

Supplementary file 1: STROBE checklist

Checklist of items that should be included in reports of cohort studies

The text in italic font is copied from the STROBE recommendations (Strengthening the Reporting of Observational Studies in Epidemiology) [1]. Using regular font, we explain how we addressed each STROBE item in our study: the parts in black are mainly copied from the manuscript; the parts in blue are written out in the present supplement but not in the main manuscript.

1. *Title and abstract. (a) Indicate the study's design with a commonly used term in the title or the abstract. (b) Provide in the abstract an informative and balanced summary of what was done and what was found.*

The title is "Prevalence and incidence of anti-SARS-CoV-2 antibodies among healthcare workers in Belgian hospitals before vaccination: a prospective cohort study". The manuscript contains a structured abstract.

2. *Background/rationale. Explain the scientific background and rationale for the investigation being reported.*

This information is given in the introduction section.

3. *Objectives. State specific objectives, including any prespecified hypotheses.*

The objectives were formulated as follows: "We thus started a prospective cohort study end of April 2020, aiming to follow the prevalence and incidence of anti-SARS-CoV-2 antibodies among Belgian hospital healthcare workers throughout the epidemic, in order to guide infection prevention and control measures in hospitals and support planning of healthcare resources. In addition, we sought to investigate the presence of symptoms, positive PCR results and neutralising antibodies in seropositive participants, and to describe these variables over time in seroconverters. In this paper we present the findings up to the end of December 2020, before the start of the vaccination of healthcare workers in Belgium mid-January 2021."

4. *Study design. Present key elements of study design early in the paper.*

Key elements of study design are given in title, abstract, the last paragraph of the introduction and the first paragraph of the methods.

5. *Setting. Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection.*

These elements are given in the subsection about the study population (methods).

6. *Participants. (a) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up. (b) For matched studies, give matching criteria and number of exposed and unexposed.*

These elements are given in the subsection about the study population (methods). The study does not involve matching.

7. *Variables. Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.*

The main outcome variable was the presence or absence of anti-SARS-CoV-2 antibodies, the assessment of which is described in the methods (under the subtitle: "Assessment of anti-SARS-CoV-2 antibodies"). Other variables of interest were the presence of symptoms compatible with COVID-19, SARS-CoV-2 molecular test results, and virus neutralisation test results. These variables were not considered to represent exposures, predictors, confounders, or effect modifiers because the study did not explore causal relations.

8. *Data sources/measurement. For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group.*

For each variable of interest, the assessment is described in a separate paragraph (under the subtitle "assessment of covariates").

9. *Bias. Describe any efforts to address potential sources of bias.*

To reduce the risk of observer bias, the people in charge of the laboratory tests did not have access to information about the participants or other test results at that time point. The only exception was the neutralisation test, which was only done when the ELISA result was positive. Furthermore, the participants filled out the online questionnaire before they knew their RT-qPCR and ELISA results at that time point, although they may have known results of tests that were done outside this study. Another potential source of bias is the lack of accuracy of the ELISA used to measure the main outcome. To address this, we present the findings over time (repeated measures) and we report symptoms, RT-qPCR, and neutralisation test results in addition to the ELISA. Finally, although this study was designed to include a representative sample of healthcare workers in Belgian hospitals, missing values may still have introduced bias.

10. *Study size. Explain how the study size was arrived at.*

The prevalence of anti-SARS-CoV-2 antibodies among HCW was unknown at the time this study was designed. The sample size calculation was based on an estimated seroprevalence of 50%, a desired absolute precision of 5%, and a design effect of 2. We set the estimated seroprevalence to 50% because that is a conservative approach (leading to a large sample size) and because we expected to find a seroprevalence of that order of magnitude. This led to a target sample size of 800 HCW, i.e. 16 clusters of 50 individuals, but in order to have an additional margin, we decided to include 17 clusters.

11. *Quantitative variables. Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why.*

The quantitative variables in this study are the participants' age, follow-up time, number of staff per hospital, ELISA results and neutralisation test results.

- Age, follow-up time, and number of staff were analysed and summarised as quantitative variables, using medians (interquartile range) and totals.
- ELISA results were dichotomised (positive or negative) as indicated by the manufacturer: if the ratio of the serum sample optical density/negative control optical density was >1.1, the corresponding serum sample was defined as ELISA positive. Otherwise, it was negative.
- Neutralisation test results were dichotomised as follows: if the serum titre of antibodies needed to neutralise 50% of the SARS-CoV-2 virus was 1:50 or higher, the corresponding serum sample was defined as positive on the neutralisation test. Otherwise, it was negative.

12. *Statistical methods. (a) Describe all statistical methods, including those used to control for confounding. (b) Describe any methods used to examine subgroups and interactions. (c) Explain how missing data were addressed. (d) If applicable, explain how loss to follow-up was addressed. (e) Describe any sensitivity analyses.*

The methods include a subsection summarising the statistical methods. Because this study did not explore associations between variables, we did not control for confounding. There were no specific examinations of subgroups. A detailed report of the missing data (and monotone/non-monotone missingness patterns) is given in a supplemental file. No sensitivity analyses were done.

We first described the prevalence and the incidence rate of positive ELISA results in the study population and inferred these findings to the target population of all healthcare workers in Belgian hospitals. Because of the sampling design, not all healthcare workers had the same probability of being selected. This was corrected at the level of the analysis by using weights consisting of a hospital and an individual component. The computation of the 95% confidence intervals also took account of the cluster design.

When we selected the hospitals for this study, the number of employees was not available for all the general hospitals in Belgium (n=103) so we used the number of beds as a proxy. For the selected hospitals (n=17), we did use the number of employees in the calculation of the weights. Hence, the weighting takes account of both the number of beds (correction for first sampling stage) and the number of staff (correction for second sampling stage) in the selected hospitals. The weights were calculated as follows:

$$w_i = \frac{52651 * \text{number of staff in hospital}_i}{\text{number of beds in hospital}_i * 17 * 50}$$

where w_i is the weight for a healthcare worker from hospital i ; 52651 is the total number of beds in all eligible hospitals; 17 is the number of sampled hospitals; and 50 is the number of healthcare workers selected in each hospital.

13. *Participants. (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed. (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram.*

This information is given in the first paragraph of the results (under the subtitle “General characteristics of hospitals and healthcare workers”) and in a supplemental file describing missingness patterns.

14. *Descriptive data. (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders. (b) Indicate number of participants with missing data for each variable of interest. (c) Summarise follow-up time (eg, average and total amount).*

The characteristics of the study participants are described in supplemental table 1. Information about follow-up is summarised in the third paragraph of the results and in the supplemental file with missingness patterns.

15. *Outcome data. Report numbers of outcome events or summary measures over time.*

This is addressed under the subtitle “Prevalence and incidence rate of anti-SARS-CoV-2 antibodies”.

16. *Main results. (a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included. (b) Report category boundaries when continuous variables were categorized. (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period.*

This is addressed under the subtitle “Prevalence and incidence rate of anti-SARS-CoV-2 antibodies”.

17. *Other analyses. Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses.*

Not applicable.

18. *Key results. Summarise key results with reference to study objectives.*

This first and the last paragraph of the discussion summarise the key results.

19. *Limitations. Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.*

The discussion contains a specific paragraph where the limitations are listed and the implications discussed.

20. *Interpretation. Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.*

This is addressed in several paragraphs of the discussion.

21. *Generalisability. Discuss the generalisability (external validity) of the study results.*

To our knowledge, this is the first study reporting on SARS-CoV-2 seroprevalence in a nationwide representative sample of hospital healthcare workers, allowing us to infer our findings to this population. Until September 2020 (the part of the study before its extension), the number of participants lost to follow up was low. Furthermore, the total duration of follow-up was long (until end December 2020), which provides a clear view of the serological status of HCW just before the start of the vaccination campaign (in January 2021).

22. *Funding. Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based.*

This study was funded by Sciensano, the Belgian institute of public health, Brussels, Belgium. Sciensano was involved in all stages of the study, from conception and implementation to analysis and reporting. Researchers were independent from funders. All authors had full access the data in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147(8): 573-577.