IMPACT OF UNDERREPORTED SARS-COV-2 INFECTIONS IN SURVEILLANCE SYSTEMS ON ASSOCIATION STUDIES, WITH AN APPLICATION TO LONG-TERM EXPOSURE TO AIR POLLUTION

Veronesi Giovanni[1], Giusti Emanuele Maria[1], Francesco Gianfagna[1,2], Marco Mario Ferrario[1]

Affiliations
1. Centro Ricerche in Epidemiologia e Medicina Preventiva (EPIMED), Dipartimento di Medicina e Chirurgia, Università degli Studi dell'Insubria, Varese
2. Mediterranea Cardiocentro, Napoli

Introduction
Administrative healthcare data are affected by SARS-CoV-2 cases underreporting, especially after the opening of the vaccination campaign: the proportion of cases that were reported to the surveillance system in Italy between January 2021 and February 2022 was about 25% [1]. Most of the prospective studies investigating the link between long-term exposure to air pollutants with SARS-CoV-2 were conducted before the vaccination campaign [2-4]. Whether air pollution increases the risk of breakthrough infections, especially in the elderly, remains to be established.

Aims
We investigate bias due to cases underreporting in simulations, and use these results to guide the interpretation of real-world estimates of the association between long-term exposure to air pollution and breakthrough SARS-CoV-2 infection in different pandemic periods, in a population-based cohort of vaccinated elderly.

Methods
Simulation study.
We simulated one continuous exposure variable X from a standard normal distribution; time-to-event from the exponential distribution assuming a true effect for X (on the log-scale) of either 0.10 or 0.25; an independent censorship mechanism; and 3 Bernoulli random variables Z_{ND}, Z_{D1} and Z_{D2} whose success probabilities p_{ND}, p_{D1} and p_{D2} represent underreporting rate. p_{ND} was independent on X (non-differential underreporting); p_{D1} was lower at lower values of X (i.e lower event underreporting amongst less exposed); p_{D2} was lower at higher X values. Events were switched to non-events for Z_{ND}, Z_{D1} or Z_{D2} =1. We considered 2 sample sizes (2500 and 5000), 3 underreporting rates (25%, 50% and 75%), and 100 runs for every combination of the above.
Real-word study.
N=1505 participants to the RoCAV (Risk Of Cardiovascular diseases and abdominal aortic Aneurysm in Varese) population-based cohort[5] underwent a re-examination visit in 2021-22. N=1363 of these with at least one COVID-19 vaccine dose (source: Local Health Agency database) and no evidence of SARS-CoV-2 infection before vaccine dose 1 qualified for the analyses. The outcome was first positive swab (molecular or antigen) after vaccination in the Local Health Agency administrative healthcare dataset. We considered two follow-up periods: restricted - until the end of the pandemic emergency (March 31st, 2022); and full - until Dec 31st, 2022. The 2019 annual mean levels of Particulate Matter ($PM_{2.5}$ and PM_{10}), Nitrogen Dioxide (NO_2) and Ozone (O_3) were estimated from combined use of chemical-transport and random-Forest models (spatial resolution: 1Kmq), and assigned based on the latest individuals' residency address.
We estimated the associations between air pollutants and infection in Cox models with time-dependent COVID-19 vaccine exposure (1-5 doses) and adjusting for age, sex, current smoking, autoimmune disease and use of immunosuppressive treatments in the last six months, in the restricted and in the full follow-up periods. Hazard Ratios (HR) and 95%CI are expressed for 1 interquartile range (IQR) increase in air pollutants (1.2, 1.0, 3.6 and 1.9 μg/m^3 for $PM_{2.5}$, PM_{10}, NO_2 and O_3, respectively).

Results

Simulation study

Table 1 reports the results of the simulations for n=2500 and true effect of X of 0.10. Under ND, point estimates are biased toward the null: in the worst underreporting rate scenario, the mean bias was -13% of the true effect size, but coverage remained as high as 96%. Under D1 (lower underreporting rates in less exposed), downward bias was substantial (-46% of the true effect at 25% underreporting rate), and coverage was low. Under D2 the associations are over-estimated, the mean bias being 50% of the true effect in the lowest underreporting rate scenario. Findings for n=5000 or true effect of X of 0.25 were consistent with these.

Real-world data

Mean±SD age was 74.5±4.8 years, 52% were men, 12% current smokers. We observed n=130 SARS-CoV-2 cases up to March 31st, 2022; and additional 246 up to Dec 31st, 2022, corresponding to event rates of 2.9 and 5.0 per 10000 person-days in the two follow-up periods, respectively.

We estimated a HR of 1.12 (95%CI: 0.94-1.33) for IQR increase in $PM_{2.5}$ in the restricted follow-up period; and of 1.04 (0.94-1.14) in the full period. Similarly, HRs for PM_{10} and NO_2 were 1.10 (0.94-1.27) and 1.05 (0.87-1.25) in the restricted; and 1.05 (0.97-1.15) and 0.96 (0.87-1.07) in the full follow-up periods, respectively. O_3 was inversely associated with breakthrough infection risk in both follow-up periods.

Table 1. Results of the simulation study, for n=2500 and true effect of X=0.10, under non-differential and differential scenarios with different underreporting rates.

Underreporting mechanism and rate		Mean Estimate	SD Estimate	Mean bias	Mean SE	Mean MSE	Coverage %
Non differential (ND)							
	Underreporting 25%	0.105	0.074	0.005	0.075	0.011	0.97
	Underreporting 50%	0.094	0.085	-0.006	0.093	0.016	0.99
	Underreporting 75%	0.087	0.124	-0.013	0.132	0.033	0.96
Differential, D1							
	Underreporting 25%	0.054	0.074	-0.046	0.074	0.013	0.89
	Underreporting 50%	0.034	0.096	-0.066	0.091	0.022	0.89
	Underreporting 75%	-0.044	0.129	-0.144	0.128	0.054	0.78
Differential, D2							
	Underreporting 25%	0.151	0.071	0.051	0.076	0.013	0.89
	Underreporting 50%	0.171	0.085	0.071	0.093	0.021	0.90
	Underreporting 75%	0.253	0.140	0.153	0.136	0.061	0.75

D1: less exposed have lower probability of underreporting than more exposed. D2: less exposed have higher probability of underreporting than more exposed. Abbreviations: SD=Standard Deviation; SE=Standard Error; MSE=Mean Squared Error. Coverage % is the proportion of simulation runs for which the true effect of X falls within the estimated 95% confidence interval.

Conclusions

Administrative healthcare datasets are affected by SARS-CoV-2 cases underreporting [1]. Under a non-differential mechanism, downward bias was limited and coverage was high, even for reduced sample sizes, small true beta effects, and high underreporting. This was not the case for differential underreporting mechanisms. Long-term exposure to air pollution has been linked with COVID-19 severity [2,4]. The shift of infection towards vaccinated individuals may have resulted in higher amount of asymptomatic or pauci-symptomatic infections, which are less likely to be detected by the surveillance system, especially so after the end of the emergency period in March 2022. This picture corresponds to our simulated D2 scenario. However, in the restricted vs. full follow-up period we generally found larger, and not lower, HR estimates, in contrast with simulation results where overestimating bias increased with increasing underreporting probability. Hence, in our population of elderly, under-reporting of positive cases through surveillance systems is more likely to have been non-differential with respect to pre-pandemic levels of air pollution, potentially resulting in downward bias of the associations. A careful case-by-case evaluation of the underreporting mechanisms is needed when using administrative healthcare dataset for SARS-CoV-2 research.

Bibliografia

[1] Marziano V, Guzzetta G, Menegale F, et al. Estimating SARS-CoV-2 infections and associated changes in COVID-19 severity and fatality. Influenza Other Respir Viruses. 2023;17(8):e13181. doi:10.1111/irv.13181

[2] Zhang J, Lim YH, So R, et al. Long-term exposure to air pollution and risk of SARS-CoV-2 infection and COVID-19 hospitalisation or death: Danish nationwide cohort study. Eur Respir J. 2023;62(1). doi:10.1183/13993003.00280-2023.

[3] Veronesi G, De Matteis S, Calori G, Pepe N, Ferrario MM. Long-term exposure to air pollution and COVID-19 incidence: a prospective study of residents in the city of Varese, Northern Italy. Occup Environ Med. 2022;79(3):192-199. doi:10.1136/oemed-2021-107833

[4] Ranzani O, Alari A, Olmos S, et al. Long-term exposure to air pollution and severe COVID-19 in Catalonia: a population-based cohort study. Nat Commun. 2023;14(1):2916. doi:10.1038/s41467-023-38469-7

[5] Gianfagna F, Veronesi G, Bertù L, et al. Prevalence of abdominal aortic aneurysms and its relation with cardiovascular risk stratification: protocol of the Risk of Cardiovascular diseases and abdominal aortic Aneurysm in Varese (RoCAV) population based study. BMC Cardiovasc Disord. 2016;16:243. doi:10.1186/s12872-016-0420-2