nature portfolio

Corresponding author(s): Nikhil Garg

Last updated by author(s): Nov 12, 2023

Reporting Summary

Nature Portfolio 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 Portfolio 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	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	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)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	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	We did not use any software to collect data: public data can be directly downloaded from their portals listed in our public repository.
Data analysis	The main data analysis softwares we used are Python 3, CmdStan 2.27, and Scikit-learn 1.1.3. A complete list of software packages is provided in the public repository accessible through the following links. 1. https://github.com/nikhgarg/spatial_underreporting_crowdsourcing, 2. https://github.com/ZhiLiu724/reporting_rate_estimation

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Public 311 data for both NYC and Chicago is available via their Open Data portals. In the GitHub repository provided in the manuscript, we have included official links to the portals and further host the exact public data used in this work. The primary NYC results in Section 4 are run on private data confidentially provided to us by

NYC DPR that cannot be shared -- that data additionally contains columns for anonymized reporter identification, to allow filtering duplicate reports by the same user. However, all results replicate with the available public data, as reported in the Appendix. The Chicago analyses are fully based on publicly available data.

The links to the public data sets used are as follows.
1. 311 service requests made to the Forestry unit of NYC Parks. This dataset is actively maintained, and the version used in the main text was retrieved 10/22/2022 at 14:45 EDT from https://data.cityofnewyork.us/Environment/Forestry-Service-Requests/mu46-p9is
2. Inspections conducted by the Forestry unit of NYC Parks. This dataset is actively maintained, and the version used in the main text was retrieved 10/22/2022 at 14:32 EDT from https://data.cityofnewyork.us/Environment/Forestry-Inspections/4pt5-3vv4
3. Work orders completed by the Forestry unit of NYC Parks. This dataset is actively maintained, and the version used in the main text was retrieved 10/22/2022 at 14:31 EDT from https://data.cityofnewyork.us/Environment/Forestry-Work-Orders/bdjm-n7q4
4. Risk assessments by the Forestry unit of NYC Parks. This dataset is actively maintained, and the version used in the main text was retrieved 10/24/2022 at 14:30 EDT from https://data.cityofnewyork.us/Environment/Forestry-Risk-Assessments/259a-b6s7
5. 311 service requests data from City of Chicago. This dataset is actively maintained, and the version used in the main text was retrieved 7/4/2022 at 21:20 EDT from https://data.cityofchicago.org/Service-Requests/311-Service-Requests/v6vf-nfxy

Reproduction code can be accessed using the following link:
https://github.com/nikhgarg/spatial_underreporting_crowdsourcing

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

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

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

Behavioural & social sciences study design

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

Study description	We study observational, quantitative data on 311 service requests submitted by residents to NYC and Chicago governments.
Research sample	The samples are all NYC and Chicago residents who filed relevant service requests using the 311 system in the time period indicated below. The data are preexisting, as described in the Data section of this summary, where links to the original data are provided.
Sampling strategy	We collect all the data available on the open data portals of NYC and Chicago, at the time of data retrieval. We do not sample the raw data.
Data collection	We do not curate the data ourselves, but only collect them after they are curated and made public by municipal governments. 311 service requests and the subsequent responses are automatically logged by municipal governments: 311 service requests are formatted and automatically collected in internal databases of municipal governments; subsequent government agency responses are logged by government employees when those responses are carried out.
Timing	NYC: 02/28/2015 - 09/01/2022; Chicago: 03/01/2019 - 07/04/2022
Data exclusions	We exclude a small portion of the service requests: for the New York City data set used to conduct analyses in the main text, out of 98,994 incidents identified, we exclude 17,356 incidents; for the Chicago data set, out of 698,365 incidents identified, we exclude 132,935 incidents. The exclusions are mainly due to missing attributes which are essential to our analysis.
Non-participation	N/A, no participants were involved.
Randomization	N/A

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	
n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☒ ☐	Animals and other organisms
☒ ☐	Clinical data
☒ ☐	Dual use research of concern

Methods	
n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging