nature portfolio

Corresponding author(s): Mei Liu

Last updated by author(s): May 25, 2025

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	A GitHub repository containing the code used for data collection is available at https://github.com/GatorAIM/AKI_Data_Extraction.git.
Data analysis	A GitHub repository containing the code used in the analysis is available at https://github.com/GatorAIM/AKI-1_Subphenotyping.

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 is not publicly available and restrictions apply to its use. The de-identified multi-center EHR datasets in PCORnet Common Data Model may be made available by the Greater Plains Collaborative and PaTH clinical research networks, subjective to individual institution's and network-wide data governance and ethical approvals.

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	Sex was determined based on biological attributes recorded in the electronic health record (EHR) at the time of patient admission. Gender information was not available in the dataset and was therefore not included in the analysis. Analyses were not stratified by sex due to the focus on SCr trajectory-based AKI-1 subphenotyping.
Reporting on race, ethnicity, or other socially relevant groupings	Race and ethnicity were extracted from EHR data based on self-report at the time of hospital admission. These variables were not used as primary stratification variables. Race and ethnicity were not used as proxies for socioeconomic status.
Population characteristics	The study included 53,565 hospitalized adult patients diagnosed with AKI-1, each matched 1:1 to a non-AKI control based on key demographic and clinical variables. Patients were drawn from eight academic medical centers between 2009 and 2022. Most patients were aged ≥60 years. In AKI-1 group, 55.57% patients were male.
Recruitment	Participants were not prospectively recruited. Data were obtained retrospectively from de-identified EHRs across eight academic medical centers.
Ethics oversight	This study was determined by the institutional review boards of the University of Florida, University of Pittsburgh Medical Center, and University of Missouri as non-human subject research because it only involved the collection of existing and de-identified patient medical data. Data use agreements have been executed with both the Greater Plains Collaborative (GPC) and the University of Pittsburgh.

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 final study cohort consisted of 53,565 unique Stage-1 Acute Kidney Injury (AKI-1) patients, along with an equal number of matched patients without any AKI, resulting in a total of 107,130 unique patients.
Data exclusions	The detailed eligibility criteria for patient selection were as follows: (1) Exclusion of patients with pre-existing end-stage renal disease (ESRD); (2) Exclusion of patients with a history of dialysis or RRT; (3) Exclusion of patients whose estimated glomerular filtration rate (eGFR) < 15 mL/min/1.73m2 or baseline SCr > 3.5 mg/dL; (4) A minimum hospital stay of two days; (5) The encounter must involve an AKI-1 as the only AKI stage (i.e. no AKI stage progression during hospitalization); (6) At least one SCr measurement on each of the three days: two days before AKI-1 onset, one day before AKI-1 onset, and the day of AKI-1 onset; (7) No more than two missing SCr measurements within the seven-day observation window. (8) Only the earliest encounter of each patient was kept, making sure each patient enrolled in the study was unique.
Replication	A GitHub repository containing the code used in the analysis is available at https://github.com/GatorAIM/AKI-1_Subphenotyping. However, the clinical data used for training and validation in this study is not publicly available and restrictions apply to its use.
Randomization	The detailed eligibility criteria for patient selection were as follows: (1) Exclusion of patients with pre-existing end-stage renal disease (ESRD); (2) Exclusion of patients with a history of dialysis or RRT; (3) Exclusion of patients whose estimated glomerular filtration rate (eGFR) < 15 mL/min/1.73m2 or baseline SCr > 3.5 mg/dL; (4) A minimum hospital stay of two days; (5) The encounter must involve an AKI-1 as the only AKI stage (i.e. no AKI stage progression during hospitalization); (6) At least one SCr measurement on each of the three days: two days before AKI-1 onset, one day before AKI-1 onset, and the day of AKI-1 onset; (7) No more than two missing SCr measurements within the seven-day observation window. (8) Only the earliest encounter of each patient was kept, making sure each patient enrolled in the study was unique. To minimize confounding variables, we devised a three-step matching framework to pair each AKI-1 patient with a non-AKI counterpart. The matching process was based on demographics, baseline SCr, and comorbidities.
Blinding	This was a retrospective observational study using de-identified electronic health record data. As such, investigators were not blinded to group allocation during data analysis. However, all data-driven subphenotype classifications were derived using unsupervised clustering algorithms prior to any outcome analyses, thereby minimizing bias. Blinding was not applicable during data collection due to the nature of the retrospective design.

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	This was a retrospective observational study using de-identified electronic health record data.
Study protocol	This was a retrospective observational study using de-identified electronic health record data.
Data collection	This was a retrospective observational study using de-identified electronic health record data.
Outcomes	This was a retrospective observational study using de-identified electronic health record data.

Plants

Seed stocks	This was a retrospective observational study using de-identified electronic health record data.
Novel plant genotypes	This was a retrospective observational study using de-identified electronic health record data.
Authentication	This was a retrospective observational study using de-identified electronic health record data.