



Article

Impact of Two Different Recruitment Procedures (Random versus Volunteer Selection) on the Results of Seroepidemiological Study (SARS-CoV-2)

Maksymilian Gajda *, Małgorzata Kowalska and Jan Eugeniusz Zejda

Department of Epidemiology, Faculty of Medical Sciences in Katowice, Medical University of Silesia in Katowice, 40-752 Katowice, Poland; mkowalska@sum.edu.pl (M.K.); jzejda@sum.edu.pl (J.E.Z.)

* Correspondence: mgajda@sum.edu.pl

Citation: Gajda, M.; Kowalska, M.; Zejda, J.E. Impact of Two Different Recruitment Procedures (Random versus Volunteer Selection) on the Results of Seroepidemiological Study (SARS-CoV-2). *Int. J. Environ. Res. Public Health* **2021**, *18*, 9928. <https://doi.org/10.3390/ijerph18189928>

Academic Editors: Stuart Gilmour and Paul B. Tchounwou

Received: 10 August 2021

Accepted: 16 September 2021

Published: 21 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstract: The proper recruitment of subjects for population-based epidemiological studies is critical to the external validity of the studies and, above all, to the sound and correct interpretation of the findings. Since 2020, the novel coronavirus SARS-CoV-2 pandemic has been a new factor that has been, additionally, hindering studies. Therefore, the aim of our study is to compare demographic, socio-economic, health-related characteristics and the frequency of SARS-CoV-2 infection occurrence among the randomly selected group and the group composed of volunteers. We compare two groups of participants from the cross-sectional study assessing the seroprevalence of SARS-CoV-2 coronavirus, which was conducted in autumn 2020, in three cities of the Silesian Voivodeship in Poland. The first group consisted of a randomly selected, nationally representative, age-stratified sample of subjects (1167 participants, “RG” group) and was recruited using personal invitation letters and postal addresses obtained from a national registry. The second group (4321 volunteers, “VG” group) included those who expressed their willingness to participate in response to an advertisement published in the media. Compared with RG subjects, volunteers were more often females, younger and professionally active, more often had a history of contact with a COVID-19 patient, post-contact nasopharyngeal swab, fewer comorbidities, as well as declared the occurrence of symptoms that might suggest infection with SARS-CoV-2. Additionally, in the VG group the percentage of positive IgG results and tuberculosis vaccination were higher. The findings of the study confirm that surveys limited to volunteers are biased. The presence of the bias may seriously affect and distort inference and make the generalizability of the results more than questionable. Although effective control over selection bias in surveys, including volunteers, is virtually impossible, its impact on the survey results is impossible to predict. However, whenever possible, such surveys could include a small component of a random sample to assess the presence and potential effects of selection bias.

Keywords: SARS-CoV-2; seroepidemiological study; recruitment; randomization; self-selection

1. Introduction

The proper recruitment of subjects for population-based epidemiological studies is critical to the external validity of the studies and, above all, to the sound and correct interpretation of the findings. The problem poses a challenge in any selection procedure that aims at the representativeness of the study group. However, it is particularly important in large surveys that rely on convenient face-to-face or telephone interviews. The latter methods differ in terms of application difficulties and cost-effectiveness [1,2]. The common dilemma of this form of research stems from the fact that such surveys are usually limited to volunteers and, thus, are affected by selection bias. Another concern is related to a usually large number of refusals, making it difficult to obtain an appropriate

sample, also in terms of its size [3]. Thus, novel recruitment strategies are sought, also with the use of new forms of communication (e.g., the internet). Voluntary recruitment is known to be an important source of non-response bias and volunteer bias [1,4]. However, the exact dimension of the bias and its consequences are seldom reported.

There are many examples of different seroprevalence studies involving volunteers. The study conducted in Italy among adults over 65 years of age showed an overall seroprevalence of anti-SARS-CoV-2 antibodies of 4.7% [5]. In the study performed in Massachusetts at the end of the first wave of COVID-19 pandemic, the incidence of infection was lower in the representative sample than in volunteers: 1.85% vs. 3.29%, respectively [6]. The occurrence of SARS-CoV-2 infection in employees of a large teaching hospital in England examined between May and July 2020 was estimated to be 17.4% [7]. The Italian cohort showed a seroprevalence of 14.4% in a period from March to June 2020 [8]. During the first months of the COVID-19 pandemic, Poland reported a lower incidence of confirmed cases compared to other European countries [9]. The seroprevalence in the population of Poznań metropolitan area in Poland was 1.67% (sample collection between the end of July and the end of September 2020), finally dropping to 0.93% after immunoblotting verification [9]. As in other countries, Polish health care workers were also tested. Samples collected from the staff working at the Children's Memorial Health Institute in Warsaw (Poland) between July and August 2020 resulted in a seroprevalence of 0.85% [10]. The IgG seropositivity of asymptomatic healthcare workers from southern Poland varied between 1.2% and 10% (July/August 2020) [11]. In our recent seroepidemiological study on SARS-CoV-2 infection, we examined randomly selected subjects using questionnaires and immunological tests [12]. Our project also made possible applying the same research tools in a large group of volunteers, recruited from the same study area and examined in the same study period. It allowed us to explore, in a real-life setting, potential differences between both groups, in terms of information provided by questionnaires and immunological tests. The objective of our study was to compare demographic, socio-economic, health-related characteristics, and the frequency of SARS-CoV-2 infection occurrence between the randomly selected group and the group composed of volunteers.

2. Materials and Methods

In autumn 2020, the cross-sectional study assessing the seroprevalence of the SARS-CoV-2 virus was conducted in three Polish cities located in the Silesian Voivodeship (Gliwice, Katowice, and Sosnowiec). The methodology of this study and some of the results have already been presented and discussed in our previous articles [12,13]. The main research tools were questionnaires and measurement of anti-SARS-CoV-2 immunoglobulins (IgG and IgM). Antibodies were measured against S1 proteins (IgG) and modified nucleocapsid protein (IgM) of SARS-CoV-2 in serum and the results were expressed as ratios (test/control extinction), according to the following scale: ratio < 0.8 = negative result, ratio 0.8–1.09 = questionable result, ratio > 1.09 = positive result. The manufacturer's log files (EuroImmun Polska Sp. z o.o., Wrocław, Poland) reported a specificity of 99% (IgG) and a maximum sensitivity of 88% (IgG). Independent values included demographic, socioeconomic, health-related characteristics, and type of sampling. However, some important points that must be mentioned to introduce to the current analysis are discussed and repeated below. Initially, the study designed assumed the acquisition of subjects using random sampling method stratified by age and gender. Limited participation was expected at the planning stage of the study and was taken into account when calculating the minimum sample size. As only 1167 people (19.5% initially invited) agreed to participate in the study, supplementary recruitment was introduced resulting in additional 4321 volunteers and a total of 5488 participants. The final sample size met the estimated minimum sample size. Therefore, we decided to compare these two groups of participants to assess the impact of the recruitment method on the results. The first group (1167 participants) included a randomly selected, sex- and age-stratified sample of subjects (which is hereinafter referred to as the random group, RG), recruited using personal

invitation letters and postal addresses obtained from a national registry. The second group (4321 volunteers, “VG” group) included those who expressed their willingness to participate in response to an advertisement published in media (including regional newspapers, TV, radio, and Facebook). Moreover, we also provided each primary care physician active in the study area with a complete kit of information on the project. Distributions of age and sex in RG did not differ statistically significantly from analogous distributions in the general population of the Silesian Voivodeship [12].

2.1. Statistical Analysis

Both groups (RG and VG) were compared in terms of the results of the questionnaire and serological examinations. The distribution of quantitative variables was initially assessed, with deviations from the normal distribution found (Shapiro–Wilk test). Therefore, between-group differences in the distribution of quantitative variables were analyzed using non-parametric tests (Mann–Whitney test). Analysis involving qualitative variables was performed using chi-square test (or Fisher’s exact test for cells with expected counts less than 5). The results of descriptive analysis included the relative frequencies (percentages) for categorical variables and the medians with interquartile ranges (IQR) for quantitative variables. Additionally, we deepened the analysis of the determinants of recruitment (binary dependent variable: random recruitment vs. self-selection) applying the multivariate logistic regression model. Only the variables considered statistically significant at the stage of univariate analysis were included in the regression model and, finally, odds ratios (OR) with 95% confidence intervals (CI) were counted. Interpretation of statistical tests was conducted according to the criterion $p < 0.05$. All analyses were performed using the R 4.1.0 statistical environment (2021, R Core Team, GNU General Public License; R Foundation for Statistical Computing, Vienna, Austria).

2.2. Ethical Approval

The study was approved by the Ethics Committee of the Medical University of Silesia in Katowice (14 November 2020; the number of approval PCN/0022/KB1/61/20) and was registered at the ClinicalTrials.gov (accessed on 26 March 2021) PRS system with NCT04627623 identifier. The protocol and the course of the study were in line with Helsinki’s declaration. All participants signed informed written consent to participate in the study.

3. Results

The RG group consisted of 1167, while the VG group included 4321 subjects. In the first stage, several factors were identified using simple difference tests, including socio-demographic features, differentiating participants depending on the recruitment model. Compared with randomly selected subjects, the volunteers were more often females, they were younger and more often professionally active (Table 1). They more often had a history of contact with a COVID-19 patient (33% vs. 12.8% in RG, $p < 0.001$), post-contact nasopharyngeal swab (17.5% vs. 13%, $p = 0.001$), fewer comorbidities (12.9% vs. 18%, $p < 0.001$; more details in Table 2), as well as declared the occurrence of symptoms that might suggest infection with SARS-CoV-2 (detailed results presented in Table 3).

Additionally, a significantly higher percentage of positive IgG results was found in this group (23.5% compared to 11.4% in RG, $p < 0.001$), denoting their contact with the SARS-CoV-2 virus (Table 4). We did not find any statistically significant differences neither in the frequency of positive IgM tests between the compared groups nor in associations between IgM results and other variables, including symptoms.

Table 1. Basic sociodemographic characteristics (relative frequencies presented as percentages or median with interquartile ranges).

| | | All Subjects n = 5488 | Randomized n = 1167 | Self-Selected n = 4321 | p |
|---|-------------------|--------------------------|------------------------|---------------------------|--------|
| Quantitative variables (median and IQR) | | | | | |
| Age | Median (IQR) | 44 (32–57) | 48 (34–63) | 43 (31–55) | <0.001 |
| Body mass index | Median (IQR) | 25.2 (22.3–28.7) | 25.6 (22.5–29.0) | 25.2 (22.2–28.5) | 0.08 |
| | Missing | 6.7% | 9.8% | 5.9% | |
| Categorical variables (percentages) | | | | | |
| Town | Gliwice | 32.5 | 33.6 | 32.2 | 0.7 |
| | Katowice | 34.2 | 33.8 | 34.3 | |
| | Sosnowiec | 33.3 | 32.6 | 33.5 | |
| Gender | Female | 58.3 | 51.3 | 60.1 | <0.001 |
| | Male | 41.7 | 48.7 | 39.9 | |
| Professional activity | Has no job | 31.2 | 43.9 | 27.7 | <0.001 |
| | Has a job | 67.6 | 55.5 | 70.9 | |
| Month of serological examination | Works and studies | 1.2 | 0.6 | 1.4 | <0.001 |
| | 10/2020 | 9.1 | 38.2 | 1.3 | |
| | 11/2020 | 74.0 | 60.1 | 77.7 | |
| | 12/2020 | 16.9 | 1.7 | 21.0 | |
| Vaccinated against tuberculosis | Yes | 76.2 | 68.7 | 78.2 | <0.001 |
| Vaccinated against seasonal influenza last year | Yes | 16.5 | 12.2 | 17.6 | <0.001 |
| Medical help following symptoms | Yes | 21.6 | 19.5 | 22.2 | 0.2 |
| | Missing | 2.5 | 3.5 | 2.2 | |
| Contact with a confirmed case of COVID-19 | Yes | 28.7 | 12.8 | 33.0 | <0.001 |
| Had a post contact test | Yes | 16.6 | 13.0 | 17.5 | 0.001 |

Legend: n—number of participants; p—p-value; IQR—interquartile range; IgG—immunoglobulin G; IgM—immunoglobulin M.

Table 2. The burden of comorbidities in the study group in the entire study group and depending on the recruitment mode (percentages).

| Comorbidity | All Subjects n = 5479 | Randomized n = 1164 | Self-Selected n = 4315 | p |
|---------------------------------------|--------------------------|------------------------|---------------------------|--------|
| Any comorbidities | 14.0 | 18.0 | 12.9 | <0.001 |
| Coronary artery disease | 0.6 | 1.1 | 0.4 | 0.01 |
| Myocardial infarction | 0.5 | 0.9 | 0.3 | 0.05 |
| Heart failure | 0.3 | 0.5 | 0.3 | 0.2 |
| Valvular heart disease | 0.4 | 0.8 | 0.3 | 0.07 |
| Stroke | 0.3 | 0.6 | 0.2 | 0.07 |
| Chronic obstructive pulmonary disease | 2.0 | 3.2 | 1.7 | 0.003 |
| Asthma | 6.7 | 7.1 | 6.6 | 0.6 |
| Chronic Allergic Disease | 11.5 | 9.9 | 11.9 | 0.06 |
| Diabetes | 6.5 | 9.7 | 5.6 | <0.001 |
| Cancer | 4.2 | 6.4 | 3.6 | <0.001 |
| Chronic Rheumatological Disease | 3.5 | 5.3 | 3.0 | <0.001 |
| Autoimmune disease | 6.7 | 6.2 | 6.9 | 0.4 |

Legend: n—number of participants; p—p-value.

Table 3. Results of epidemiological interview, including prevalence of declared symptoms (percentages) with *p*-values of difference tests.

| Symptoms | All Subjects n = 5479 | Randomized n = 1164 | Self-Selected n = 4315 | <i>p</i> |
|--------------------------|--------------------------|------------------------|---------------------------|----------|
| Fever 38 °C | 18.7 | 15.1 | 19.7 | <0.001 |
| Chills | 20.7 | 14.7 | 22.3 | <0.001 |
| Fatigue | 47.9 | 36.6 | 51.0 | <0.001 |
| Muscle ache (myalgia) | 31.5 | 23.7 | 33.7 | <0.001 |
| Sore throat | 35.7 | 29.3 | 37.4 | <0.001 |
| Cough | 36.5 | 33.0 | 37.4 | 0.005 |
| Runny nose (rhinorrhoea) | 46.6 | 43.4 | 47.5 | 0.01 |
| Shortness of breath | 16.1 | 13.5 | 16.8 | 0.006 |
| Wheezing | 7.7 | 7.0 | 7.9 | 0.4 |
| Chest pain | 14.3 | 13.0 | 14.6 | 0.2 |
| Headache | 44.3 | 33.6 | 47.2 | <0.001 |
| Conjunctivitis | 5.1 | 4.2 | 5.4 | 0.1 |
| Nausea/vomiting | 7.9 | 6.7 | 8.2 | 0.09 |
| Abdominal pain | 13.7 | 11.5 | 14.3 | 0.01 |
| Diarrhea | 15.6 | 11.9 | 16.6 | <0.001 |
| Loss of smell/taste | 13.5 | 9.2 | 14.6 | <0.001 |
| Other symptoms | 6.0 | 5.0 | 6.3 | 0.08 |

Legend: n—number of participants; *p*—*p*-value.

Table 4. The prevalence of positive serological results and ratios of IgG and IgM: percentages or median values with interquartile ranges (IQR).

| Variables | All Subjects n = 5479 | Randomized n = 1164 | Self-Selected n = 4315 | <i>p</i> |
|--------------------|--------------------------|------------------------|---------------------------|----------|
| <i>IgG</i> | | | | |
| % positive | 21.0 | 11.4 | 23.5 | <0.001 |
| ratio-median (IQR) | 0.2 (0.1–0.5) | 0.2 (0.1–0.3) | 0.1 (0.1–0.9) | <0.001 |
| <i>IgM</i> | | | | |
| % positive | 5.0 | 4.8 | 5.1 | 0.8 |
| ratio-median (IQR) | 0.1 (0.1–0.2) | 0.1 (0.1–0.3) | 0.1 (0.1–0.2) | 0.004 |

Legend: n—number of participants; *p*—*p*-value; IQR—interquartile range; IgG—immunoglobulin G; IgM—immunoglobulin M.

Furthermore, most of them were vaccinated against tuberculosis (78.2% vs. 68.7%, *p* < 0.001), but with no differences for the influenza seasonal vaccination. There were no differences in the recruitment structure to the inhabited city, seek for medical help to symptoms, and subject's body mass index (BMI). Detailed characteristics are presented in Tables 1–4.

Most of these relationships were confirmed in the multivariable regression model (Table 5), which had a McFadden pseudo R² index of 0.3 and Cragg and Uhler's pseudo R² of 0.4. Based on these results, subjects who were self-selected to the study were less often male (OR = 0.76, 95% CI 0.64–0.89), more often younger (OR = 0.98; 95% CI 0.98–0.99), recruited in December (OR = 411.58, 95% CI 246.24–727.09), professionally active (OR = 1.92; 95% CI 1.62–2.28), had contact with a COVID-19 case (OR = 1.49; 95% CI 1.15–1.95), and had a higher IgG ratio (OR = 1.08; 95% CI 1.03–1.15). According to the multivariable model, the presence of comorbidities did not significantly affect study participation.

Table 5. Characteristics of the respondents in univariate analyses and multivariate binary logistics regression model in relation to the recruitment mode as dependent variable.

| Binary Logistic Regression-Dependent Variable: Mode of Recruitment (Random = 0 vs. Self-Selected = 1) | | | | | | |
|--|-------------|------------------|----------|---------------|------------------|----------|
| Explanatory variables | Univariable | | | Multivariable | | |
| | OR | 95% CI | <i>p</i> | OR | 95% CI | <i>p</i> |
| Age | 0.99 | 0.98 to 0.99 | * | 0.98 | 0.98 to 0.99 | * |
| Body mass index | 0.99 | 0.98 to 1.00 | NS | X | X | X |
| IgM ratio | 0.99 | 0.90 to 1.09 | NS | X | X | X |
| IgG ratio | 1.20 | 1.14 to 1.26 | * | 1.08 | 1.03 to 1.15 | * |
| Gender male vs. female | 0.70 | 0.61 to 0.80 | * | 0.76 | 0.64 to 0.89 | * |
| Comorbidities Yes vs. No | 0.68 | 0.57 to 0.81 | * | 0.90 | 0.71 to 1.12 | NS |
| Month of serological examination | | | | | | |
| <i>November vs. October</i> | 38.09 | 28.75 to 51.35 | * | 35.52 | 26.48 to 48.78 | * |
| <i>December vs. October</i> | 361.41 | 218.45 to 624.13 | * | 411.58 | 246.24 to 727.09 | * |
| Professional activity | | | | | | |
| <i>Has a job vs. has no job</i> | 2.01 | 1.77 to 2.31 | * | 1.92 | 1.62 to 2.28 | * |
| <i>Works and studies vs. has no job</i> | 3.60 | 1.75 to 8.70 | * | 2.53 | 1.13 to 6.80 | 0.039 |
| Contact with a confirmed COVID-19 case | | | | | | |
| <i>No vs. do not know</i> | 0.61 | 0.52 to 0.73 | * | 0.67 | 0.55 to 0.83 | * |
| <i>Yes vs. do not know</i> | 2.36 | 1.89 to 2.95 | * | 1.49 | 1.15 to 1.95 | 0.002 |
| Post-contact test | | | | | | |
| <i>No vs. do not know</i> | 0.98 | 0.60 to 1.54 | NS | X | X | X |
| <i>Yes vs. do not know</i> | 1.39 | 0.84 to 2.26 | NS | X | X | X |
| Vaccinated against tuberculosis | | | | | | |
| <i>No vs. do not know</i> | 0.79 | 0.60 to 1.03 | 0.08 | 0.98 | 0.69 to 1.40 | NS |
| <i>Yes vs. do not know</i> | 1.54 | 1.31 to 1.80 | * | 1.43 | 1.18 to 1.75 | * |
| Vaccinated against seasonal influenza last year | | | | | | |
| <i>No vs. do not know</i> | 1.16 | 0.38 to 3.00 | NS | X | X | X |
| <i>Yes vs. do not know</i> | 1.79 | 0.57 to 4.69 | NS | X | X | X |

Legend: OR—odds ratio; CI—confidence interval; *p*—level of significance; *—*p*-value ≤ 0.001; NS—*p*-value > 0.1; X—not included in the multivariable model.

4. Discussion

The objective of our study was to compare demographic, socio-economic, health-related characteristics, and the anti-SARS-CoV-2 immunoglobulin IgG occurrence among the randomly selected group and the group composed of volunteers. The major finding of our study was the identification of a selection bias, defined as the difference in many characteristics between randomly selected subjects and volunteers. Our findings not only showed between-group statistically significant differences in the distribution and associations of many important variables, but also allowed us to measure the size of the bias.

As expected, our study confirmed the between-group differences in the distribution of pertinent variables and perhaps the most important difference concerned the frequency of seropositivity with an apparent overestimation of the frequency of SARS-CoV-2 infection in the general population. However, the bias was seen in relation to all relevant aspects of the study: description, analysis differences, and associations.

With regard to the structure of examined groups, our results confirmed a larger participation of women and younger people among volunteers and this finding was consistent with previous observations in this regard [1,14]. This finding was also in line with the results of other cross-sectional studies concerned with different goals of public health,

in which females engage more frequently [14–17]. Importantly, within our study and compared with randomly selected subjects, the volunteers had a significantly higher percentage of positive IgG test results, probably due to their more frequent contact with the SARS-CoV-2 coronavirus. Such an association was suggested by the information provided by the questionnaire. It could not be excluded that the people who had had contact with individuals positively diagnosed with novel coronavirus infections were more likely to use the offered opportunity to check their serological status. Another self-selection mechanism might have resulted from a greater personal interest in the subject of the study, and a higher awareness and perception of the impact of the COVID-19 pandemic. As mentioned above, the volunteer bias shown in our study affected all pertinent aspects of research, starting with the baseline description of the subjects. Such an observation was similar to the results of the New Zealand study in which volunteers and sampled subjects differed significantly, mainly in socio-behavioral respects [15]. In another current publication, the authors concluded that depending on the sampling location and time, people who are present to be sampled may be at a higher or lower risk of COVID-19 than the average risk in the source population [18]. Therefore, our results provided by simple analyses were verified with the use of a multivariable logistic regression model which allowed to identify the factors characterizing the group of volunteers vis-à-vis the randomly selected group. It was established that they included: gender (female), age (younger), employment status (active), history of contact with COVID-19 case (positive), and IgG ratio (higher). These findings were broadly consistent with what has been reported so far, especially with regard to the dominance of women among volunteers [1,2,4].

The results of another recently published study showed that the two different sampling methods had a significant impact on the reporting of COVID-19 symptoms leading to different frequencies (symptomatic subjects: 28.2% in open invitation group and 16.2% in random sample), which was in line with our results. Moreover, the overall prevalence rate of active COVID-19 cases within the open invitation sample (13.3%) was almost twice as big as that found in the random sample (6.9%) [19].

The method of recruiting volunteer participants in our study was similar to the methods used in other population-based studies [9]. In our study, we employed a variety of means in order to reach the population, ranging from traditional press, through television, and ending with the internet. It could be expected that such a “diversified” approach increases the probability of reaching different groups of potential participants in comparison with only one type of information channel.

There is ample evidence that voluntary recruitment may lead to a distorted assessment of the problem under investigation [4,20–22]. The issue is far from being resolved particularly in the field of population-based surveys targeting important public health topics. Moreover, survey response rates in cross-sectional studies have been declining for decades [23] and there is a need to develop better methods and practices of surveying in epidemiological studies. Some possibilities arise from the advent of new technologies and the introduction of effective social network support. Such methods should be reviewed and verified in real-life studies with a view of “good epidemiological practice”. The lack of effective population-based recruitment strategies in large surveys affects the reliability of the results and significantly hampers the proper interpretation of the research findings [24,25]. Our study showed that the practical method of such an evaluation may involve a direct comparison of the results obtained using novel recruitment techniques and random sampling, which remain essential components of standard epidemiological studies designs [18,26].

Our study had some limitations. First of all, it was unjustified to claim that our selection procedure of the random sample had resulted in a fully representative sample of the source population in all aspects pertinent to the study objectives. Even if the sample had met the requirements of sex and age distribution and its size had been satisfactory in terms of the desired study power, the participation rate was rather low (19%). However, the problem was general and with several assumptions, it was a random sampling that allowed

inferences to a source population. Another issue that potentially hampered the conclusion regarding the exact impact of volunteer bias as analyzed in our study stemmed from a low participation rate in the random-selection phase of the project. The problem was universal and it could not be excluded that a better participation would have resulted in more reliable estimates. However, the findings of our study reflected a real life scenario and, with all potential pitfalls, the results of our comparison between representative and volunteers groups described the presence of a volunteer bias. Moreover, the random selection might have missed the individuals who had been treated for COVID-19 or on quarantine and not responded to the invitation. The methods used to address the objective of our study allowed conclusive real-life comparisons which was the strength of our investigation. Both groups of subjects were inhabitants of the same precisely defined area (three towns; population 600,000) and all subjects were examined by one team using the same methods (questionnaire, IgG test), including one diagnostic laboratory. Moreover, both groups were examined in the same study period (October–November 2020). Additionally, the fear of being infected during the procedures carried out in the study could lead to the exclusion of some potential subjects from the participation (non-responder bias), which may be one of the limitations of the study.

5. Conclusions

The findings of the study confirmed that surveys limited to volunteers are biased. The presence of the bias may seriously affect and distort inference and make the generalizability of the results more than questionable. The impact of bias on the external validity of the study depends on its size. Specific outcomes of comparisons performed within our study showed that, with regard to such an important issue as SARS-CoV-2 infection, the much-needed evidence on the description and cause–effect associations unequivocally depend on the recruitment procedure. Effective control over selection bias in surveys, including volunteers, is virtually impossible and its impact on the survey results is impossible to predict. However, whenever possible, such surveys could include a small component of a random sample to assess the presence and potential effects of selection bias.

Author Contributions: Conceptualization, M.K. and J.E.Z.; methodology, M.K. and M.G.; formal analysis, M.G.; investigation, M.K., J.E.Z., and M.G.; resources, J.E.Z.; data curation, M.G.; writing—original draft preparation, M.G.; writing—review and editing, M.K., J.E.Z., and M.G.; supervision, M.K. and J.E.Z.; project administration, M.K. and J.E.Z.; funding acquisition, J.E.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The presented results were derived from data obtained during the project titled: “Prevalence and Risk Factors of COVID-19 in the Upper Silesian Agglomeration (EpiSARS2)”, which was supported by the Medical Research Agency in Poland (2020/ABM/COVID19/0044).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of the Medical University of Silesia in Katowice, Poland (approval number: PCN/0022/KB1/61/20; 14 July 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are available for research upon approval from the Medical Research Agency, Poland. To obtain data, researchers need to submit an analysis proposal to the corresponding author for evaluation and processing the request to the Medical Research Agency, Poland.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ganguli, M.; Lytle, M.E.; Reynolds, M.D.; Dodge, H.H. Random versus volunteer selection for a community-based study. *J. Gerontol. A Biol. Sci. Med. Sci.* **1998**, *53*, M39–M46.
2. Rosnow, R.L.; Rosenthal, R. The volunteer subject revisited. *Aust. J. Psychol.* **1976**, *28*, 97–108.
3. Gajda, M.; Kowalska, M. The role of an electronic questionnaires in epidemiologic studies aimed at cancer prevention. *Przegl. Epidemiol.* **2016**, *70*, 479–489.

4. Cheung, K.L.; Ten Klooster, P.M.; Smit, C.; de Vries, H.; Pieterse, M.E. The impact of non-response bias due to sampling in public health studies: A comparison of voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. *BMC Public Health* **2017**, *17*, 276.
5. Polvere, I.; Parrella, A.; Casamassa, G.; D'Andrea, S.; Tizzano, A.; Cardinale, G.; Voccola, S.; Porcaro, P.; Stilo, R.; Vito, P.; et al. Seroprevalence of Anti-SARS-CoV-2 IgG and IgM among Adults over 65 Years Old in the South of Italy. *Diagnostics* **2021**, *11*, 483.
6. Bruno-Murtha, L.A.; Osgood, R.; Lan, F.-Y.; Buley, J.; Nathan, N.; Weiss, M.; MacDonald, M.; Kales, S.N.; Sayah, A.J. SARS-CoV-2 antibody seroprevalence after the first wave among workers at a community healthcare system in the Greater Boston area. *Pathog. Glob. Health* **2021**, 1–4, <https://doi.org/10.1080/20477724.2021.1901041>.
7. Shorten, R.J.; Haslam, S.; Hurley, M.A.; Rowbottom, A.; Myers, M.; Wilkinson, P.; Orr, D. Seroprevalence of SARS-CoV-2 infection in healthcare workers in a large teaching hospital in the North West of England: A period prevalence survey. *BMJ Open* **2021**, *11*, e045384.
8. De Santi, M.; Diotallevi, A.; Brandi, G. Seroprevalence of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) infection in an Italian cohort in Marche Region, Italy. *Acta Biomed.* **2021**, *92*, e2021070.
9. Lorent, D.; Nowak, R.; Roxo, C.; Lenartowicz, E.; Makarewicz, A.; Zaremba, B.; Nowak, S.; Kuszel, L.; Stefaniak, J.; Kierzek, R.; et al. Prevalence of Anti-SARS-CoV-2 Antibodies in Poznań, Poland, after the First Wave of the COVID-19 Pandemic. *Vaccines* **2021**, *9*, 541.
10. Kasztelewicz, B.; Janiszewska, K.; Burzyńska, J.; Szydłowska, E.; Migdał, M.; Dzierżanowska-Fangrat, K. Prevalence of IgG antibodies against SARS-CoV-2 among healthcare workers in a tertiary pediatric hospital in Poland. *PLoS ONE* **2021**, *16*, e0249550.
11. Bułdak, R.J.; Woźniak-Grygiel, E.; Wąsik, M.; Kasperczyk, J.; Gawrylak-Dryja, E.; Mond-Paszek, R.; Konka, A.; Badura-Brzoza, K.; Fronczek, M.; Golec, M.; et al. SARS-CoV-2 Antibody Screening in Healthcare Workers in Non-Infectious Hospitals in Two Different Regions of Southern Poland (Upper Silesia and Opole Voivodeships): A Prospective Cohort Study. *Int. J. Environ. Res. Public Health* **2021**, *18*, 4376.
12. Zejda, J.E.; Brożek, G.M.; Kowalska, M.; Barański, K.; Kaleta-Pilarska, A.; Nowakowski, A.; Xia, Y.; Buszman, P. Seroprevalence of Anti-SARS-CoV-2 Antibodies in a Random Sample of Inhabitants of the Katowice Region, Poland. *Int. J. Environ. Res. Public Health* **2021**, *18*, 3188.
13. Kowalska, M.; Niewiadomska, E.; Barański, K.; Kaleta-Pilarska, A.; Brożek, G.; Zejda, J.E. Association between Influenza Vaccination and Positive SARS-CoV-2 IgG and IgM Tests in the General Population of Katowice Region, Poland. *Vaccines* **2021**, *9*, 415.
14. Stopponi, M.A.; Alexander, G.L.; McClure, J.B.; Carroll, N.M.; Divine, G.W.; Calvi, J.H.; Rolnick, S.J.; Strecher, V.J.; Johnson, C.C.; Ritzwoller, D.P. Recruitment to a randomized web-based nutritional intervention trial: Characteristics of participants compared to non-participants. *J. Med. Internet Res.* **2009**, *11*, e38.
15. Broad, J.B.; Wu, Z.; Bloomfield, K.; Hikaka, J.; Bramley, D.; Boyd, M.; Tatton, A.; Calvert, C.; Peri, K.; Higgins, A.-M.; et al. Health profile of residents of retirement villages in Auckland, New Zealand: Findings from a cross-sectional survey with health assessment. *BMJ Open* **2020**, *10*, e035876.
16. Bjertness, E.; Sagatun, A.; Green, K.; Lien, L.; Sjøgaard, A.J.; Selmer, R. Response rates and selection problems, with emphasis on mental health variables and DNA sampling, in large population-based, cross-sectional and longitudinal studies of adolescents in Norway. *BMC Public Health* **2010**, *10*, 602.
17. Xiaoming, X.; Ming, A.; Su, H.; Wo, W.; Jianmei, C.; Qi, Z.; Hua, H.; Xuemei, L.; Lixia, W.; Jun, C.; et al. The psychological status of 8817 hospital workers during COVID-19 Epidemic: A cross-sectional study in Chongqing. *J. Affect. Disord.* **2020**, *276*, 555–561.
18. Accorsi, E.K.; Qiu, X.; Rumpler, E.; Kennedy-Shaffer, L.; Kahn, R.; Joshi, K.; Goldstein, E.; Stensrud, M.J.; Niehus, R.; Cevik, M.; et al. How to detect and reduce potential sources of biases in studies of SARS-CoV-2 and COVID-19. *Eur. J. Epidemiol.* **2021**, *36*, 179–196.
19. A/Qotba, H.A.; Al Nuaimi, A.S.; Al Mujalli, H.; Zainel, A.A.; Khudadad, H.; Marji, T.; Veettil, S.T.; Syed, M.A. COVID-19 Surveillance in the Primary Health Care Population of Qatar: Experience of Prioritizing Timeliness Over Representativeness When Sampling the Population. *Front. Public Health* **2021**, *9*, 654734.
20. Jordan, S.; Watkins, A.; Storey, M.; Allen, S.J.; Brooks, C.J.; Garaiova, I.; Heaven, M.L.; Jones, R.; Plummer, S.F.; Russell, I.T.; et al. Volunteer bias in recruitment, retention, and blood sample donation in a randomised controlled trial involving mothers and their children at six months and two years: A longitudinal analysis. *PLoS ONE* **2013**, *8*, e67912.
21. Higgins, J.P.; Altman, D.G.; Sterne, J.A. Chapter 8: Assessing Risk of Bias in Included Studies. In *Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series Version 5.2.0 (Updated June 2017)*; Higgins, J., Churchill, R., Chandler, J., Cumpston, M., Eds; Cochrane: 2017; pp. 187–241. Available online: https://training.cochrane.org/sites/training.cochrane.org/files/public/uploads/resources/Handbook5_1/Chapter_8_Handbook_5_2_8.pdf (accessed on 10 September 2021).
22. Kho, M.E.; Duffett, M.; Willison, D.J.; Cook, D.J.; Brouwers, M.C. Written informed consent and selection bias in observational studies using medical records: Systematic review. *BMJ* **2009**, *338*, b866.
23. Hutchinson, M.K.; Sutherland, M.A. Conducting surveys with multidisciplinary health care providers: Current challenges and creative approaches to sampling, recruitment, and data collection. *Res. Nurs. Health* **2019**, *42*, 458–466.
24. Travlos, V.; Patman, S.; Wilson, A.; Simcock, G.; Downs, J. Quality of Life and Psychosocial Well-Being in Youth With Neuromuscular Disorders Who Are Wheelchair Users: A Systematic Review. *Arch. Phys. Med. Rehabil.* **2017**, *98*, 1004–1017.e1.
25. Guo, X.; Vittinghoff, E.; Olgin, J.E.; Marcus, G.M.; Pletcher, M.J. Volunteer Participation in the Health eHeart Study: A Comparison with the US Population. *Sci. Rep.* **2017**, *7*, 1956.
26. Wang, X.; Cheng, Z. Cross-Sectional Studies: Strengths, Weaknesses, and Recommendations. *Chest* **2020**, *158*, S65–S71.