nature portfolio

Corresponding author(s): Neil Sehgal

Last updated by author(s): Feb 26, 2026

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	All data used in this study are publicly available from Reddit.com, or are accessible via Pushshift and Arctic Shift.
Data analysis	All data was analyzed using Python version 3.1

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

All data used in this study are publicly available from Reddit.com, or are accessible via Pushshift and Arctic Shift.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	NA
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	As data were public and non-identifiable, IRB review was not required.

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	Quantitative, observational cross-sectional study using publicly available social media text. We analyzed Reddit posts and comments to identify self-disclosed use of semaglutide or tirzepatide and to extract self-reported side effects mapped to MedDRA Preferred Terms. Analyses are descriptive (prevalence of reported PTs among users who disclosed ≥1 side effect; plus descriptive comparisons by medication formulation among users who mentioned only one formulation).
Research sample	Public Reddit posts/comments from nine subreddits focused on GLP-1 medications and weight/weight loss communities: r/mounjaro, r/ozempic, r/fasting, r/intermittentfasting, r/keto, r/loseit, r/Semaglutide, r/SuperMorbidlyObese, r/PlusSize. The analytic population consists of unique Reddit users who self-disclosed taking semaglutide or tirzepatide. No individual-level demographics (age/sex/gender/race/location) were reliably available and were not used for inclusion/stratification. The sample is not representative of medication users in the population; it reflects self-selected Reddit participants.
Sampling strategy	Convenience sample of posts/comments drawn from the specified subreddits within the study window. Posts/comments were retrieved using keyword matching on brand/generic names and common misspellings, then filtered using an LLM-based classifier to identify self-disclosed medication use; among self-use posts, a second LLM-based classifier extracted side effects and mapped them to MedDRA PTs. All eligible posts within the collection window were included.
Data collection	Reddit posts/comments accessed via Pushshift and Arctic Shift for the collection period. Data consist only of public text content and associated timestamps/metadata available in the dataset.
Timing	Posts/comments collected for January 1, 2015 through June 30, 2025; in practice, all posts meeting inclusion criteria were dated May 2019 through June 2025.
Data exclusions	Exclusions were applied at the post/comment level based on content and eligibility: 1. Posts/comments that mentioned the medications but did not contain a self-disclosure of personal use were excluded from downstream side-effect analyses. 2. Users were included in side-effect prevalence calculations only if they self-disclosed at least one side effect. 3. For the medication-comparison sub-analysis, users who mentioned both semaglutide and tirzepatide were excluded (restricted to users exclusively mentioning one formulation). Across the May 2019–June 2025 period during which eligible posts appeared, 172,679 of 410,198 Reddit posts mentioning semaglutide or tirzepatide indicated personal use by 67,008 users. Among them, 43.5% (29,172 users) reported at least one side effect. In exploratory descriptive analyses, we examined side effect frequencies separately for users who exclusively mentioned semaglutide (n=17,937) versus tirzepatide (n=7,125).

Non-participation	NA
Randomization	NA

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
☒	☐ Plants

Methods

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