions: Implications for Survey Participation and Data Quality

According to the *theory of fundamental causes of health inequalities* [4], individuals deploy a wide range of resources – including knowledge, money, power,

prestige, and beneficial social connections – that can be used individually or collectively to maximise their well-being. Indeed, it is known that low SES individuals and ethnic minorities are less likely to enrol in medical research (e.g., screening, clinical trials), primarily due to medical mistrust [38]-[40], leveraged by lack of information, limited awareness, and understanding of the availability and value of medical research. Similarly, we argue that institutional distrust may have played a role in limiting participation in our study. The distribution of postcards coincided with a period of minimal institutional consensus in Italy, stemming from the containment measures implemented during various stages of the COVID-19 pandemic [41]-[44], as well as the Italian political crisis that occurred in July 2022 and the repercussions of global events such as international conflicts and energy crisis, which led to rising inflation. It is plausible that this widespread mistrust of societal institutions impeded civic engagement [45], and affected survey participation, which was already challenging. The fact that civicness follows a social gradient [46] may partially explain the significant SES bias observed in our survey.

Prior research has shown that addressing facilitators (e.g., benefits to participants) and barriers (e.g., mistrust, stigma, and competing demands) can help improve the recruitment of underrepresented populations and some have posited the use of web-survey might mitigate the issue of disproportionate representation among different populations [47]. Regarding benefits to participants, interesting contributions come from the application of *social exchange theory* to the case of survey participation, according to which social behaviour is the outcome of an exchange process based on maximizing benefits and minimizing costs [48]. In this view, incentives are often adopted to increase response rates, although their interaction with SES is often overlooked. Research has shown that non-monetary incentives tend to be more effective for individuals with higher education levels, while monetary incentives are associated with smaller SES differences and generally outperform non-monetary incentives [49]. To address non-response bias, an additional strategy involves over-sampling groups that are expected to have lower participation rates based on their sociodemographic characteristics, counterbalancing the effects of non-response. Furthermore, post-survey adjustments can be made by applying sampling weights to align the characteristics of the respondents with the sampled population [50]. However, the lack of socioeconomic information in the Italian civil registries prevented the adoption of these strategies.

4.5. Addressing SES Bias in Survey Research: Challenges and Opportunities

Despite its importance, the existing literature on SES bias in survey non-response is limited. The available material primarily focuses on quantitatively assessing the characteristics of individuals and groups that are less likely to participate in surveys and determining the impact of non-response bias on the representative-

ness of research results. There have been seminal theoretical frameworks proposed to understand the decision-making processes involved in survey participation [51] and practical guidance on addressing non-response and non-response bias [52]. However, there has been comparatively less attention given to exploring the individual and structural mechanisms that contribute to non-response differences based on SES. In the literature, significant efforts have been made to engage “hard-to-reach” populations, which typically lack a sampling frame and face various structural barriers such as language, literacy, cultural differences, power imbalances, and legal issues, making the sampling and enrolment process challenging [53]. However, this is not the case for low SES individuals who are present in population registries and can be easily reached for survey invitations. The lower participation of low SES individuals is influenced by a combination of individual predispositions, beliefs, attitudes, and structural constraints. Therefore, while this contribution is inevitably limited, it represents one of the initial attempts to explore and frame the issue within a broader theoretical framework, serving as a starting point for further empirical investigations.

5. Conclusion

In this paper we presented a project based on a primary data collection by means of a web survey administered to a random sample of the Milanese population, representative at the neighbourhood level. For several reasons, the data collection led to noteworthy non-response and non-response bias problems, making the data gathered unusable for purposes of locally tailored health promotion and prevention interventions aimed at increasing population well-being and reducing social inequalities in health, particularly focusing on the territorial dimension. By demonstrating the extent of SES bias in our data, we have highlighted the methodological and conceptual pitfalls that can hinder successful data collection, also proposing reflections based on sociological theories that can help prevent such distortions in future research. It is important to recognize that even with acceptable response rates, failing to account for the unequal distribution of respondents across social strata can have detrimental consequences, reinforcing and widening existing socioeconomic and health inequalities if policy interventions are based on information deriving from the unequal balance of power, knowledge, and resources existing between different socioeconomic groups.

Fundings

The study has been conducted in the context of the project ‘Enhancing health-care and well-being through the potential of big data: an integration of survey, administrative, and open data to assess health risk in the City of Milan with data science’ (Project Code 2020-4269) funded by Fondazione Cariplo (<https://www.fondazione-cariplo.it/en/index.html>) within the call “Data Science for science and society”. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Ethical Statement

The study protocol was submitted for ethical approval to the Ethical Committee “Milano Area 2” in charge c/o “Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico di Milano” and approved on May 10th 2022, ID experimentation n. 2804, study n. 6156.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Marmot, M. (2005) Social Determinants of Health Inequalities. *The Lancet*, **365**, 1099-1104. [https://doi.org/10.1016/s0140-6736\(05\)71146-6](https://doi.org/10.1016/s0140-6736(05)71146-6)
- [2] Fehr, R. and Capolongo, S. (2016) Healing Environment and Urban Health. *Epidemiologia & Prevenzione*, **40**, 151-152.
- [3] Marques, T.S., Ferreira, M., Saraiva, M., Forte, T. and Santinha, G. (2021) Mapping Health Vulnerabilities: Exploring Territorial Profiles to Support Health Policies. *Ciência & Saúde Coletiva*, **26**, 2459-2470. <https://doi.org/10.1590/1413-81232021266.1.40862020>
- [4] Phelan, J.C., Link, B.G. and Tehranifar, P. (2010) Social Conditions as Fundamental Causes of Health Inequalities: Theory, Evidence, and Policy Implications. *Journal of Health and Social Behavior*, **51**, S28-S40. <https://doi.org/10.1177/0022146510383498>
- [5] Oberti, M. and Préteceille, E. (2016) La ségrégation urbaine. *La Découverte*. <https://doi.org/10.3917/dec.edmon.2016.01>
- [6] Pearce, J., Witten, K., Hiscock, R. and Blakely, T. (2008) Regional and Urban-Rural Variations in the Association of Neighbourhood Deprivation with Community Resource Access: A National Study. *Environment and Planning A: Economy and Space*, **40**, 2469-2489. <https://doi.org/10.1068/a409>
- [7] Galster, G.C. (2011) The Mechanism(s) of Neighbourhood Effects: Theory, Evidence, and Policy Implications. In: van Ham, M., et al., Eds., *Neighbourhood Effects Research: New Perspectives*, Springer, 23-56. https://doi.org/10.1007/978-94-007-2309-2_2
- [8] Macintyre, S., Maciver, S. and Sooman, A. (1993) Area, Class and Health: Should We Be Focusing on Places or People? *Journal of Social Policy*, **22**, 213-234. <https://doi.org/10.1017/s0047279400019310>
- [9] Cochran, W.G. (1977) *Sampling Techniques*. John Wiley & Sons.
- [10] Pebley, A. and Sastry, N. (2009) Our Place: Perceived Neighborhood Size and Names in Los Angeles. UCLA CCPR Population Working Papers. <http://128.97.186.17/index.php/pwp/article/view/839/222>
- [11] Weiss, L., Ompad, D., Galea, S. and Vlahov, D. (2007) Defining Neighborhood Boundaries for Urban Health Research. *American Journal of Preventive Medicine*, **32**, S154-S159. <https://doi.org/10.1016/j.amepre.2007.02.034>
- [12] Rosano, A., Pacelli, B., Zengarini, N., Costa, G., Cislighi, C. and Caranci, N. (2020) Update and Review of the 2011 Italian Deprivation Index Calculated at the Census Section Level. *Epidemiologia e Prevenzione*, **44**, 162-170.
- [13] Dimick, S., Stefanacci, R.G. and Gugliucci, M.R. (2022) Addressing Social Determinants of Health (SDoH) in Primary Care: A Pilot Study. *Open Journal of Social Sciences*, **10**, 233-246. <https://doi.org/10.4236/jss.2022.104017>

- [14] Groves, R.M., Fowler Jr., F.J., Couper, M.P., Lepkowski, J.M., Singer, E. and Tourangeau, R. (2011) *Survey Methodology*. John Wiley & Sons.
- [15] Shih, T. and Fan, X.T. (2008) Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis. *Field Methods*, **20**, 249-271.
<https://doi.org/10.1177/1525822x08317085>
- [16] Fan, W. and Yan, Z. (2010) Factors Affecting Response Rates of the Web Survey: A Systematic Review. *Computers in Human Behavior*, **26**, 132-139.
<https://doi.org/10.1016/j.chb.2009.10.015>
- [17] Edwards, P.J., Roberts, I., Clarke, M.J., DiGuseppi, C., Wentz, R., Kwan, I., et al. (2009) Methods to Increase Response to Postal and Electronic Questionnaires. *Cochrane Database of Systematic Reviews*, **2009**, MR000008.
<https://doi.org/10.1002/14651858.mr000008.pub4>
- [18] Ryu, E., Couper, M.P. and Marans, R.W. (2006) Survey Incentives: Cash vs. In-Kind; Face-to-Face vs. Mail; Response Rate vs. Nonresponse Error. *International Journal of Public Opinion Research*, **18**, 89-106.
<https://doi.org/10.1093/ijpor/edh089>
- [19] Singer, E. and Ye, C. (2012) The Use and Effects of Incentives in Surveys. *The ANNALS of the American Academy of Political and Social Science*, **645**, 112-141.
<https://doi.org/10.1177/0002716212458082>
- [20] Jang, M. and Vorderstrasse, A. (2019) Socioeconomic Status and Racial or Ethnic Differences in Participation: Web-Based Survey. *JMIR Research Protocols*, **8**, e11865.
<https://doi.org/10.2196/11865>
- [21] Blumenberg, C., Zugna, D., Popovic, M., Pizzi, C., Barros, A.J.D. and Richiardi, L. (2018) Questionnaire Breakoff and Item Nonresponse in Web-Based Questionnaires: Multilevel Analysis of Person-Level and Item Design Factors in a Birth Cohort. *Journal of Medical Internet Research*, **20**, e11046.
<https://doi.org/10.2196/11046>
- [22] Couper, M.P., Kapteyn, A., Schonlau, M. and Winter, J. (2007) Noncoverage and Nonresponse in an Internet Survey. *Social Science Research*, **36**, 131-148.
<https://doi.org/10.1016/j.ssresearch.2005.10.002>
- [23] Ekholm, O., Gundgaard, J., Rasmussen, N.K.R. and Hansen, E.H. (2010) The Effect of Health, Socio-Economic Position, and Mode of Data Collection on Non-Response in Health Interview Surveys. *Scandinavian Journal of Public Health*, **38**, 699-706.
<https://doi.org/10.1177/1403494810382474>
- [24] Goyder, J., Warriner, K. and Miller, S. (2002) Evaluating Socio-Economic Status (SES) Bias in Survey Nonresponse. *Journal of Official Statistics-Stockholm*, **18**, 1-12.
- [25] Khazaal, Y., van Singer, M., Chatton, A., Achab, S., Zullino, D., Rothen, S., et al. (2014) Does Self-Selection Affect Samples' Representativeness in Online Surveys? An Investigation in Online Video Game Research. *Journal of Medical Internet Research*, **16**, e164. <https://doi.org/10.2196/jmir.2759>
- [26] Eurostat (2021) EU Survey on the Use of Information and Communication Technologies (ICT) in Households and by Individuals.
https://ec.europa.eu/eurostat/cache/metadata/en/isoc_i_esms.htm
- [27] Morris, A. and Brading, H. (2007) E-Literacy and the Grey Digital Divide: A Review with Recommendations. *Journal of Information Literacy*, **1**, 13-28.
<https://doi.org/10.11645/1.3.14>
- [28] Sala, E., Gaia, A. and Cerati, G. (2020) The Gray Digital Divide in Social Networking Site Use in Europe: Results from a Quantitative Study. *Social Science Computer Review*, **40**, 328-345. <https://doi.org/10.1177/0894439320909507>

- [29] Israel, B.A., Schulz, A.J., Parker, E.A. and Becker, A.B. (1998) Review of Community-Based Research: Assessing Partnership Approaches to Improve Public Health. *Annual Review of Public Health*, **19**, 173-202. <https://doi.org/10.1146/annurev.publhealth.19.1.173>
- [30] Minkler, M. and Wallerstein, N. (2011) Community-Based Participatory Research for Health: From Process to Outcomes. Jossey-Bass.
- [31] Jelks, N.O., Jennings, V. and Rigolon, A. (2021) Green Gentrification and Health: A Scoping Review. *International Journal of Environmental Research and Public Health*, **18**, Article No. 907. <https://doi.org/10.3390/ijerph18030907>
- [32] Mehdipanah, R., Marra, G., Melis, G. and Gelormino, E. (2017) Urban Renewal, Gentrification and Health Equity: A Realist Perspective. *European Journal of Public Health*, **28**, 243-248. <https://doi.org/10.1093/eurpub/ckx202>
- [33] Schnake-Mahl, A.S., Jahn, J.L., Subramanian, S.V., Waters, M.C. and Arcaya, M. (2020) Gentrification, Neighborhood Change, and Population Health: A Systematic Review. *Journal of Urban Health*, **97**, 1-25. <https://doi.org/10.1007/s11524-019-00400-1>
- [34] Cole, H.V.S., Garcia Lamarca, M., Connolly, J.J.T. and Anguelovski, I. (2017) Are Green Cities Healthy and Equitable? Unpacking the Relationship between Health, Green Space and Gentrification. *Journal of Epidemiology and Community Health*, **71**, 1118-1121. <https://doi.org/10.1136/jech-2017-209201>
- [35] Kawachi, I. (2010) Social Capital and Health. In: Bird, C.E., Conrad, P., Fremont, A.M. and Timmermans, S., Eds., *Handbook of Medical Sociology*, 6th Edition, Vanderbilt University Press, 18-32. <https://doi.org/10.2307/j.ctv16h2n9s.5>
- [36] Sampson, R.J., Raudenbush, S.W. and Earls, F. (1997) Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy. *Science*, **277**, 918-924. <https://doi.org/10.1126/science.277.5328.918>
- [37] O'Connor, J.S. and Olsen, G.M. (1998) Power Resource Theory and the Welfare State: A Critical Approach. University of Toronto Press.
- [38] George, S., Duran, N. and Norris, K. (2014) A Systematic Review of Barriers and Facilitators to Minority Research Participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. *American Journal of Public Health*, **104**, e16-e31. <https://doi.org/10.2105/ajph.2013.301706>
- [39] Scharff, D.P., Mathews, K.J., Jackson, P., Hoffsuemmer, J., Martin, E. and Edwards, D. (2010) More than Tuskegee: Understanding Mistrust about Research Participation. *Journal of Health Care for the Poor and Underserved*, **21**, 879-897. <https://doi.org/10.1353/hpu.0.0323>
- [40] Sharrocks, K., Spicer, J., Camidge, D.R. and Papa, S. (2014) The Impact of Socioeconomic Status on Access to Cancer Clinical Trials. *British Journal of Cancer*, **111**, 1684-1687. <https://doi.org/10.1038/bjc.2014.108>
- [41] Battiston, P., Kashyap, R. and Rotondi, V. (2021) Reliance on Scientists and Experts during an Epidemic: Evidence from the COVID-19 Outbreak in Italy. *SSM-Population Health*, **13**, Article 100721. <https://doi.org/10.1016/j.ssmph.2020.100721>
- [42] Bordandini, P., Santana, A. and Lobera, J. (2020) La fiducia nelle istituzioni ai tempi del COVID-19. *Polis*, **35**, 203-213.
- [43] Gualano, M.R., Lo Moro, G., Voglino, G., Bert, F. and Siliquini, R. (2022) Is the Pandemic Leading to a Crisis of Trust? Insights from an Italian Nationwide Study. *Public Health*, **202**, 32-34. <https://doi.org/10.1016/j.puhe.2021.10.015>
- [44] Lovari, A. (2020) Spreading (Dis)trust: Covid-19 Misinformation and Government Intervention in Italy. *Media and Communication*, **8**, 458-461.

- <https://doi.org/10.17645/mac.v8i2.3219>
- [45] Putnam, R. D. (1993) Making Democracy Work: Civic Traditions in Modern Italy. Princeton University Press.
- [46] Assirelli, G. (2014) Studiare di più rende cittadini migliori? Analisi della relazione tra istruzione e civicsness in Italia. *Scuola Democratica*, **1**, 29-52.
- [47] Alessi, E.J. and Martin, J.I. (2010) Conducting an Internet-Based Survey: Benefits, Pitfalls, and Lessons Learned. *Social Work Research*, **34**, 122-128. <https://doi.org/10.1093/swr/34.2.122>
- [48] Dillman, D.A., Smyth, J.D. and Christian, L.M. (2014) Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method. Wiley. <https://doi.org/10.1002/9781394260645>
- [49] Seymer, A. and Weichbold, M. (2018) Social Inequalities and the Effects of Incentives on Survey Participation: A Recruitment Experiment. *Österreichische Zeitschrift für Politikwissenschaft*, **47**, 5-20. <https://doi.org/10.15203/ozp.2393.vol47iss2>
- [50] Toepoel, V. and Schonlau, M. (2017) Dealing with Nonresponse: Strategies to Increase Participation and Methods for Postsurvey Adjustments. *Mathematical Population Studies*, **24**, 79-83. <https://doi.org/10.1080/08898480.2017.1299988>
- [51] Groves, R.M., Cialdini, R.B. and Couper, M.P. (1992) Understanding the Decision to Participate in a Survey. *Public Opinion Quarterly*, **56**, 475-495. <https://doi.org/10.1086/269338>
- [52] Groves, R.M. (2006) Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, **70**, 646-675. <https://doi.org/10.1093/poq/nfl033>
- [53] Marpsat, M. and Razafindratsima, N. (2010) Survey Methods for Hard-to-Reach Populations: Introduction to the Special Issue. *Methodological Innovations Online*, **5**, 3-16. <https://doi.org/10.4256/mio.2010.0014>