nature portfolio

Corresponding author(s): Li Yang,Jundong Jiao,Li Wang,Jicheng Lv,

Last updated by author(s): 2024/11/2

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 developed Java program to extract data from Oracle database (12c) 12.1.0.2.0 where dataset is stored.
Data analysis	All analyses were performed using Python version 3.9.12 (coded by scikit-learn 1.3.0) and SPSS version 24. Figures were generated by Python version 3.9.12 and R version 4.3.2. The code is available on Code Ocean (https://codeocean.com/capsule/9261428/tree/v1).

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 clinical data used for training and validation in this study are protected and are not publicly available due to data privacy laws. The source data for all figures generated in this study are provided in Source data files. Demo data are provided for reproducibility of code available on Code Ocean (https://codeocean.com/capsule/9261428/tree/v1)..

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	The findings apply to both gender; Gender was considered in study design, and gender was determined based on self reporting in medical records; No individual level data has not been collected; Gender based analysis were performed and reported in the manuscript.
Reporting on race, ethnicity, or other socially relevant groupings	No socially constructed or socially relevant categorization variables are used in the manuscript.
Population characteristics	The study involves all adult patients admitted to five hospitals during the study period. The mean age ranges from 59.1- 66.7 in five cohorts, with 35.5-43.7% being female. The majority of patients are admitted in Internal Medicine ward and surgical ward, with a small proportion being admitted in ICU.
Recruitment	All adult patients admitted to five hospitals during the study period were enrolled.
Ethics oversight	The study was approved by the Research Ethics Committee of Peking University First Hospital (Site 1, approval number 2023[069-001]), Sichuan Provincial People's Hospital (Site 2, approval number 2023[175]), the Second Affiliated Hospital of Harbin Medical University (Site 3, approval number 2023[069]), Beijing Miyun District Hospital (Site 4, approval number 2023[006-001]), and Taiyuan Central Hospital (Site 5, approval number 2023006).

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	This is a big data research and we included all adult patients admitted during the study period to ensure a sample size as large as possible.
Data exclusions	Exclusion criteria in our study are pre-established and described in detail in Methods section (Patients were excluded if they had less than 2 documented serum creatinine measurements during hospitalization, were diagnosed with end-stage renal disease, were maintained on dialysis or had an initial Scr greater than or equal to 4.0 mg/dL at admission, developed AKI prior to admission or within 24 hours after admission, had a length of stay shorter than 24 hours, underwent kidney transplantation or nephrectomy during hospitalization, or had all Scr measurements lower than or equal to 0.6 mg/dL from 90 days prior to admission until discharge.). The exclusion criteria exclude those who will never develop AKI and those who were not suitable to be diagnosed with AKI using the current international diagnostic criteria of AKI.
Replication	The findings of the study are replicable using code and original data.
Randomization	This is a big data research for model building and randomization is not relevant to our study.
Blinding	This is a retrospective big data study for model building and validation, there is no group allocation in the study, so blinding was not relevant to our 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

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, mosiacism, off-target gene editing) were examined.