PDF version

Ali Davoudi Kiakalayeh,[1] Morteza Rahbar Taramsari,[2] Reza Mohammadi,[3] Sajad Davoudi Kiakalayeh [4] and Hassan Kavakpour [4]

[1]Department of Preventive and Social Medicine, School of Medicine, Guilan University of Medical Sciences. Rasht, Islamic Republic of Iran (Correspondence to: A. Davoudi Kiakalayeh: davoudikiakalayeh@gmail.com). [2]Department of Forensic Medicine, School of Medicine, Guilan University of Medical Sciences, Rasht, Islamic Republic of Iran.

[3]Division of Family Medicine and Primary Care, Division of Social Medicine, Department of NVS, Karolinska Institutet, Stockholm, Sweden.

[4]Guilan Road Trauma Research Center, Guilan University of Medical Sciences, Rasht, Islamic Republic of Iran.

Abstract

Background: Reliable estimation of prevalence is important for monitoring and evaluation of COVID-19 prevention programmes among at-risk populations.

Aims: We compared the capture–recapture method with a seroprevalence survey for accurate estimation of the prevalence of COVID-19 during a 1-year period in Guilan Province, northern Islamic Republic of Iran.

Methods: We used the capture–recapture method to estimate the prevalence of COVID-19. Records from the primary care registry system and the Medical Care Monitoring Center were compared, using 4 matching approaches based on combinations of the following variables: name, age, gender, date of death, positive or negative cases, and alive or dead cases.

Results: The estimated prevalence of COVID-19 in the study population from the beginning of

the pandemic in February 2020 until the end of January 2021 was 16.2–19.8%, depending on the matching approach used, which was lower than in previous studies.

Conclusion: The capture–recapture method may provide better accuracy than seroprevalence surveys in measuring the prevalence of COVID-19. This method may also reduce the bias in the estimation of prevalence and correct the misconception of policymakers about seroprevalence survey results.

Keywords: COVID-19, capture–recapture, seroprevalence, prevalence, Islamic Republic of Iran

Citation: Davoudi Kiakalayeh A; Rahbar Taramsari M; Mohammadi R; Davoudi Kiakalayeh S; Kavakpour H. Comparison of the capture–recapture method and seroprevalence survey for estimation of COVID-19 prevalence in the Islamic Republic of Iran. East Mediterr Health J. 2023;29(2):126–131.https://doi.org/10.26719/emhj.23.010
Received: 06/12/20; accepted: 05/10/22

Copyright © Authors 2023; Licensee: World Health Organization. EMHJ is an open access journal. This paper is available under the Creative Commons Attribution Non-Commercial ShareAlike 3.0 IGO licence (CC BY-NC-SA 3.0 IGO; https://creativecommons.org/licenses/by-nc-sa/3.0/igo).

Introduction

A novel coronavirus pneumonia caused by SARS-CoV-2 was identified in Wuhan, China in December 2019 (1–3). On 3 February 2020, there were simultaneous outbreaks of viral pneumonia in Guilan Province in northern Islamic Republic of Iran and other provinces, and the infection spread rapidly throughout the country. On 10 February, the disease had spread into more than 212 countries around the world, with more than 103 million cases (4). On 11 February, the World Health Organization named the new disease COVID-19.

Estimations of COVID-19 prevalence are useful in predicting epidemic trends that can help policy-makers with informed decision-making. Since its inception, researchers have used models to estimate the size and trends of the COVID-19 pandemic. However, to date, the

seroprevalence of SARS-CoV-2 infection has not been fully assessed (5). The capture–recapture method has been applied in epidemiological and surveillance studies to correct under-ascertainment in epidemiological surveillance (6–8).

In this study, we used the capture–recapture method to analyse the data from overlapping lists of cases from 2 sources to generate estimates of missing cases in Guilan Province from February 2020 to February 2021.

Methods

Study area and population

The study was based on annual outpatient and inpatient data from Guilan Province reported in the Primary Care Registry and Medical Care Monitoring Center. Guilan is located in northern Islamic Republic of Iran and has about 2 530 657 inhabitants, and is one of the main destinations for tourists. Denominators were determined using population data obtained from the Statistical Center of Iran. The Iranian Ministry of Health and Medical Education established the Primary Care Registry and Medical Care Monitoring Center in early 2020 to register COVID-19 infection across the country. These 2 registries are comprehensive sources of data on COVID-19 that enhance the use of capture–recapture methods to provide a clear estimate of the prevalence of COVID-19 cases in northern Islamic Republic of Iran. Records included in the Medical Care Monitoring Center are based on information from inpatient and outpatient facilities, and emergency medical services. Positive or negative reverse transcription-polymerase chain reaction (RT-PCR) results for SARS-CoV-2 in COVID-19 cases were often included in the reports prepared by the emergency departments of hospitals and outpatient facilities. The Primary Care Registry is managed by the Ministry of Health and Medical Education, and is currently used across the country. Records for this surveillance system are generated using information from outpatients and inpatients, and for COVID-19 cases, a positive or negative result for SARS-COV-2 by RT-PCR. Quality control on COVID-19 data gathered by both registries is conducted at district level by frequent checking of the records by trained health officers in each district; at the provincial level by programme managers; and at the national level by employees in the Ministry of Health.

Capture–recapture method

COVID-19 cases were first identified or captured using Medical Care Monitoring Center records, and then recaptured using Primary Care Registry records. This method was based on matching 2 independent data sources to arrive at an overall estimation. Capture–recapture methods that use results from more than 1 surveillance source can provide more reliable estimates of communicable and noncommunicable diseases. Another potential use of this method is the refinement of prevalence estimates of population-based studies. For application of the capture–recapture method, 4 assumptions must be considered that underpin the 2 data sources (9). Firstly, 2 comprehensive datasets in a closed population (constant population during the

study period) must be used. The chance of being referred to either database must be equal, and datasets must be relatively independent and do not refer cases to each other. Cases must be matched confidently and accurately between sources (10). Ascertainment rates in active surveillance systems, such as population-based registries, are acknowledged to be better than those obtained with passive surveillance systems (11,12).

In the current study, records from both databases were merged into a single file using Excel 2019, and sorted for cases that were reported in both systems, using multiple screenings. For the first matching exercise, 6 variables were used for matching records: name, age, gender, date of death, positive or negative cases, and alive or dead cases. The strategy adopted for subsequent matching exercises used 4 separate matching approaches, A–D. Each of the approaches required matching of the 6 variables. Approach A required 6 matching variables in order to consider that the records from both systems identified the same cases. During subsequent matching with approaches B–D, the number of required matching variables was progressively reduced. In the final approach D, only 3 variables were required: gender, positive or negative cases, and dead or alive cases.

An estimation of the total number of cases including the missing cases was derived using the simplest form of a 2-sample capture–recapture model developed by the researchers (13,14). Thus, COVID-19 prevalence rates and 95% confidence interval (CI) were calculated by the estimated total number of COVID-19 cases (N) using the following formula: $N = [(S1+1) \times (S2+1)/ (C+1)]-1$

where S1 represented the number of records in the Primary Care Registry and S2 the number of records in the Medical Care Monitoring Center. The overlap between these samples was named C, which represented those common to both sources.

Variance $N = [(S1 + 1) (S2 + 1)(S1 - C) (S2 + C)/(C + 1)2 (C + 2)]$

95% CI = $n \pm 1.96$

The (prevalence) infection rate per 100 000 population was calculated by dividing the number of cases (N) by the population of Guilan Province (based on Statistical Center of Iran) (15).

Results

The capture–recapture method using approach D identified 110 477 COVID-19 cases in the Primary Care Registry data, resulting in an infection rate of 4315 per 100 000 population (Table 1). The Medical Care Monitoring Center database identified 34 833 COVID-19 cases, leading to an infection rate of 1360 per 100 000 population. There were no significant differences between the variables describing demographic characteristics of the cases recorded in both surveillance systems. There were 9475 cases common to both datasets. Application of the method described in this study indicated an estimated total of 406 162 COVID-19 cases, corresponding to an infection rate of 15 865 per 100 000 population (95% CI: 15 783–15 948). By combining these 2 datasets and ignoring the overlapping cases, we identified an aggregate of 135 835 cases, or 33.4% ascertainment-corrected rate for COVID-19 in the population. The different matching approaches and the more relaxed matching criteria of approach D (Table 1) yielded an estimated prevalence of 16.2%. The more restrictive approach A (Table 2) resulted in a larger estimated prevalence of 19.8%.

Discussion

COVID-19 is a global public health crisis, therefore, there is a need to monitor and estimate its prevalence for effective infection control and management. Different statistical models, such as the seroprevalence method, have been used for estimation of COVID-19 prevalence since February 2020 (15–17). To our knowledge, this is the first study based on an official dataset for prediction of COVID-19 prevalence using the capture–recapture method, over a 1-year period after the onset of the pandemic in low- and middle-income countries. Our analysis demonstrated that the estimated prevalence of COVID-19 in the study population, from the beginning of the pandemic in February 2020 to February 2021, ranged from 16.2% to 19.8% depending on the matching approach used.

Shakiba et al. assessed the seropositivity of COVID-19 in Guilan Province using a population-based cluster random sampling method with a design effect of 1.24 among 198 households and using rapid test kits

(VivaDiag COVID-19 IgM/IgG) (16). They reported prevalence of 33% for COVID-19 from February to April 2020 and estimated that between 518 000 and 777 000 people had been infected. Poustchi et al. determined the seropositivity of COVID-19 in 18 Iranian cities using population-based cluster random sampling of 244 samples and PishtazTeb SARS-CoV2 ELIZA

kits approved by the Iranian Food and Drug Administration (17). They reported a prevalence of 72.6% (53.9–92.8%) for COVID-19 from February to 2 June 2020 in Rasht City (capital of Guilan Province). It is important to mention that many researchers and participants were involved in both studies. The estimated prevalence of both studies was higher than that of our study. There are a number of possible explanations for the higher prevalence in the studies by Shakiba et al. (16) and Poustchi et al. (17): design effect, statistical error, and high sensitivity and specificity of the test kit in the former study, and high uncertainty of sampling, problems in calculating the design effect, and lack of clustering control in estimating sampling error in the latter study. Similar issues were reported by Cassaniti et al. (18).

Khalagi et al. described the prevalence of COVID-19 among the general Iranian population using stratified random sampling of 858 samples from February to 20 August 2020 (19). They used the PishtazTeb SARS-CoV2 ELIZA kits and reported 8% (5–12.5%) prevalence for COVID-19. This under-reporting and discrepancy with our study may contribute to under-estimation of the number of infected people with negative serological test results, which will increase as the pandemic continues. The higher prevalence of 16.2% reported in our study was probably because the capture–recapture analysis was based on valid and reliable data. The Medical Care Monitoring Center is the most reliable source of inpatient hospital data in the Islamic Republic of Iran. The Primary Care Registry uses outpatient primary healthcare data, and there is weekly quality control of records by The Ministry of Health and Medical Education at the national and provincial levels. Therefore, optimal data sources with a high rate of variable completion were used in our study. However, the validity of our results could have been improved by greater overlap across sources.

There were several assumptions with regard to the 4 criteria for appropriate 2-source capture–recapture analyses listed under the capture–recapture method. The first assumption was that all capture and recapture samples were chosen for the same 1-year period. The second assumption was that records for both surveillance systems were gathered by separate reporting sources. However, if there was a large overlap between sources, then this would have resulted in an underestimation of prevalence (data sources were positively dependent). Conversely, if there was little overlap between the sources (data sources were negatively dependent), then an overestimation of the prevalence would have resulted. The estimated prevalence of COVID-19 was 16.2–19.8% and the overlap between the sources was 2.3%, meaning that there may have been an overestimation. The third assumption was that the Medical Care Monitoring Center data focused on inpatient COVID-19 cases and may have been captured first by the Primary Care Registry because both registries were present in each district of the study area. The final assumption was that the capture history of all cases was accurate.

A limitation of the study was that the Medical Care Monitoring Center only focused on inpatients and included an insufficient number of outpatient cases. We confirm that the records captured

by each of these official sources were accurately recorded. The capture–recapture method requires that record matching must be performed appropriately. In this study, we used a combination of exact matching with relaxed matching approaches, using key variables from records of both sources to perform the matching.

Conclusion

The ongoing COVID-19 pandemic requires simple and accurate techniques such as capture–recapture to achieve better estimates of prevalence. The method described in this study can be easily replicated in most settings, even using the less restrictive matching criteria. The estimated COVID-19 prevalence in our study was lower than that reported by seroprevalence methods. The prevalence measured by the seroprevalence method was likely to be higher than that based on official COVID-19 case numbers in the study area. Despite the large uninfected population in the study area and the increasing number of daily confirmed and suspected COVID-19 cases, the results of this study can correct the misunderstanding of policy-makers about the results of seroprevalence surveys.

Funding: None

Competing interests: None declared.

Comparaison de la méthode de capture-recapture et de l'enquête de séroprévalence pour l'estimation de la prévalence de la COVID-19 en République islamique d'Iran

Résumé

Contexte : Une estimation fiable de la prévalence est importante pour le suivi et l'évaluation des programmes de prévention de la COVID-19 au sein des populations à risque.

Objectifs : Nous avons comparé la méthode de capture-recapture avec une enquête de séroprévalence visant à obtenir une estimation précise de la prévalence de la COVID-19 sur une période d'un an dans la province de Guilan, au nord de la République islamique d'Iran.

Méthodes : Nous avons utilisé la méthode de capture-recapture pour estimer la prévalence de la COVID-19. Les dossiers du système de registre des soins de santé primaires et du Medical

WHO EMRO | Comparison of the capture–recapture method and seroprevalence survey for estimation of

Care Monitoring Center (Centre de suivi des soins médicaux) ont été comparés à l'aide de quatre approches d'appariement basées sur des combinaisons des variables suivantes : nom, âge, genre, date de décès, cas positifs ou négatifs et cas vivants ou morts.

Résultats : Selon l'approche d'appariement utilisée, la prévalence estimée de la COVID-19 dans la population de l'étude depuis le début de la pandémie en février 2020 jusqu'à la fin de janvier 2021 était comprise entre 16,2 % et 19,8 %, soit un niveau inférieur à celui des études précédentes.

Conclusion : La méthode de capture-recapture pourrait fournir une meilleure précision que les enquêtes de séroprévalence pour mesurer la prévalence de la COVID-19. Elle pourrait également permettre de réduire le biais dans l'estimation de la prévalence et de corriger l'idée erronée que se font les responsables de l'élaboration des politiques des résultats des enquêtes de séroprévalence.

مقارنة بين طريقة "الالتقاط وإعادة الالتقاط" ومسح الانتشار المصلي في تقدير معدل انتشار كوفيد-19 في جمهورية إيران الإسلامية

علي كياكالايه، مرتضى تارامساري، رضا محمدي، سجاد كياكالايه، حسن كاواكبور

الخلاصة

: من المهم تقدير معدل انتشار كوفيد -19 على نحو موثوق فيه، من أجل رصد برامجالخلفية
وقاية الفئات السكانية المعرضة للخطر وتقييم هذه البرامج .

: هدفت هذه الدراسة الى اجراء مقارنة بين طريقة " الالتقاط وإعادة الالتقاط "الأهداف
ومسح الانتشار المصلي من ناحية مدى دقة تقدير معدل انتشار كوفيد -19 خلال سنة واحدة في
محافظة جيلان بشمال جمهورية إيران الإسلامية .

: استخدمنا طريقة " الالتقاط وإعادة الالتقاط " لتقدير معدل انتشار كوفيد -19.طرق البحث
وأجرينا مقارنة لسجلات مأخوذة من نظام تسجيل الرعاية الأولية ومركز رصد الرعاية
الطبية، باستخدام 4 أساليب للمطابقة تستند إلى مزيج من المتغيرات التالية : الاسم،
والعمر، ونوع الجنس، وتاريخ الوفاة، والحالات الإيجابية أو السلبية، والحالات الحية
أو الميتة .

8 / 10

[illegible] 19- [illegible] : [illegible] 2021 [illegible] /[illegible] 2020 [illegible] /[illegible] [illegible] .19.8% 16.2% .[illegible]

[illegible] "[illegible]" [illegible] : [illegible] .19- [illegible] [illegible] .[illegible]

References

1. Chen Z-M, Fu J-F, Shu Q, Chen Y-H, Hua C-Z, Li F-B, et al. Diagnosis and treatment recommendations for pediatric respiratory infection caused by the 2019 novel coronavirus. World J Pediatr. 2020 Jun;16(3):240–6. https://doi.org/10.1007/s12519-020-00345-5 PMID:32026148

2. Yu A, Wang Z, Ren W, Wu Z, Hu Z, Li L, et al. Epidemic analysis of COVID-19 in China after Wuhan was restricted. Res Square. 2020 Feb 24. https://doi.org/10.21203/rs.2.24289/v1

3. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020 Feb 15;395(10223):507–13. https://doi.org/10.1016/S0140-6736(20)30211-7 PMID:32007143

4. Coronavirus disease (COVID-2019) situation reports [website]. Geneva: World Health Organization (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports, accessed 23 November 2022).

5. Population-based age-stratified seroepidemiological investigation protocol for coronavirus 2019 ([illegible]COVID-19) [illegible] infection, 26 May 2020. Geneva: World Health Organization; 2020 (https://apps.who.int/iris/handle/10665/332188, accessed 23 November 2022).

6. Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. Epidemiol Rev. 1995;17(2):243–64. https://doi.org/10.1093/oxfordjournals.epirev.a036192 PMID:8654510

7. Capture-recapture and multiple-record systems estimation I: history and theoretical development. International Working Group for Disease Monitoring and Forecasting. Am J Epidemiol. 1995 Nov 15;142(10):1047–58. PMID:7485050

8. Capture-recapture and multiple-record systems estimation II: applications in human diseases. International Working Group for Disease Monitoring and Forecasting. Am J Epidemiol. 1995 Nov 15;142(10):1059–68. PMID:7485051

9. Rahi J, Dezateux C. Measuring and interpreting the incidence of congenital ocular anomalies: lessons from a national study of congenital cataracts in the UK. Invest Opthalmol Vis Sci. 2001 Jun;42(7):1444–8. PMID:11381045

10. Davoudi-Kiakalayeh A, Mohammadi R, Stark-Ekman D, Yousefzade-Chabok S, Behboudi F, Jansson B. Estimating drowning deaths in Northern Iran using capture-recapture method. Health Policy. 2011 May;100(2–3):290–6. https://doi.org/10.1016/j.healthpol.2010.09.005 PMID:20951461

11. Piriyawat P, Smajsov M, Smith MA, Pallegar S, Al-Wabil A, Garcia NM, et al. Comparison of active and passive registry for cerebrovascular disease. Am J Epidemiol. 2002 Dec 1;156(11):1062–9. https://doi.org/10.1093/aje/kwf152

12. Hsu VP, Staat MA, Roberts N, et al. Use of active registry to validate international classification of diseases code estimates of rotavirus hospitalizations in children. Pediatrics. 2005 Jan;115(1):78–82. https://doi.org/10.1542/peds.2004-0860 PMID:15629984

13. Wittes JT, Sidel VW. A generalization of the simple capture–recapture model with applications to epidemiological research. J Chronic Dis. 1968 Aug;21(5):287–301. https://doi.org/10.1016/0021-9681(68)90038-6 PMID:5675416

14. Zavareh D K, Mohammadi R, Laflamme L, Naghavi M, Zarei A, BJA Haglund. Estimating road traffic mortality more accurately: use of the capture –recapture method in the West Azarbaijan Province of Iran. Int J Inj Contr Saf Promot. 2008 Mar;15(1):9–17. https://doi.org/10.1080/17457300701794105 PMID:18344091

15. Statistical Center of Iran 2015, from http://www.sci.org.ir/ portal/faces/public/sci_ en/sci_en. Selected data

16. Shakiba M, Nazari SSH, Mehrabian F, Rezvani SM, Ghasempour Z, Heidarzadeh A. Seroprevalence of COVID-19 virus infection in Guilan province, Iran. MedRxiv. 2020 May 1. https://doi.org/10.1101/2020.04.26.20079244.

17. Poustchi H, Darvishian M, Mohammadi Z, Shayanrad A, Delavari A, Bahadorimonfared A, et al. SARS-CoV-2 antibody seroprevalence in the general population and high-risk occupational groups across 18 cities in Iran: a population-based cross-sectional study. Lancet Infect Dis. 2021 Apr;21(4):473–81. https://doi.org/10.1016/S1473-3099(20)30858-6 PMID:33338441

18. Cassaniti I, Novazzi F, Giardina F, Salinaro F, Sachs M, Perlini S, et al. Performance of VivaDiag COVID-19 IgM/IgG Rapid Test is inadequate for diagnosis of COVID-19 in acute patients referring to emergency room department. J Med Virol. 2020 Oct;92(10):1724–7. https://doi.org/10.1002/jmv.25800 PMID:32227490

19. Khalagi K, Gharibzadeh S, Khalili D, Mansournia MA, Samiee SM, Aghamohamadi S, et al. Prevalence of COVID-19 in Iran: Results of the first survey of the Iranian COVID-19 Serological Surveillance program. Clin Microbiol Infect. 2021 Nov;27(11):1666–71. https://doi.org/10.1016/j.cmi.2021.06.002 PMID:34111585

Friday 12th of July 2024 03:01:27 AM