



Community-level evidence for SARS-CoV-2 vaccine protection of unvaccinated individuals

Oren Milman^{1,5}, Idan Yelin^{1,5}, Noga Aharony¹, Rachel Katz², Esma Herzel², Amir Ben-Tov^{1,2,3}, Jacob Kuint^{2,3}, Sivan Gazit^{1,2}, Gabriel Chodick^{1,2,3}, Tal Patalon^{1,2}✉ and Roy Kishony^{1,4}✉

Mass vaccination has the potential to curb the current COVID-19 pandemic by protecting individuals who have been vaccinated against the disease and possibly lowering the likelihood of transmission to individuals who have not been vaccinated. The high effectiveness of the widely administered BNT162b vaccine from Pfizer-BioNTech in preventing not only the disease but also infection with SARS-CoV-2 suggests a potential for a population-level effect, which is critical for disease eradication. However, this putative effect is difficult to observe, especially in light of highly fluctuating spatiotemporal epidemic dynamics. Here, by analyzing vaccination records and test results collected during the rapid vaccine rollout in a large population from 177 geographically defined communities, we find that the rates of vaccination in each community are associated with a substantial later decline in infections among a cohort of individuals aged under 16 years, who are unvaccinated. On average, for each 20 percentage points of individuals who are vaccinated in a given population, the positive test fraction for the unvaccinated population decreased approximately twofold. These results provide observational evidence that vaccination not only protects individuals who have been vaccinated but also provides cross-protection to unvaccinated individuals in the community.

The Pfizer-BioNTech COVID-19 BNT162b vaccine is highly effective at preventing disease and infection at the individual and community levels, as demonstrated in clinical trials, as well as in real-world vaccination campaigns^{1–5}. Furthermore, among vaccinated individuals infected with SARS-CoV-2, a lower viral load was observed⁶. Reduced infection and viral load suggest that reduced transmission occurred. However, vaccination could, in principle, also increase transmission due to behavioral effects, as vaccinated individuals may not quarantine after contacting a patient with COVID-19, or may be less mindful of social-distancing measures⁷. It is therefore unclear whether, overall, vaccination reduces transmission at the population level, thereby conferring protection for those who are not vaccinated, such as individuals who are immunodeficient or those who are currently ineligible for vaccination^{4,8–10}. Moreover, as the reproduction number varies with sociobehavioral and environmental factors even in the absence of vaccination and because global disease rates represent both vaccinated and unvaccinated individuals, it has proven challenging to determine the effect of vaccination on community-level SARS-CoV-2 transmission^{11–13}.

The rapid vaccine rollout in Israel, initiated on 19 December 2020 and covering almost 50% of the population within 9 weeks, presents a unique opportunity to test these questions using real-world

data. Capitalizing on differences in vaccination rates among geographically distinct communities, and on the availability of an unvaccinated bystander cohort of individuals below 16 years of age for whom the vaccine was not authorized in the first stages of vaccine rollouts, we asked whether and to what extent the fraction of patients vaccinated in each community affects the risk of infection in an unvaccinated cohort of individuals under 16 years old within this same community.

We focused our analysis on the vaccination rates and test results of 177 distinct communities with a presumed low rate of natural immunization as inferred by a low fraction of individuals infected with SARS-CoV-2. We retrieved the vaccination dates and test results, from 9 December 2020 to 9 March 2021, of members of Maccabi Healthcare Services (MHS), Israel's second largest healthcare maintenance organization. We defined geographical communities based on residence codes, and identified 246 communities that each comprised a sufficient number of tests and people (Methods). As both vaccination and natural infection could render individuals immunized, thereby possibly conferring protection to unvaccinated individuals, high infection rates could mask the effect of vaccination-induced immunity. To minimize the confounding effect of natural immunization, we considered the test record since 1 March 2020, and included only communities in which the fraction of people tested positive by 9 March 2021 was lower than 10% (177 communities; Methods and Extended Data Fig. 1). This exclusion of communities did not introduce notable biases in the age distributions of neither the vaccinated individuals nor the patients with a confirmed SARS-CoV-2 infection (Extended Data Fig. 2).

To control for the spatiotemporally dynamic nature of the epidemic, we focused on relative changes in positive test fraction within each community between fixed time intervals, thus accounting for global temporal patterns as well as for intrinsic differences among communities. First, for each community, the mean cumulative fraction of patients inoculated with the first dose of the vaccine was calculated for two consecutive 3-week intervals (V_1 , V_2) for individuals aged 16–50 years, who are assumed to represent the population that is likely to interact with the bystander unvaccinated cohort of under 16 years of age¹⁴ (Fig. 1a). Second, for each such vaccination time interval, a corresponding time-shifted 3-week testing interval was defined with a delay of 28 d, to allow for the putative immunization of the vaccinated individuals and subsequent cross-protection of unvaccinated individuals to take effect (Fig. 1a,b). For these two testing intervals, the positive test fraction among the unvaccinated bystander cohort (P_1 , P_2) was calculated (Fig. 1b). Comparing periods 1 and 2, we then considered the relationship between the change

¹Faculty of Biology, Technion-Israel Institute of Technology, Haifa, Israel. ²Maccabitech, Maccabi Healthcare Services, Tel Aviv-Yafo, Israel. ³Sackler Faculty of Medicine, Tel-Aviv University, Tel Aviv-Yafo, Israel. ⁴Faculty of Computer Science, Technion-Israel Institute of Technology, Haifa, Israel. ⁵These authors contributed equally: Oren Milman, Idan Yelin. ✉e-mail: patalon_t@mac.org.il; rkishony@technion.ac.il

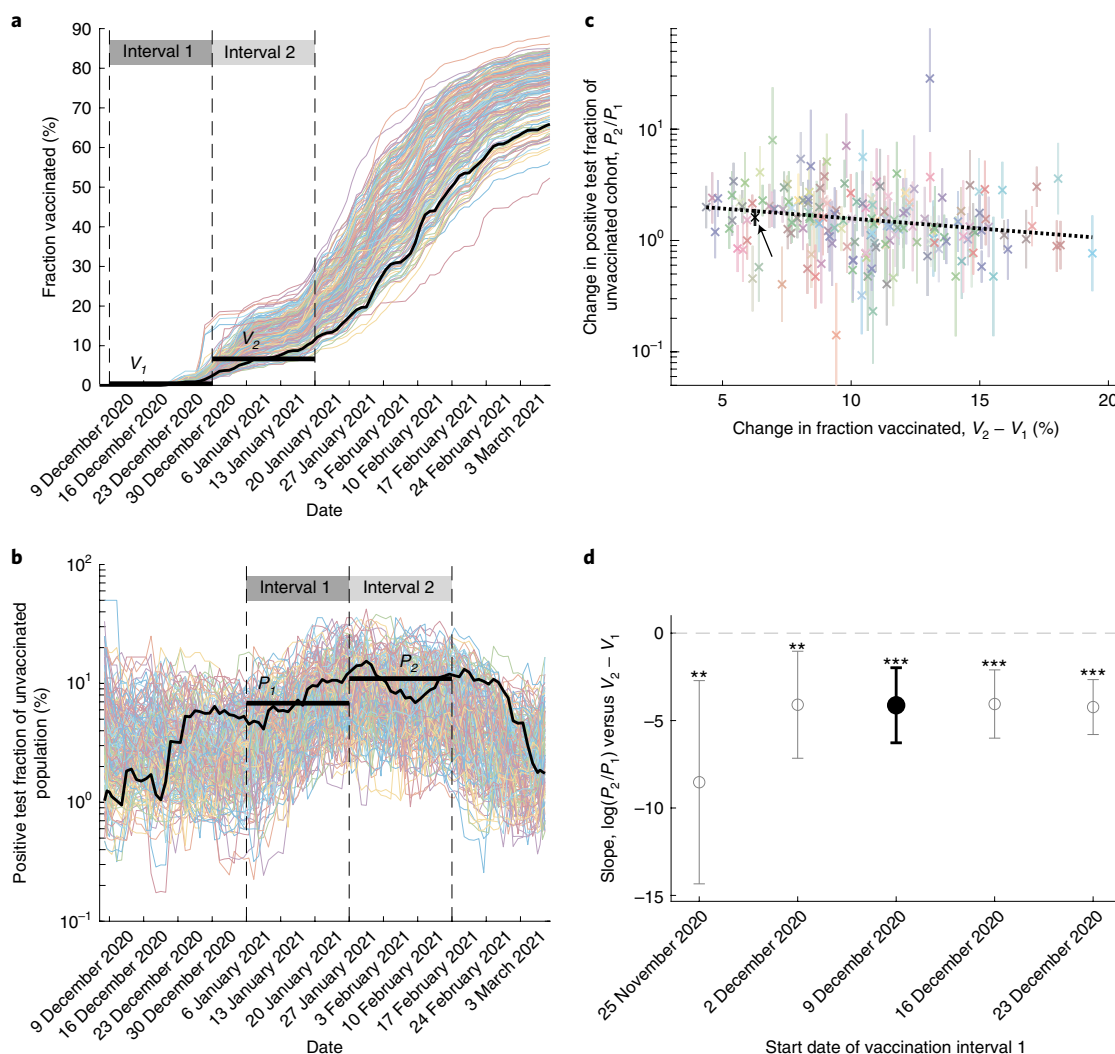


Fig. 1 | An increase in the fraction of vaccination in a community is associated with a later reduction in the number of confirmed infections of the unvaccinated cohort in the same community. a, b, The accumulated fraction vaccinated among 16–50-year-old individuals (**a**) and the positive test fraction among the 0–15-year-old bystander unvaccinated cohort (**b**, 7-d moving window average; to avoid a logarithm of zero, time windows with zero positive tests are plotted as having a single positive test) are shown for each community as a function of time. The vaccination intervals (**a**) and their corresponding time-shifted testing intervals (**b**) are indicated above each panel (dark gray and light gray). The black trajectory highlights an example community, indicating its mean vaccination fraction for each of the intervals (V_1 , V_2), and the mean positive test fractions in the testing intervals (P_1 , P_2) (see Methods; the example community is highlighted only for demonstration). **c,** For each community ($n=177$), the change in positive test fraction between time intervals (P_2/P_1) is shown as a function of the change in the vaccinated fraction in the corresponding vaccination intervals ($V_2 - V_1$). Data are mean ± 1 s.e.m. The dotted line shows linear fit ($P=0.0003$) (see Methods). The community highlighted in **a** and **b** is indicated (black x, error bar and arrow). **d,** The association is robust to a shift in interval dates. For the same communities ($n=177$), slopes and P values are shown as a function of the first vaccination interval start date (keeping the same delay of 28 d between vaccination and test intervals). Circles and error bars represent maximum likelihood estimator (MLE) ± 2 s.e. (95% confidence intervals). The slope corresponding to the intervals used in **a–c** is indicated (filled black circle). ** $P < 0.005$; *** $P < 0.0005$.

in positive test fraction of the unvaccinated cohort (P_2/P_1) and the change in vaccination ($V_2 - V_1$) (Fig. 1c).

The positive test fraction in the unvaccinated cohort decreased in proportion to the rate of vaccination in each community. A linear fit showed a strong negative association between the log-transformed odds ratio of the positive test fraction and the increase in vaccination fraction ($\log(P_2/P_1) = a(V_2 - V_1) + b$, where $a = -4.1$ (95% confidence interval, -6.3 to -2), $P = 0.0003$) (Fig. 1c; see Methods, ‘Correlation analysis’). For each 20 percentage points of vaccinated individuals, the positive test fraction for unvaccinated individuals decreased approximately twofold; such amplification could appear owing to multiple rounds of infection cycles. These results were

robust to changes in interval length ($a = -3.6$ (95% confidence interval, -4.8 to -2.3), $P < 10^{-4}$ for 4-week intervals), interval dates (Fig. 1d), and member and community exclusion thresholds (Extended Data Figs. 3, 4). Furthermore, consistent with the known time for vaccine-induced immunity¹ plus one or more infection cycles, the effect appears robustly for time shifts beyond 24 d (Extended Data Fig. 5). Considering the age of the vaccinated cohort, we find that the association was stronger when the considered vaccinated cohort was younger than the reference vaccinated cohort of 16–50 years of age, and became non-significant for vaccination cohorts of individuals aged 66 years and older (Extended Data Fig. 6), consistent with an assumed lower level of social interactions between the two

age groups¹⁴. We also observed a strong negative correlation when considering, for each community, the positive test fraction per capita instead of per test ($a = -8.7$ (95% confidence interval, -11.3 to -6.1), $P < 10^{-4}$). Taken together, our analysis shows a strong and robust negative association between the vaccination of adults in the community and a later decrease in infection of the younger bystander cohort.

Our study has several important limitations. First, beyond vaccine-based immunity, infections of unvaccinated individuals could also be affected by naturally acquired immunity. While we minimize this confounding factor by focusing on communities with low cumulative prior infections, future studies could more directly control for this effect—for example, by including seroprevalence data. Second, individual behavior and public policy guidelines, including enforcement and compliance with a lockdown imposed between 8 January and 7 February 2021 (refs. ^{15–17}), might correlate with vaccination rates and also affect the infection potential of the unvaccinated group. Third, the proportion between the measured positive test fraction and the actual infection rate could be different in each community and might also vary with time. Using the logarithmic derivatives $\log(P_i/P_j)$, our analysis accounts for the possibility of different proportions between the infection rate and the positive test fraction in each community as well as for the possibility that this proportion changes uniformly with time, but it cannot account for the possibility that these proportions change over time in different ways for different communities. Finally, our study population is limited to MHS members, who are representative of only part of the overall population for each community.

In this study, we identified a strong negative association between vaccination rate at the community level and the risk of infection for unvaccinated members of the community. We find that high vaccination rates were associated with lower infection rates at later time points among the unvaccinated cohort. Although the observed vaccine-associated protection of the unvaccinated population is encouraging, further studies are required to understand whether and how vaccination campaigns might support the prospect of herd immunity and disease eradication.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-021-01407-5>.

Received: 26 March 2021; Accepted: 26 May 2021;
Published online: 10 June 2021

References

- Polack, F. P. et al. Safety and efficacy of the BNT162b2 mRNA COVID-19 vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020).
- Chodick, G. et al. The effectiveness of the first dose of BNT162b2 vaccine in reducing SARS-CoV-2 infection 13–24 days after immunization: real-world evidence. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.01.27.21250612> (2021).
- Dagan, N. et al. BNT162b2 mRNA COVID-19 Vaccine in a nationwide mass vaccination setting. *N. Engl. J. Med.* **384**, 1412–1423 (2021).
- Yelin, I. et al. Associations of the BNT162b2 COVID-19 vaccine effectiveness with patient age and comorbidities. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.03.16.21253686> (2021).
- Rossman, H. et al. COVID-19 dynamics after a national immunization program in Israel. *Nat. Med.* <https://doi.org/10.1038/s41591-021-01337-2> (2021).
- Levine-Tiefenbrun, M. et al. Initial report of decreased SARS-CoV-2 viral load after inoculation with the BNT162b2 vaccine. *Nat. Med.* <https://doi.org/10.1038/s41591-021-01316-7> (2021).
- Aschwanden, C. Five reasons why COVID herd immunity is probably impossible. *Nature* **591**, 520–522 (2021).
- Kadkhoda, K. Herd immunity to COVID-19. *Am. J. Clin. Pathol.* **155**, 471–472 (2021).
- Buss, L. F. et al. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science* **371**, 288–292 (2021).
- Fontanet, A. & Cauchemez, S. COVID-19 herd immunity: where are we? *Nat. Rev. Immunol.* **20**, 583–584 (2020).
- Delamater, P. L., Street, E. J., Leslie, T. F., Yang, Y. T. & Jacobsen, K. H. Complexity of the basic reproduction number (R_0). *Emerg. Infect. Dis.* **25**, 1–4 (2019).
- Hibino, S., Hayashida, K., Ahn, A. C. & Hayashida, Y. COVID-19 seropositivity changes in asymptomatic individuals during the second and third waves of COVID-19 in Tokyo. Preprint at *medRxiv* <https://doi.org/10.1101/2020.09.21.20198796> (2021).
- Harvey, R. A. et al. Association of SARS-CoV-2 seropositive antibody test with risk of future infection. *JAMA Intern. Med.* **181**, 672–679 (2021).
- Béraud, G. et al. The French connection: the first large population-based contact survey in France relevant for the spread of infectious diseases. *PLoS ONE* **10**, e0133203 (2015).
- Ministry of Health Coronavirus Cabinet approves the Ministry of Health's outline for the gradual reopening of the education system (9 February 2021); <https://www.gov.il/en/departments/news/09022021-01>
- Ministry of Health The lockdown in its current format has been extended until Sunday, 7.2.2021 at 07:00 AM (5 February 2021); <https://www.gov.il/en/departments/news/05022021-01>
- Ministry of Health The Government approved the Ministry of Health new restrictions in addition to the ones in force (7 January 2021); <https://www.gov.il/en/departments/news/07012021-01>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

Data collection. Anonymized electronic health records were retrieved from MHS database for the study period: 6 December 2020–9 March 2021. These records include the following information. (1) Patient demographics, which included for each MHS member: a random ID used to link records, year of birth, and a coded geographical location of residence. For anonymization purposes, location of residence was given as random codes for the city or town and geographical statistical area (GSA; the country is divided by the National Bureau of Statistics into GSAs, which correspond to areas or neighborhoods). (2) The test results, which included for any SARS-CoV-2 reverse-transcription quantitative PCR test performed for MHS members: the patient random ID, the sample date and an indication of positive and negative results (total, 2,137,063 tests with 110,788 positive results). (3) Vaccination, which included the patient by random ID and date of inoculation with the first dose of the BNT162b2 mRNA COVID-19 vaccine (1.37 million vaccinated individuals).

Ethics committee approval. The study protocol was approved by the ethics committee of The MHS. The IRB-approved protocol includes an exempt from informed consent.

Vaccination and testing intervals. We defined two consecutive 3-week testing intervals between 6 January 2021 and 16 February 2021 (3-week-long intervals were long enough to allow for a sufficient number of tests; see ‘Defining communities’). For each of these two testing intervals and each community, the positive test fraction among unvaccinated individuals below 16 years old was calculated (Fig. 1b; P_1 , P_2), and their standard errors were evaluated assuming Poisson statistics. MHS members were excluded from this calculation if in total (at any time) they were tested more than 20 times (to avoid patients participating in extensive surveys; a sensitivity analysis is included in Extended Data Fig. 4). In addition, for each MHS member, any test following a positive test was excluded from this calculation. To account for the expected time after the first dose for a vaccinated individual to acquire immunity and to potentially prevent further infections, each 3-week testing interval was matched with a corresponding 3-week vaccination interval that was time-shifted backwards by 28 d. For each vaccination interval and each community, the mean cumulative fraction of individuals aged 16–50 years vaccinated with the first vaccine dose (out of all MHS members aged 16–50 years in the community) was calculated (Fig. 1a; V_1 , V_2).

Defining communities. Communities were defined as groups of members residing in the same GSA geographical code. For each city code, all GSAs with fewer than 2,000 members were joined into a single community. Communities were excluded if (1) they contained fewer than 2,000 members; (2) the number of tests for any 7-d window was smaller than 2 (too noisy to analyze the positive test fraction over time, Fig. 1b); (3) the number of tests in any 3-week time window was smaller than 40 (too noisy for correlation analysis, see below); (4) the number of positive tests in any 3-week time window was zero (which would cause a division by zero or a logarithm of zero when calculating y values for the correlation analysis); (5) they contained fewer than 20 members aged 16–50 years (too noisy for correlation analysis); or (6) by 9 March 2021, more than 10% of the community members tested positive (too high a presumed natural immunization confounder; a sensitivity analysis is included in Extended Data Fig. 3). This analysis resulted in 177 communities (Extended Data Fig. 7). Members not residing in one of these eligible communities were excluded from our analysis.

Statistical methods. Correlation analysis. A linear fit was performed on the log-transformed ratio of the positive test fraction $y = \log(P_2/P_1)$ and the change in vaccination $x = V_2 - V_1$ using the built-in function `glmfit` in MATLAB. The P value of the slope was calculated as the fraction of smaller or equal slopes calculated in 10,000 xy permutation bootstrapping simulations.

Sensitivity analysis. For each main parameter, a correlation analysis was performed with different values of this parameter, while all other parameters were kept fixed. The fixed parameters were those used to produce Fig. 1c. Each of these runs produced the slope and the P value of the slope (Fig. 1c; see ‘Correlation analysis’). The specific choices of parameters that were tested and their values are indicated in Extended Data Figs. 3–6.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

According to the regulations of the Israel Ministry of Health, individual-level data cannot be shared openly. Specific requests for remote access to de-identified community-level data should be referred to Maccabitech, Maccabi Healthcare Services Institute for Research & Innovation.

Code availability

Code used for data analysis is freely available upon request.

Acknowledgements

This work was supported by the Israel Science Foundation (grant no. 3633/19 to R. Kishony and G.C.) as part of the KillCorona-Curbing Coronavirus Research Program.

Author contributions

Study design: O.M., I.Y., T.P., S.G., G.C. and R. Kishony. Data collection: O.M., I.Y., E.H. and R. Katz. Data analysis: O.M. and R. Kishony. Data interpretation: O.M., I.Y., N.A., J.K., A.B.-T., T.P., S.G., G.C. and R. Kishony. Writing: O.M., I.Y., N.A. and R. Kishony, with comments from all authors.

Competing interests

The authors declare no competing interests.

Additional information

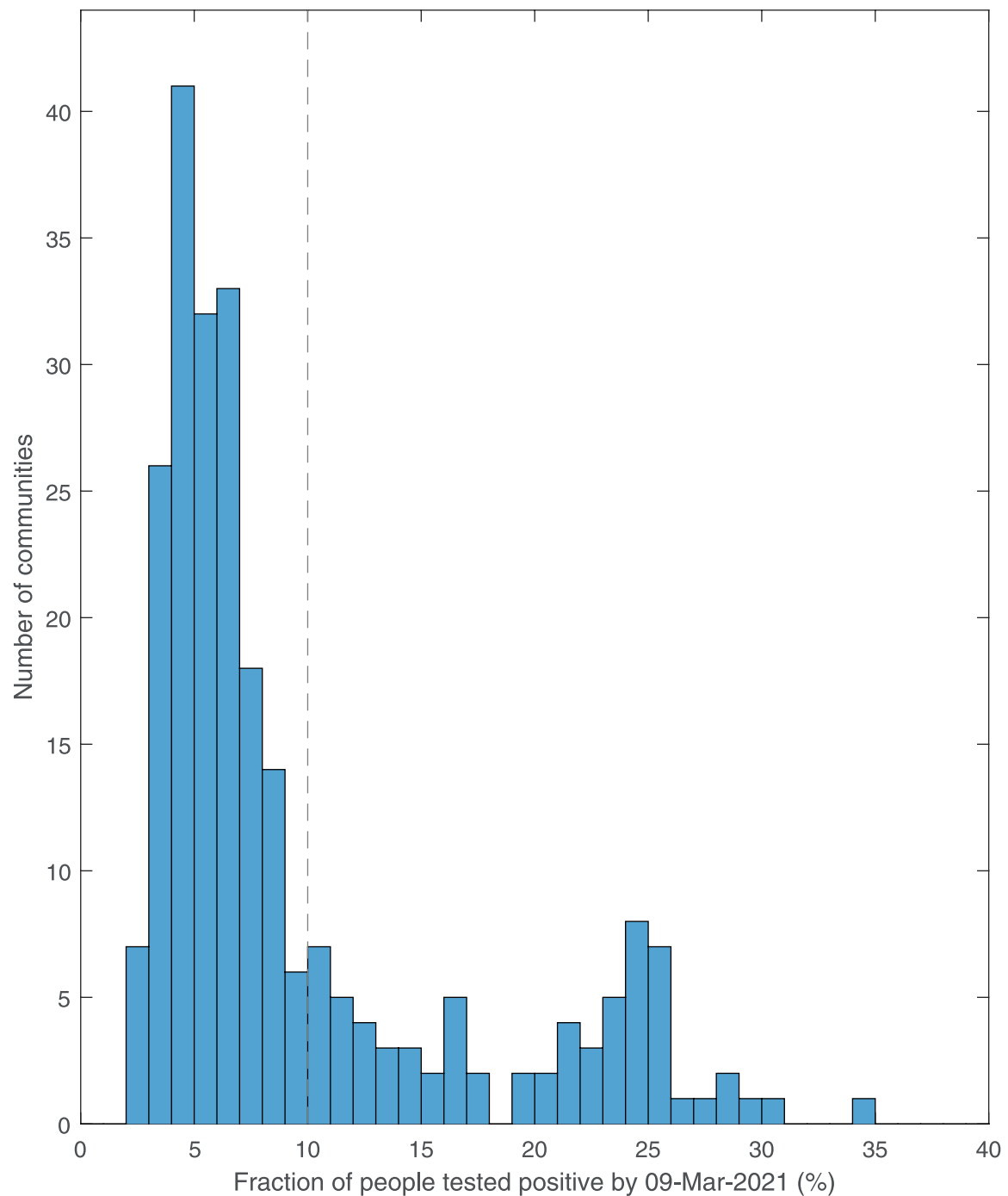
Extended data is available for this paper at <https://doi.org/10.1038/s41591-021-01407-5>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41591-021-01407-5>.

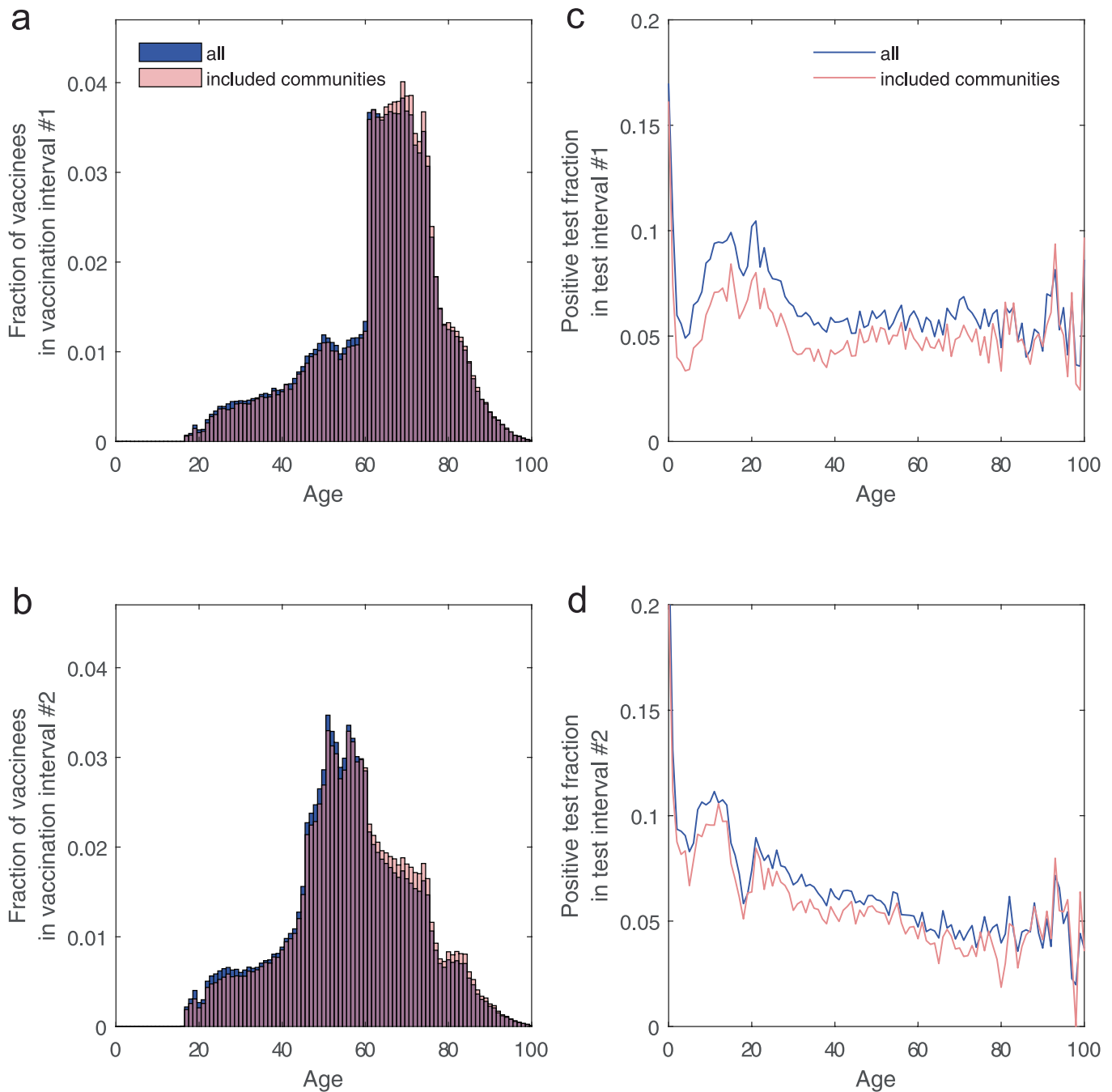
Correspondence and requests for materials should be addressed to T.P. or R.K.

Peer review information *Nature Medicine* thanks Benjamin Althouse, Ben Lopman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Jennifer Sargent was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

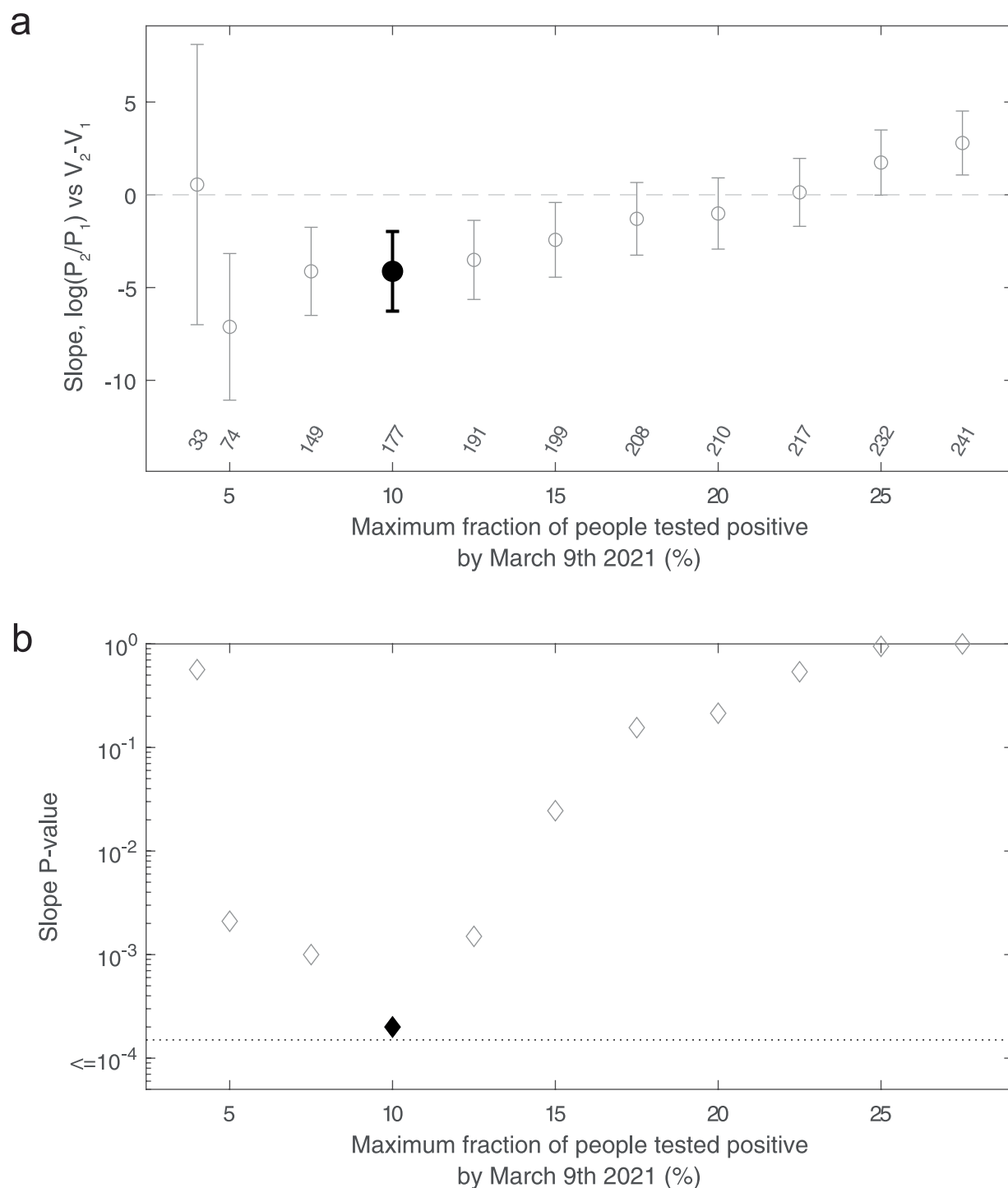
Reprints and permissions information is available at www.nature.com/reprints.



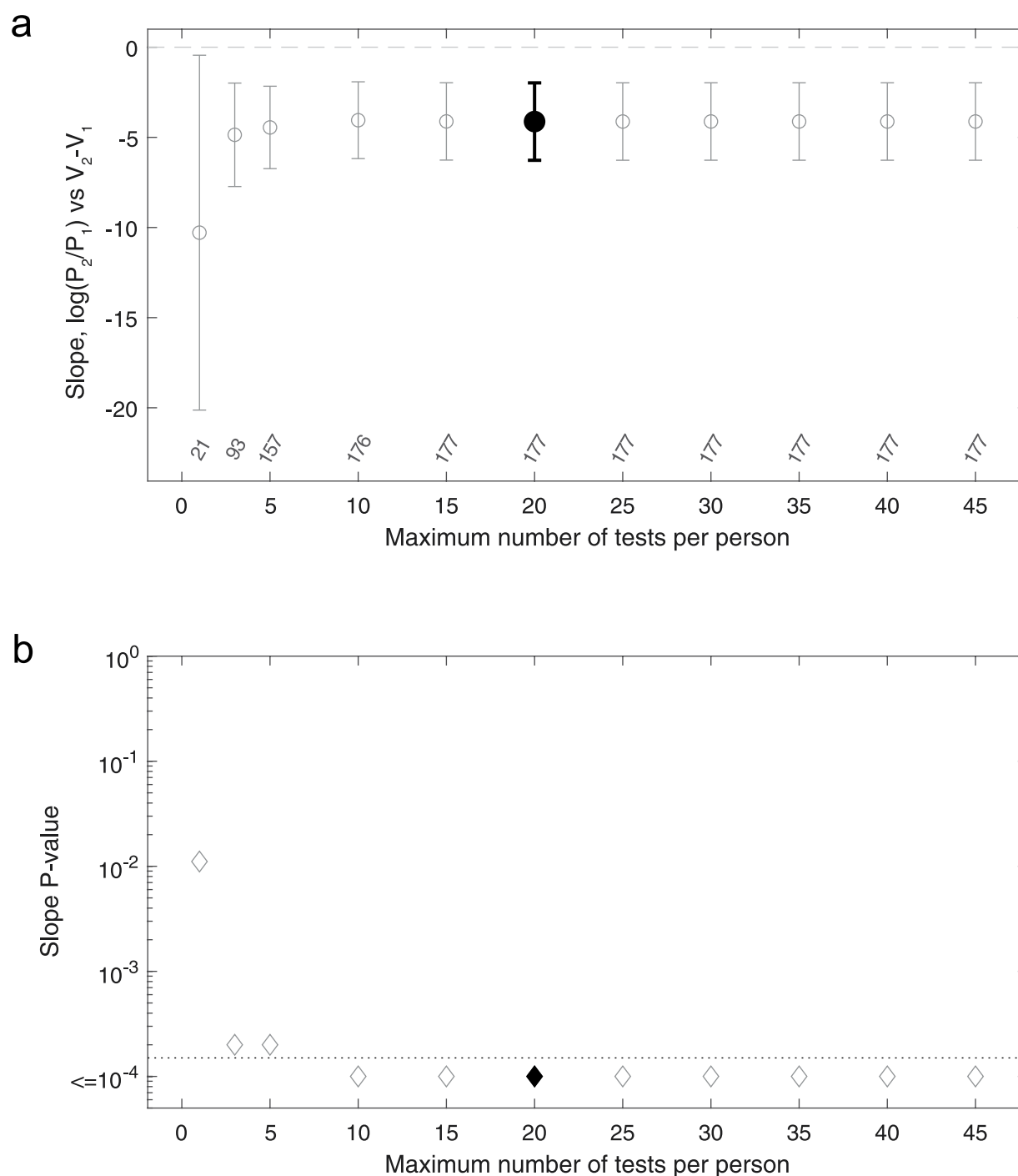
Extended Data Fig. 1 | Discarding communities with a large fraction of confirmed infections. A histogram of the cumulative fraction of people tested positive by March 9th 2021 in each community. To remove possible confounding effects of natural immunity, communities with large infection fractions were excluded (above 10%, dashed vertical line).



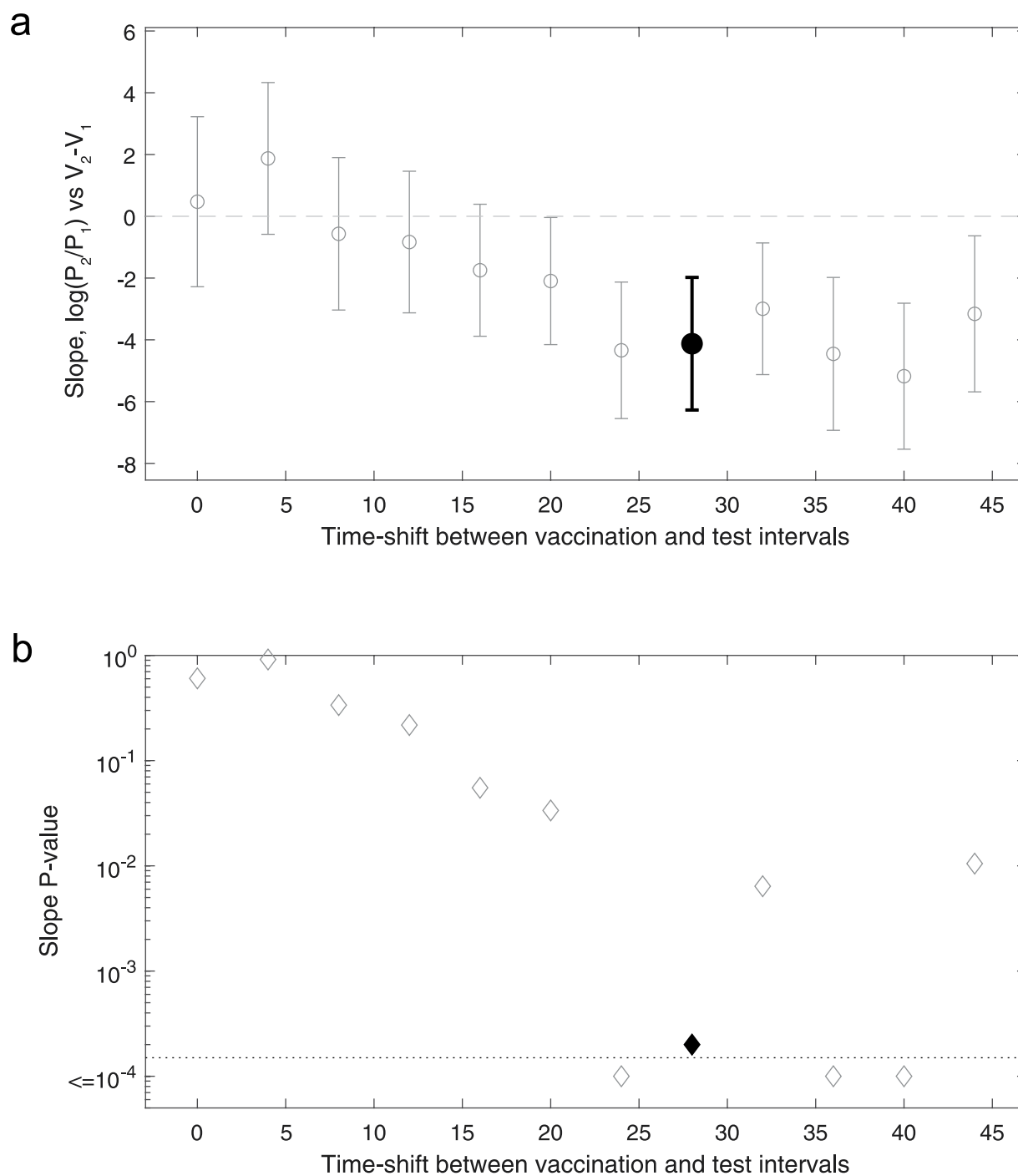
Extended Data Fig. 2 | Effect of community-exclusion on the age distribution of vaccination and positive tests. a,b, Histograms showing the age distribution of vaccinees who received the first vaccine dose in the first (**a**) and second (**b**) vaccination intervals for all MHS members (blue) and for the subset of communities included in the correlation analysis (pink). **c,d,** Positive test fraction as a function of age in the first (**c**) and second (**d**) testing intervals for all tests (blue) and for the subset of communities included in the correlation analysis (pink).



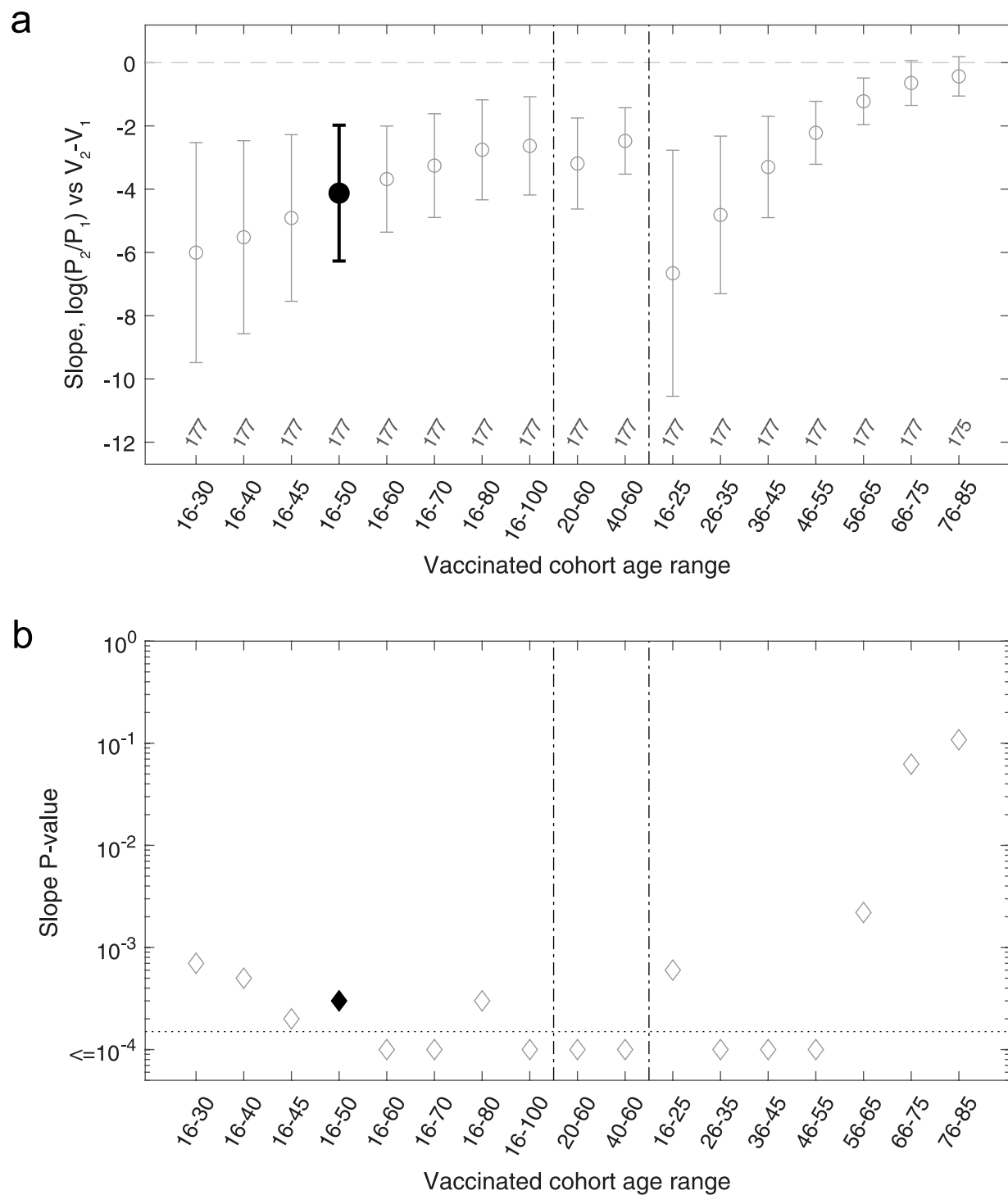
Extended Data Fig. 3 | Effect of community exclusion threshold on the correlation analysis results. Slopes (**a**) and P-values (**b**) of the correlation analysis (similar to Fig. 1c, see Methods) for different values of the threshold of the fraction of people tested positive in a community (Methods, defining communities and sensitivity analysis). The results for the chosen threshold value of 10% (used in Fig. 1) are indicated (filled black symbols). Circles and error bars represent $\text{MLE} \pm 2\text{SE}$ (95% confidence intervals). The number of communities for each threshold value is indicated at the bottom of panel **a**.



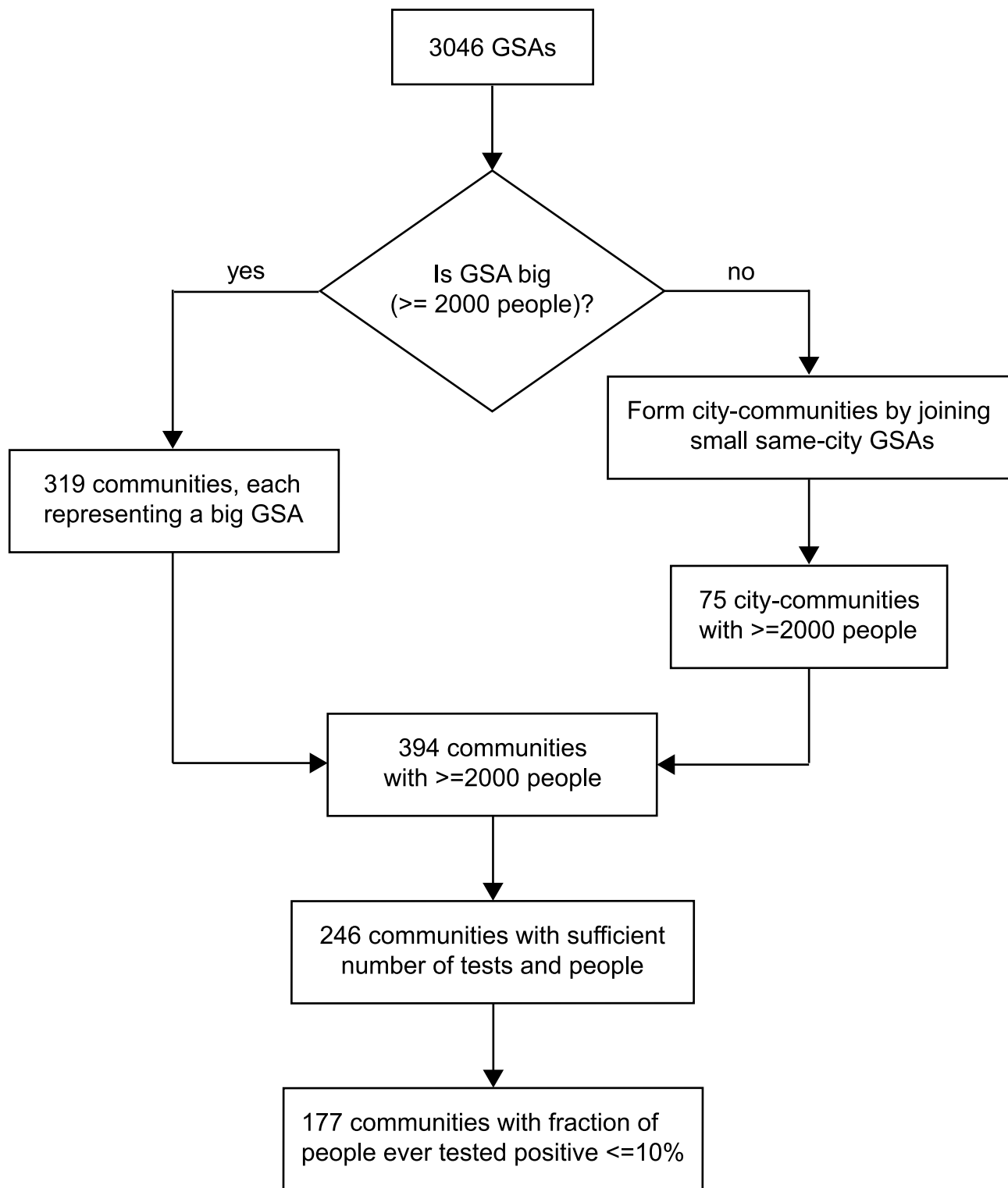
Extended Data Fig. 4 | Effect of threshold on number of tests per individual. Slopes (**a**) and P-values (**b**) of the correlation analysis (similar to Fig. 1c, see Methods) were calculated for different values of the threshold for excluding people with atypical number of tests (Methods, Vaccination and testing intervals and sensitivity analysis). The results for the chosen threshold value of 20 (used in Fig. 1) are indicated (filled black symbols). Circles and error bars represent $\text{MLE} \pm 2\text{SE}$ (95% confidence intervals). The number of communities for each threshold value is indicated at the bottom of panel **a**.



Extended Data Fig. 5 | Effect of time-shift between vaccination and test intervals. Slopes (**a**) and P-values (**b**) of the correlation analysis (similar to Fig. 1c, $n=177$, see Methods) are shown for different time-shifts (delay) between vaccination and test intervals (Methods, sensitivity analysis). The results for the chosen time-shift of 28 days (used in Fig. 1) are indicated (filled black symbols). Circles and error bars represent $\text{MLE} \pm 2\text{SE}$ (95% confidence intervals).



Extended Data Fig. 6 | The association of the positive test fraction of the young cohort with vaccination of different age groups. Slopes (**a**) and P-values (**b**) of the correlation analysis (similar to Fig. 1c, see Methods) are shown for different vaccinated cohort age groups (Methods, *sensitivity analysis*). The results for the age group of 16–50 years old (used in Fig. 1) are indicated (filled black symbols). Circles and error bars represent $\text{MLE} \pm 2\text{SE}$ (95% confidence intervals). The number of communities for each age group is indicated at the bottom of panel **a**.



Extended Data Fig. 7 | Definition and exclusion of communities. 319 big GSAs (at least 2000 members) were each considered a community. For each city, all small GSAs (less than 2000 members) were joined, forming city-communities, of which only 75 had ≥ 2000 members. Next, communities were filtered to ensure only communities with a sufficient number of tests and potential vaccinees were included in the analysis (Methods). This left 246 communities, of which 177 had less than 10% of their members test positive by March 9th 2021. These 177 communities were considered in the nominal correlation analysis (Fig. 1).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection No software was used to collect data.

Data analysis MATLAB functions: pdist, linkage and glmfit.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

According to the Israel Ministry of Health regulations, individual-level data cannot be shared openly. Specific requests for remote access to de-identified community-level data should be referred to Maccabitech, Maccabi Healthcare Services Institute for Research & Innovation.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was determined by availability of electronic health records. Exclusion is detailed in the methods section.
Data exclusions	Patients with more than 20 tests were excluded from the study. Additionally, communities were excluded according to criteria detailed in the methods section under "Defining communities".
Replication	Non applicable.
Randomization	Non applicable for a Big-data observational study.
Blinding	Non applicable for a Big-data observational study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms		
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern		

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	MHS members were included in two separate age cohorts: a test cohort of MHS members 0-15 years old and a vaccination cohort of MHS members 16-50 years old.
Recruitment	This is an observational study involving electronic health records analysis.
Ethics oversight	Ethics committee of Maccabi Healthcare Services.

Note that full information on the approval of the study protocol must also be provided in the manuscript.