

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 (<i>n</i>) 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 <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☐	☒ 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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ 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 <i>d</i> , Pearson's <i>r</i>), 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	The custom computer code for data collection is publicly available on GitHub (https://github.com/v1xerunt/DocTr) and archived on Zenodo with the identifier doi: 10.5281/zenodo.18207089
Data analysis	The custom computer code for data analysis is publicly available on GitHub (https://github.com/v1xerunt/DocTr) and archived on Zenodo with the identifier doi: 10.5281/zenodo.18207089

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](#)

The OpenPayments data used in this study are publicly available from the Centers for Medicare & Medicaid Services (CMS) at <https://openpaymentsdata.cms.gov/>.

Clinical trial data analyzed in this study are publicly available from ClinicalTrials.gov. Census data are available from the United States Census Bureau (<https://www.census.gov/>). Processed data generated during this study are available in the associated code repository.

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	This study does not involve human participants.
Reporting on race, ethnicity, or other socially relevant groupings	This study does not involve human participants.
Population characteristics	This study does not involve human participants.
Recruitment	This study does not involve human participants.
Ethics oversight	All patient data within this study is statistically de-identified and certified as compliant with the Health Insurance Portability and Accountability Act (HIPAA). As this study utilized fully anonymized, pre-existing data, it was considered non-human subjects research, and thus a waiver of consent and specific institutional review board (REB) approval for this secondary analysis were 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

Life sciences study design

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

Sample size	The study utilized a final integrated dataset encompassing 5,210 clinical trials and 24,984 clinicians across the United States. The sample size was not predetermined by statistical calculation but was determined by the availability of high-quality linked data integrating four primary sources: ClinicalTrials.gov, IQVIA medical claims, OpenPayments, and US Census data.
Data exclusions	The study is limited to US clinicians due to data constraints; clinicians outside the US were excluded.
Replication	To ensure robustness and reproducibility, the model was evaluated under two different experimental settings: a random split (8:1:1) and a temporal split (training on trials from 1989-2019, testing on trials from 2020-2026). The temporal split specifically tests generalization to unseen data. The code for implementing the model and processing the data has been made publicly available on GitHub to facilitate replication.
Randomization	As this is a retrospective computational analysis of existing datasets, participants (clinicians) were not allocated into experimental groups. However, for model evaluation, the dataset was split into training, validation, and testing sets. This included a random 8:1:1 data split where unique clinician-trial relationships were exclusive to one set.
Blinding	Blinding was not applicable in the context of clinical intervention as this was a non-human subjects research study utilizing anonymized, pre-existing data.

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

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.