INFORMATICS PROFESSIONALS. LEADING THE WAY.

Associations with spontaneous and indicated preterm birth in a densely phenotyped EHR cohort

Uncovering Associations and Subpopulations with EHR Data

S08

Jean Costello & Hannah Takasuka

Jacquelyn Roger, Ophelia Yin, Alice Tang, Tomiko Oskotsky, Marina Sirota*, John A Capra*

University of California, San Francisco

jean.costello@ucsf.edu | hannah.takasuka@ucsf.edu

#IS24

Preterm Birth (PTB)

Preterm birth is a live birth occurring <37 weeks of gestation

- Affects appr. 10% of pregnancies worldwide
- Leading cause of neonatal mortality and morbidity worldwide
- Preterm birth has a negative impact on maternal well-being

image from Perinatal Services BC

Preterm Birth Subtypes

- **Indicated (iatrogenic)**: planned preterm delivery due to maternal or fetal causes

- **Spontaneous**: unplanned preterm delivery

Many studies combine spontaneous and indicated preterm birth due to a lack of data.

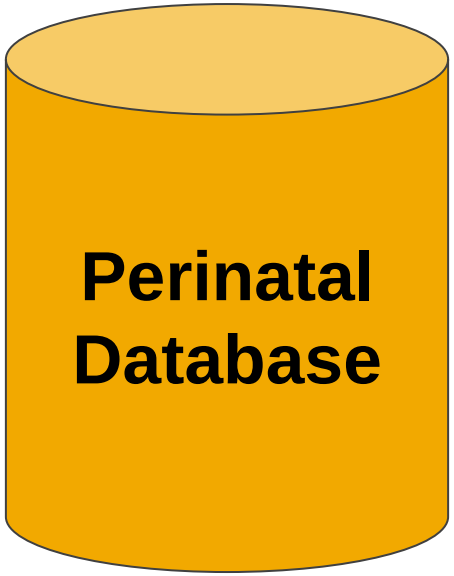
Motivation

- What biological processes cause preterm birth?
- Are the causes of spontaneous and indicated births similar?
- Can we predict pregnancies at risk of spontaneous preterm birth?

Our approach: conduct a hypothesis-generation study that investigates associations between health conditions and preterm birth by subtype

preprint: https://www.medrxiv.org/content/10.1101/2023.11.29.23299216v1

Data sources



- Accurate outcome classification
- PTB subtype
- n=47,375

- Conditions before pregnancy
- n=10,644

Study outline

Identify cohort with diagnoses prior to conception

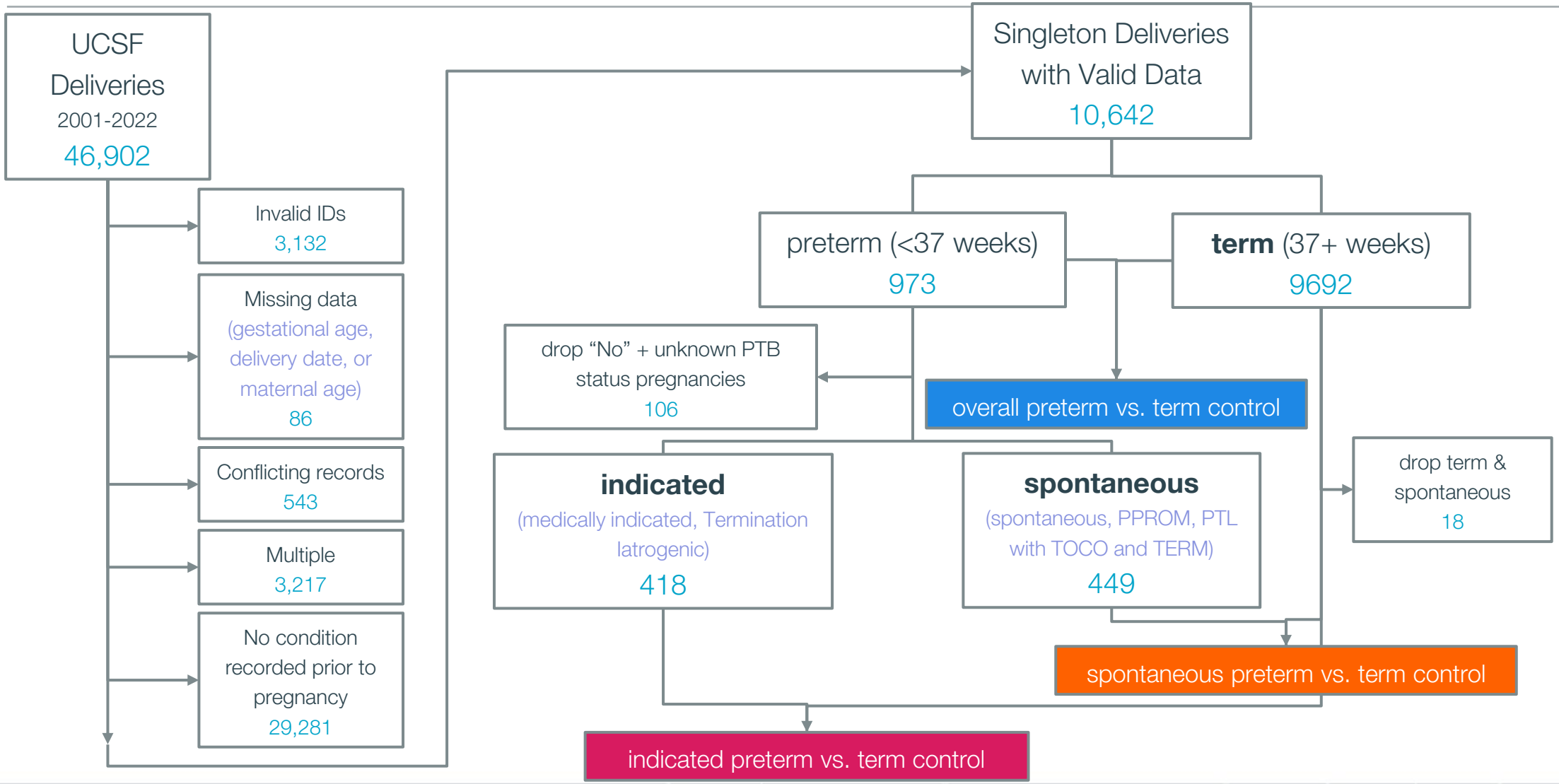
Group diagnoses into phecodes

Test for association with PTB overall and by subtype

Evaluate robustness of associations

Interpret findings: confirm known results and generate new hypotheses

Cohort construction



Testing conditions for significance

Logistic regression

- preterm ~ phecode_m + covariates

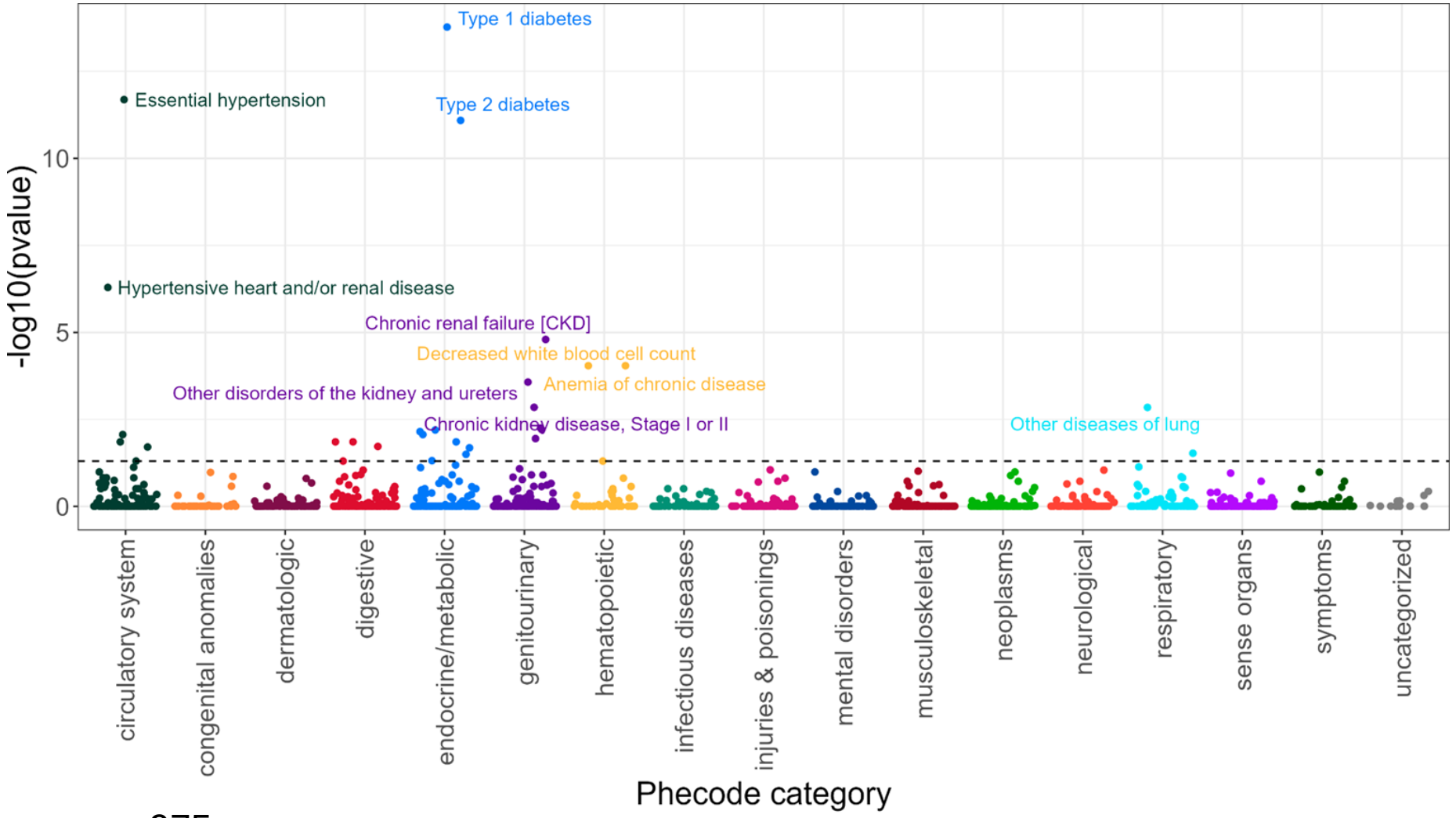
We chose our covariates *a priori* based on established associations for preterm birth:

- Maternal age (allowing for non-linearity)
- Maternal education (<12 years, 12 years, and >12 years)
- Insurance (private vs public)
- Race and ethnicity (9 categories)

p-values adjusted for multiple hypothesis testing (Benjamini-Hochberg)

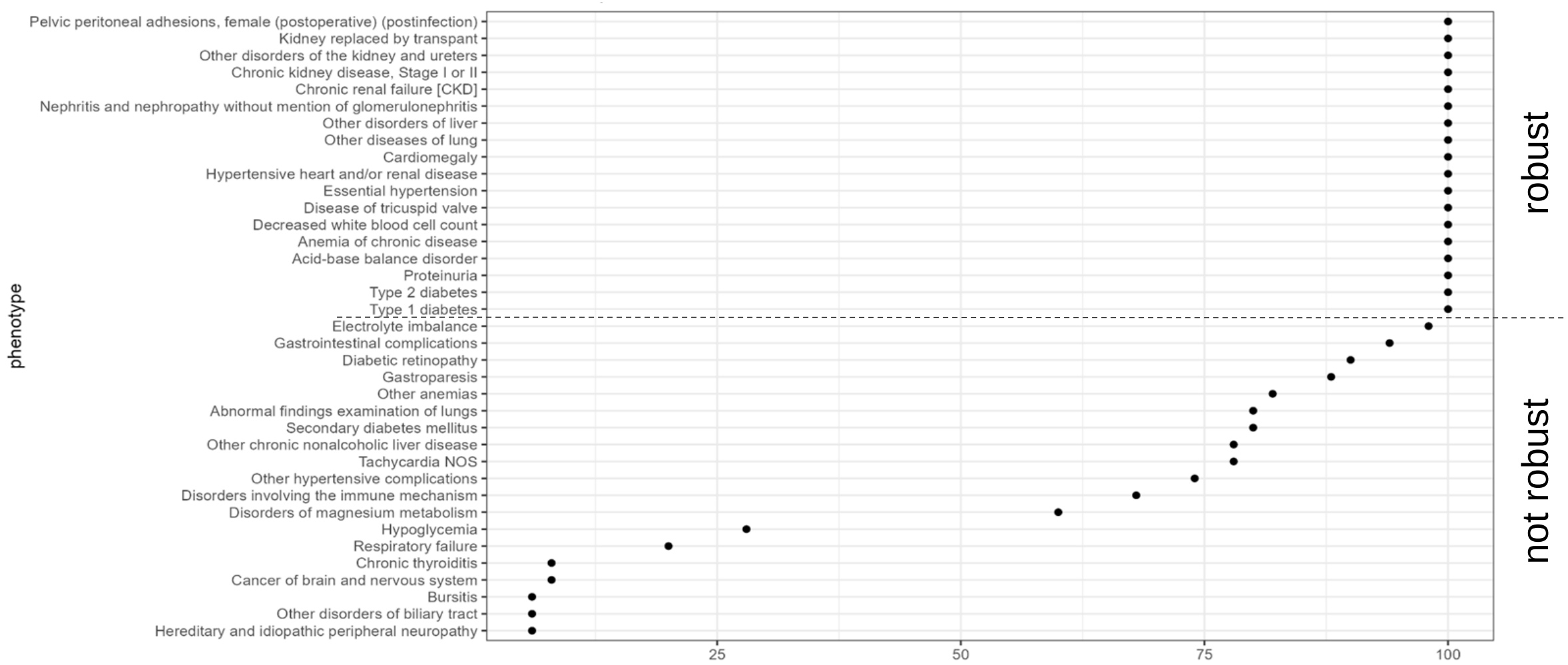
Bootstrap test to ensure robustness

37 conditions significantly associated with PTB



n_{term} = 9692 / $n_{preterm}$ = 975

18 conditions are significant and robust

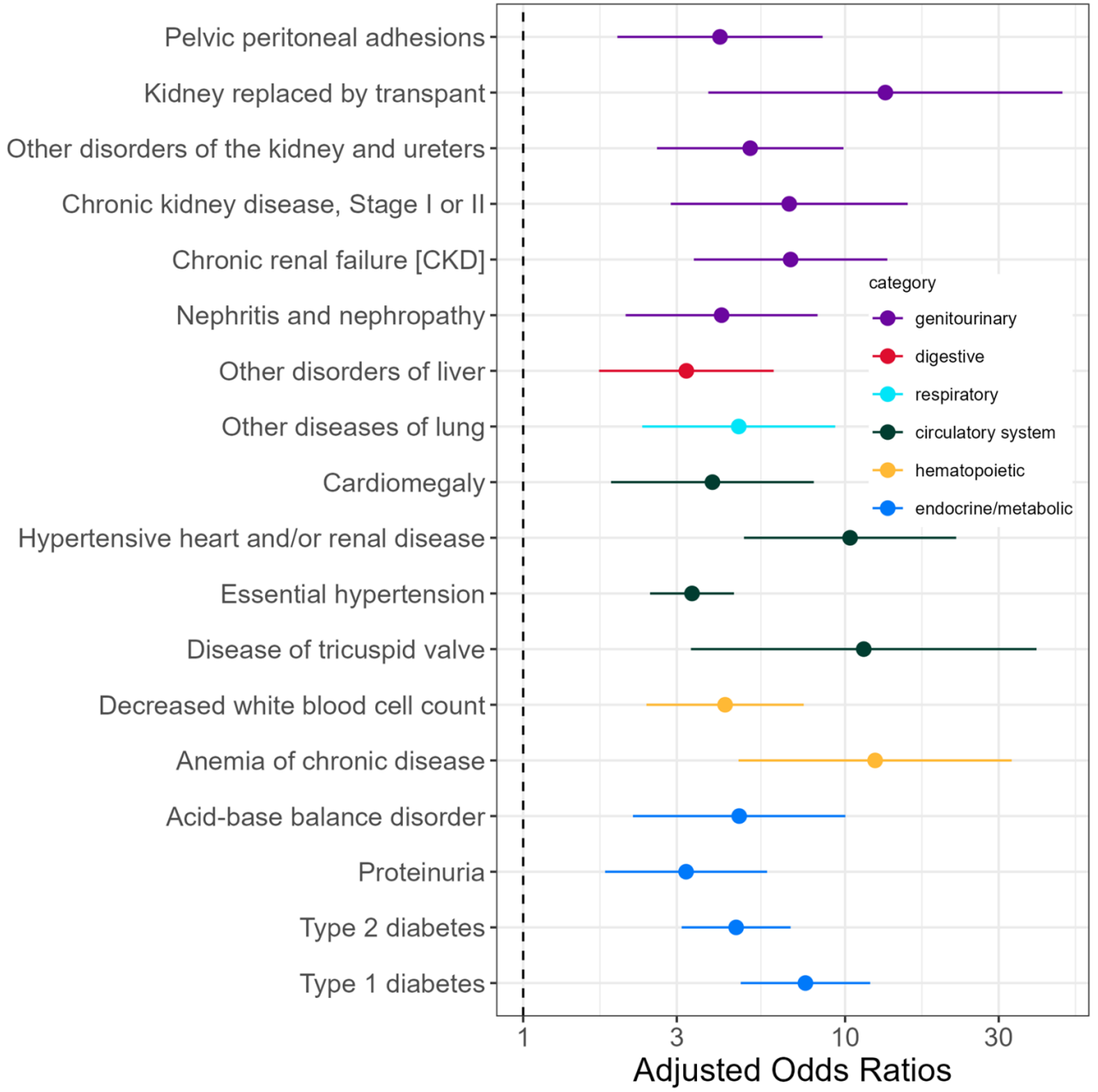


n_{term} = 9692 / $n_{preterm}$ = 975

% of significant ($p_{BH} < 0.05$) iterations

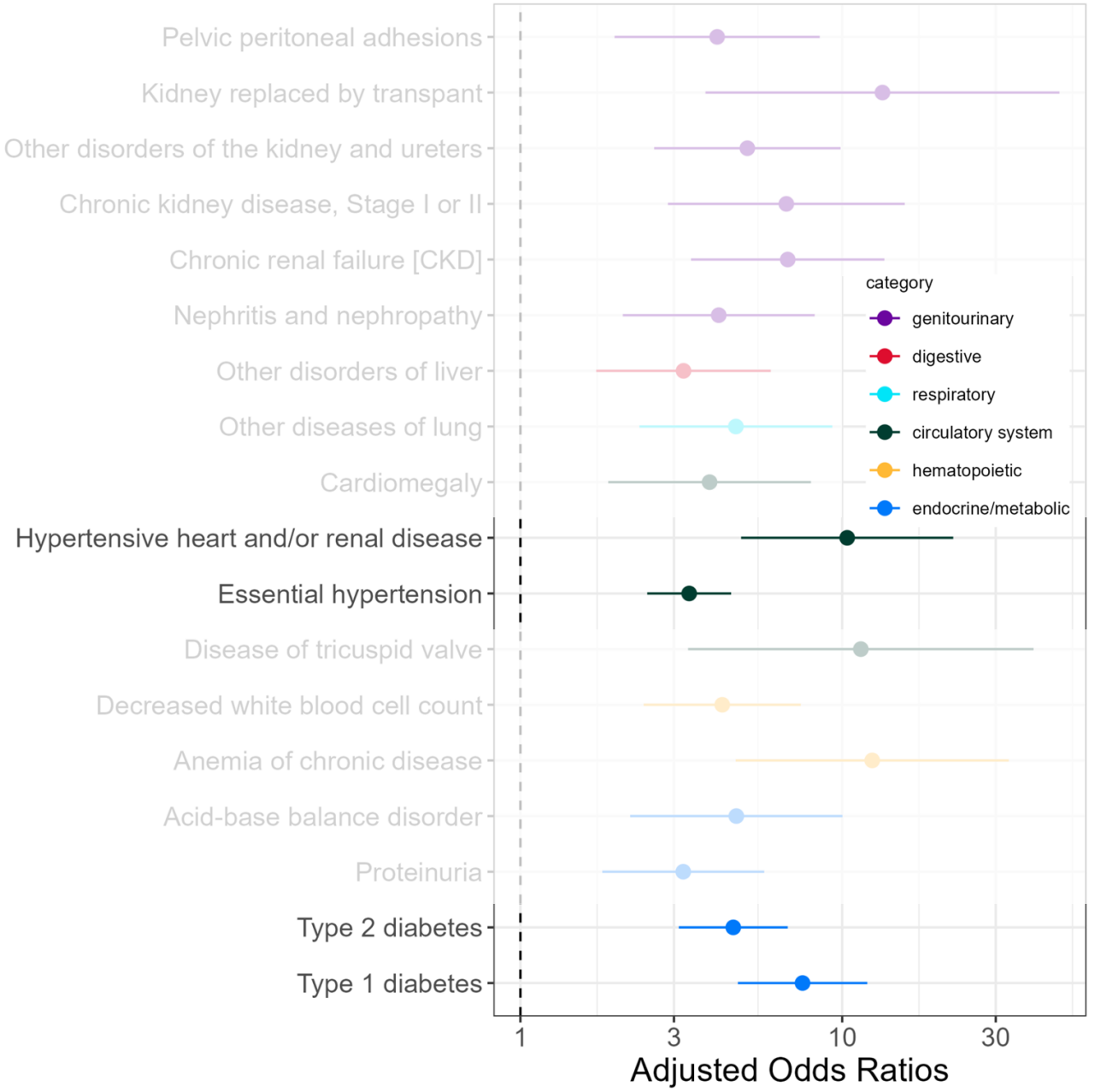
18 conditions are risk factors

None are protective



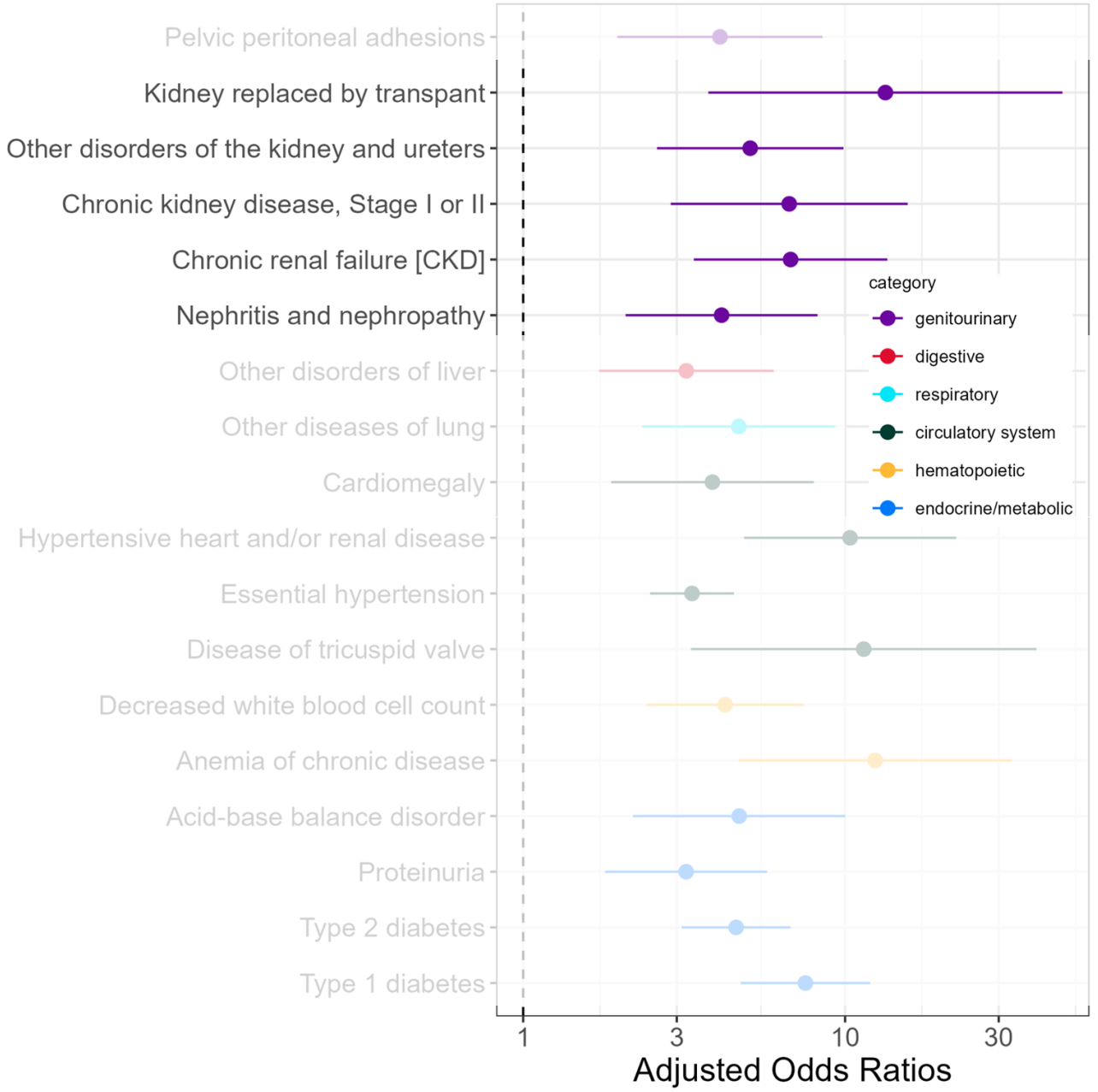
n_{term} = 9692 / $n_{preterm}$ = 975

Known preterm birth risk factors reproduced



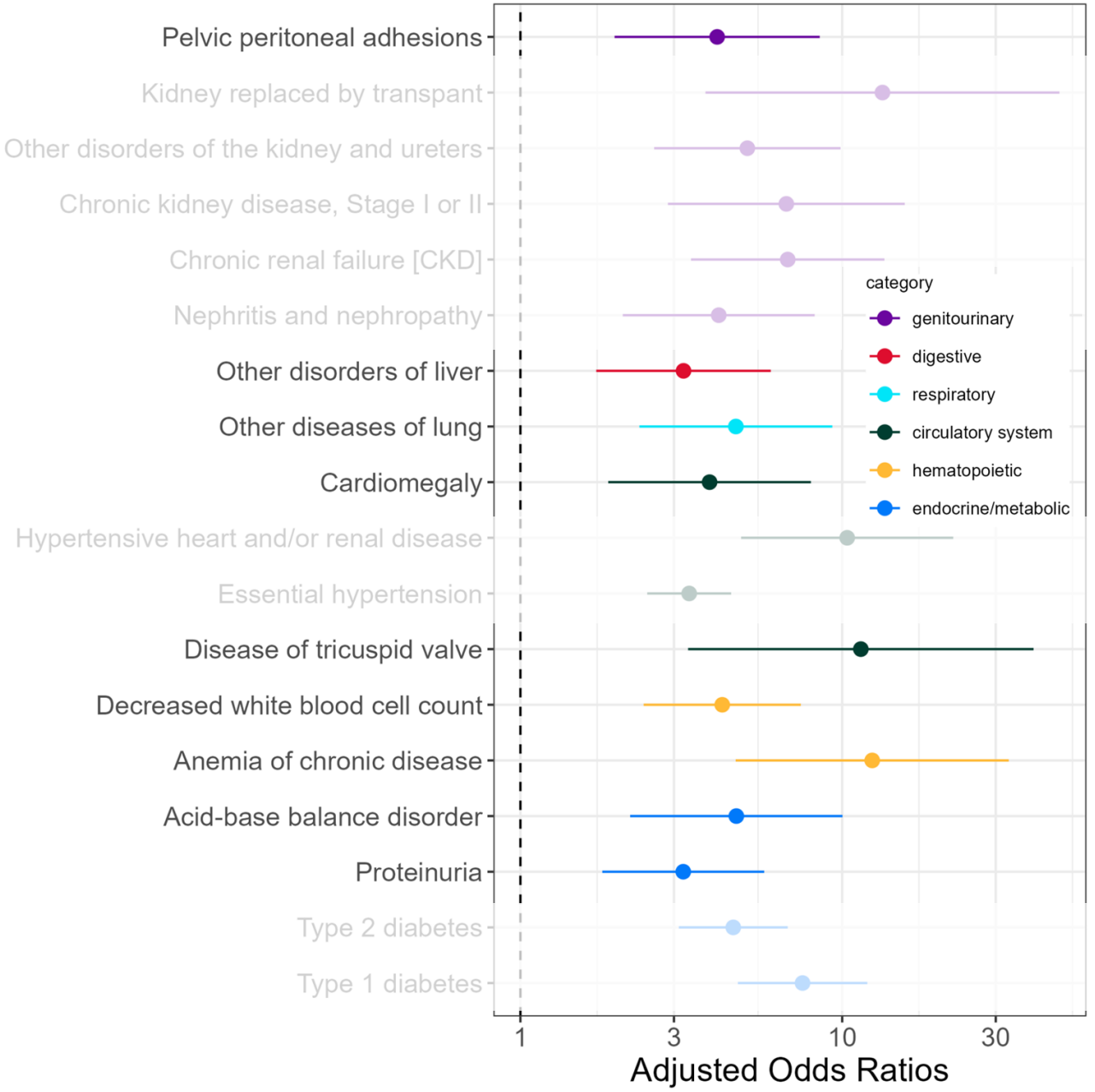
n_{term} = 9692 / $n_{preterm}$ = 975

Known preterm birth risk factors reproduced



n_{term} = 9692 / $n_{preterm}$ = 975

Less established risk factors found

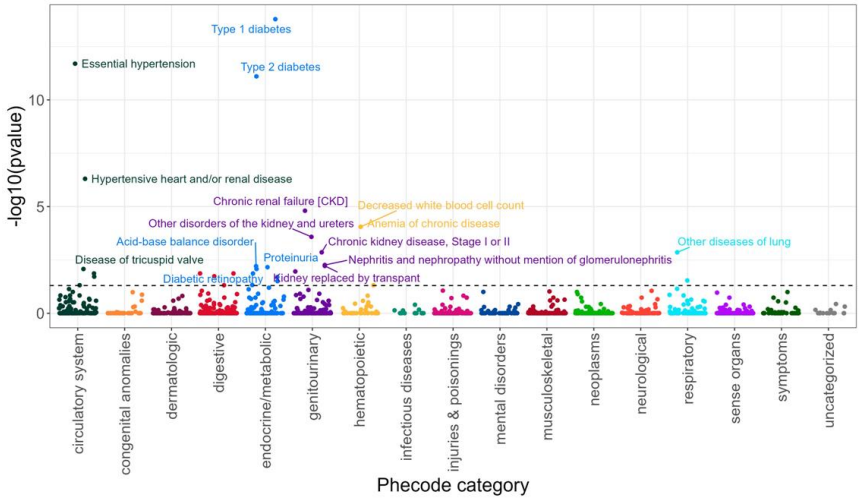


n_{term} = 9692 / $n_{preterm}$ = 975

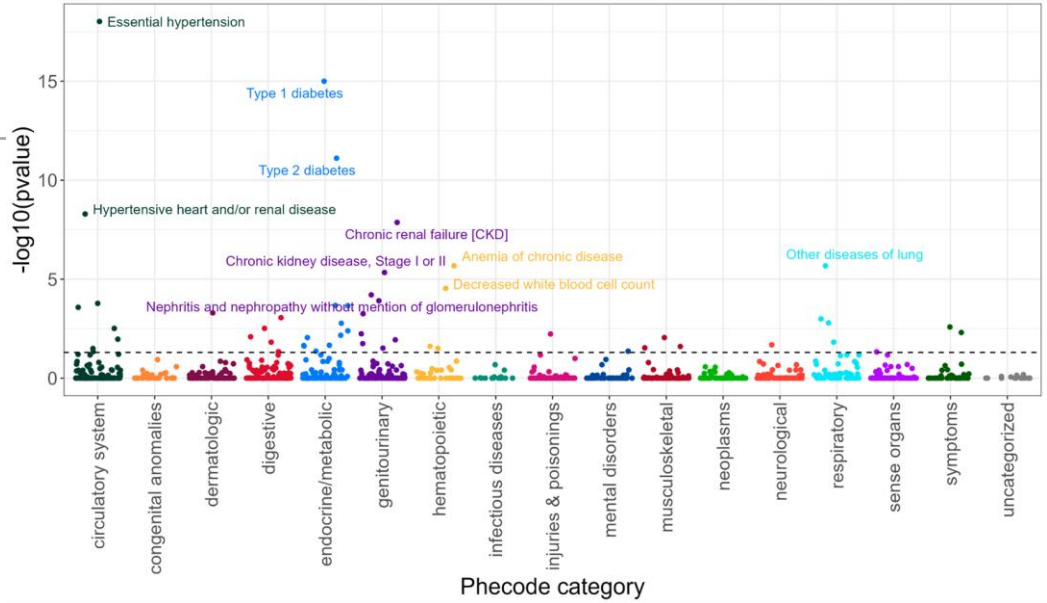
Are there associations with **spontaneous** and **indicated**?

Many associations with **indicated** preterm birth

All Preterm Births
n = 975



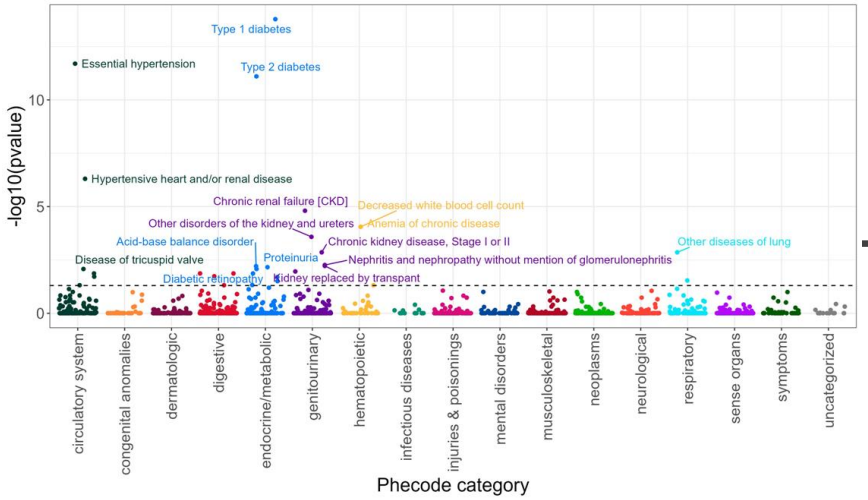
Indicated Preterm Births
n = 421



n_{term} = 9692

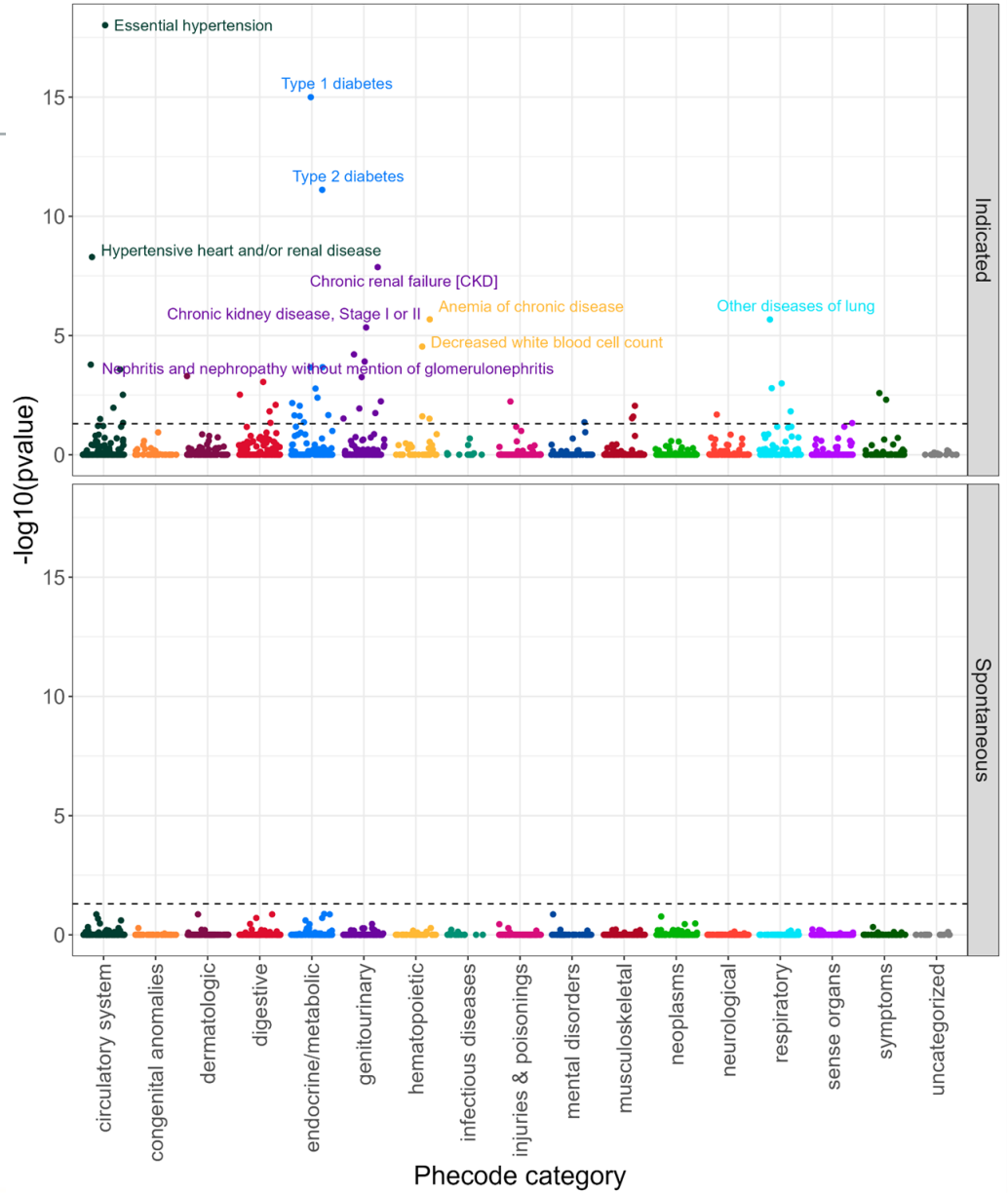
No associations with **spontaneous** preterm birth

All Preterm Births
n = 975



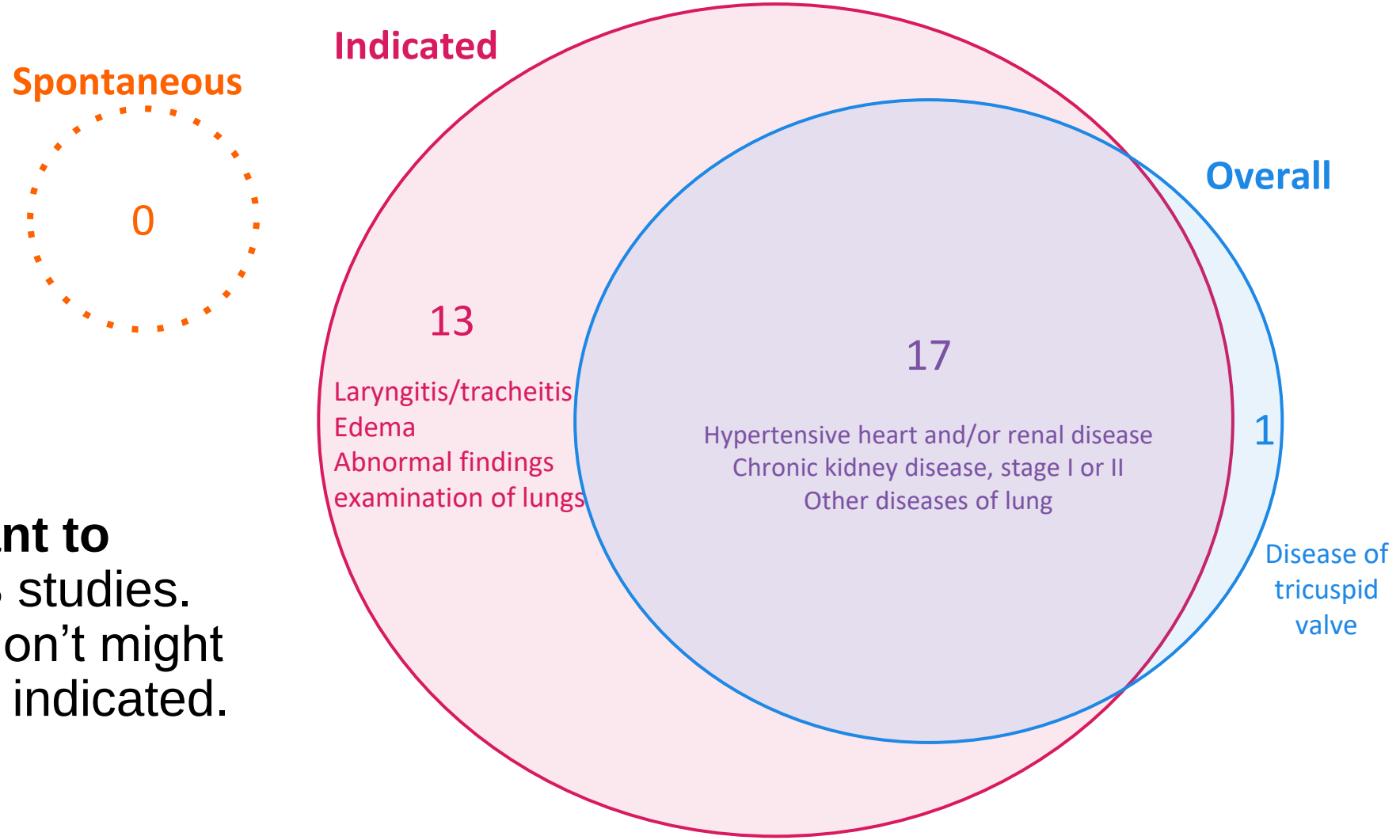
Indicated Preterm Births
n = 421

Spontaneous Preterm Births
n = 447



n_{term} = 9692

Overall preterm birth diagnosis risk factors **overlap** with **indicated**



It is **important to stratify** PTB studies. Those that don't might only capture indicated.

The absence of spontaneous associations is because of **heterogeneity**

Spontaneous individuals had a similar number of diagnoses as indicated individuals, and visit distributions are similar.

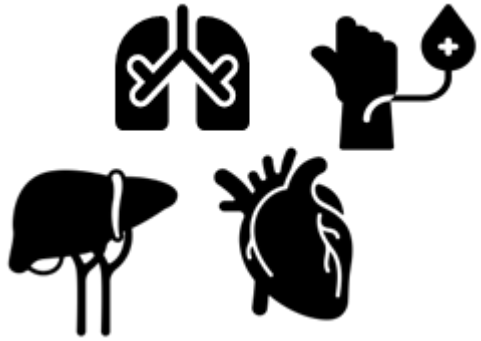
Indicated pre-conception diagnoses

Spontaneous pre-conception diagnoses

Conclusions

Spontaneous risk factors are **more heterogeneous** and/or are outside the factors we measured.

Blood disorders, **cardiac** conditions, **pulmonary** conditions, **liver** conditions, **electrolyte** imbalances, and **digestive** conditions are less established associations that should be further studied.



Icon Credit: thenounproject.com/icon/blood-4508816/, Victoruler; thenounproject.com/browse/collection-icon/internal-human-organs-solid-123918

Conclusions

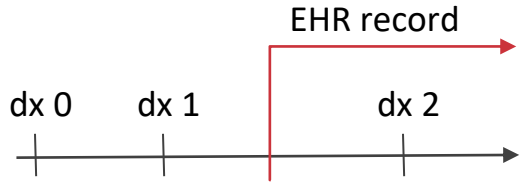
- Hypothesis generation studies should separate spontaneous and indicated preterm births if possible

- Even in data-driven studies, subject matter expertise is critical!

Limitations

UCSF patients are not representative of greater populations.

EHR is not a complete medical history.



Diagnoses are not **binary** (e.g. hypertension, diabetes can be can be well-managed).

Icon Credit: thenounproject.com/icon/blood-4508816/, commons.wikimedia.org/wiki/File:Blank_US_Map,_Mainland_with_no_States.svg

Check out our preprint

Associations with spontaneous and indicated preterm birth in a densely phenotyped EHR cohort

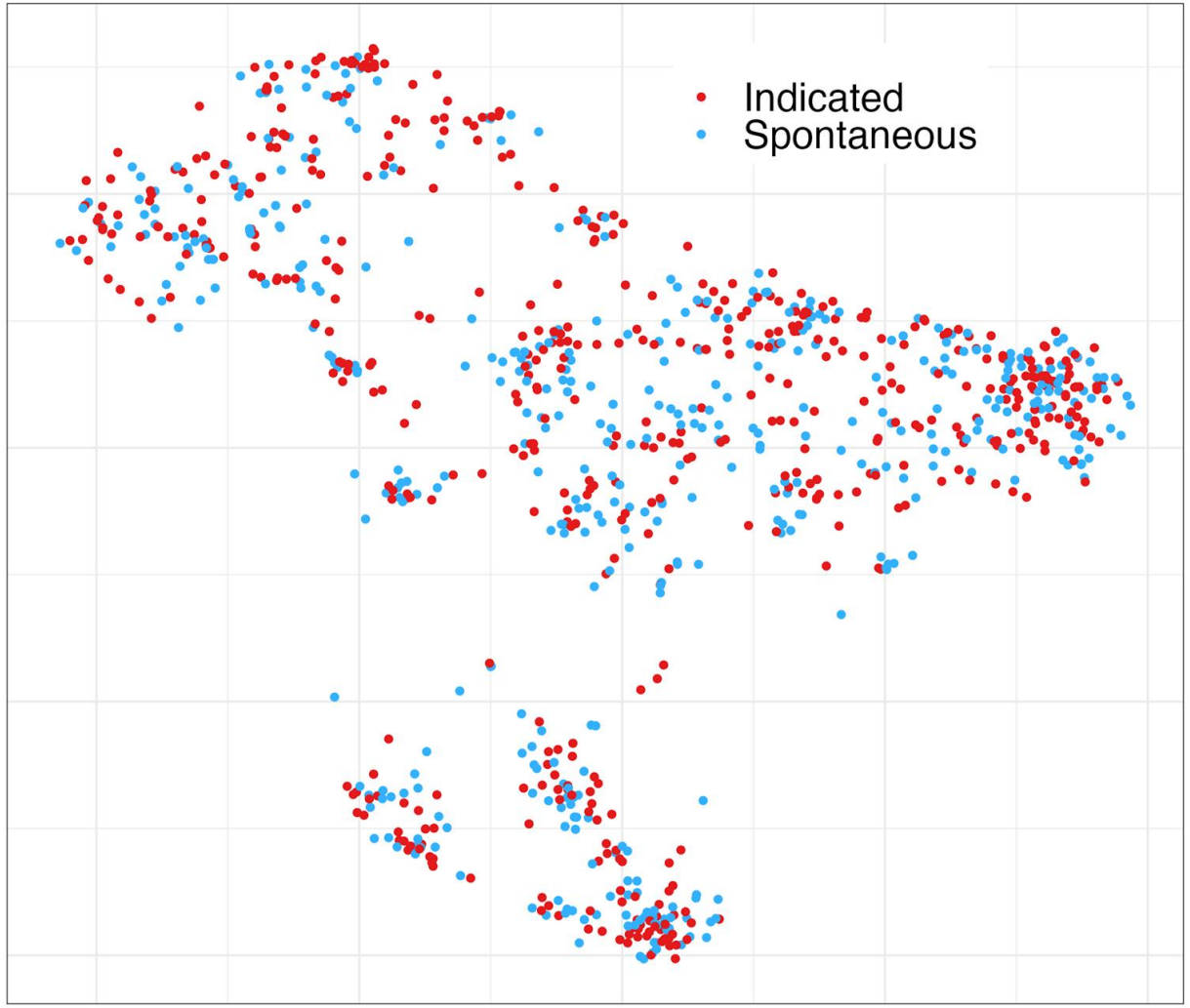
Jean M. Costello, Hannah Takasuka, Jacquelyn Roger, Ophelia Yin, Alice Tang, Tomiko Oskotsky, Marina Sirota, John A. Capra

doi: https://doi.org/10.1101/2023.11.29.23299216

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should *not* be used to guide clinical practice.

Next Steps

- **Validate** findings in another dataset
- Identify **subtypes** beyond spontaneous vs indicated
  - Latent factor analysis
  - Clustering approaches
  - N = 973



Acknowledgements

Capra Lab

Colin Brand
Yaen Chen
Sebastián Cruz González
Grace Ramey
Ava Xu
Tony Capra

PDB Team

Melissa Rosenstein
Cintia Blat

Collaborators

Tim Wen
Safyer Mackenzie-Sampson
KI Team

Sirota Lab

Tomiko Oskotsky
Jacquelyn Roger
Alice Tang
Ophelia Yin
Marina Sirota

Stanford MoD Center

Nima Aghaeepour
David K. Stevenson
Gary Shaw
Ron Gibbs
Feng Xie
Joe Phongpreecha
Davide de Francesco

Funding

March of Dimes
NIH Institutional Training Grant T32-DE007306

Thank you!

Questions?

jean.costello@ucsf.edu

hannah.takasuka@ucsf.edu

Appendix

Identifying conditions

- We only consider conditions prior to conception.
- We allow for multiple deliveries per person.
- We exclude conditions immediately following each pregnancy.

Preterm birth subtypes

Indicated preterm birth

Risk factors include

- Hypertension
- Diabetes
- Advanced maternal age

ACOG & SMFM: Medically Indicated Late-Preterm and Early-Term Deliveries

Baer et al, 2018. Pre-pregnancy or first-trimester risk scoring to identify women at high risk of preterm birth.

Spontaneous preterm birth

Risk factors include:

- short interpregnancy interval
- prior spontaneous PTB
- cervical factors
- Advanced maternal age

Cobo et al, 2020. Risk factors for spontaneous preterm delivery.

Preterm birth subtypes

Indicated (iatrogenic) preterm birth

- a preterm delivery decided upon by a clinician

Spontaneous preterm birth

- preterm delivery *not* indicated by a clinician
- preterm labor, PPROM, etc.

By not distinguishing between indicated and spontaneous, we risk replicating the guidelines used to indicate delivery.

Diabetes		
Pregestational diabetes well-controlled[†]	Full term	39 0/7–39 6/7 weeks of gestation
Pregestational diabetes with vascular complications, poor glucose control, or prior stillbirth	Late preterm/early term	36 0/7–38 6/7 weeks of gestation
Gestational: well controlled on diet and exercise	Full term	39 0/7–40 6/7 weeks of gestation
Gestational: well controlled on medications	Full term	39 0/7–39 6/7 weeks of gestation
Gestational: poorly controlled	Late preterm/early term	Individualized

from ACOG guidelines on indicated delivery

Combining EHR data and subtype information

We have a curated perinatal database (PDB) with detailed delivery information, including whether a preterm delivery was spontaneous or indicated.

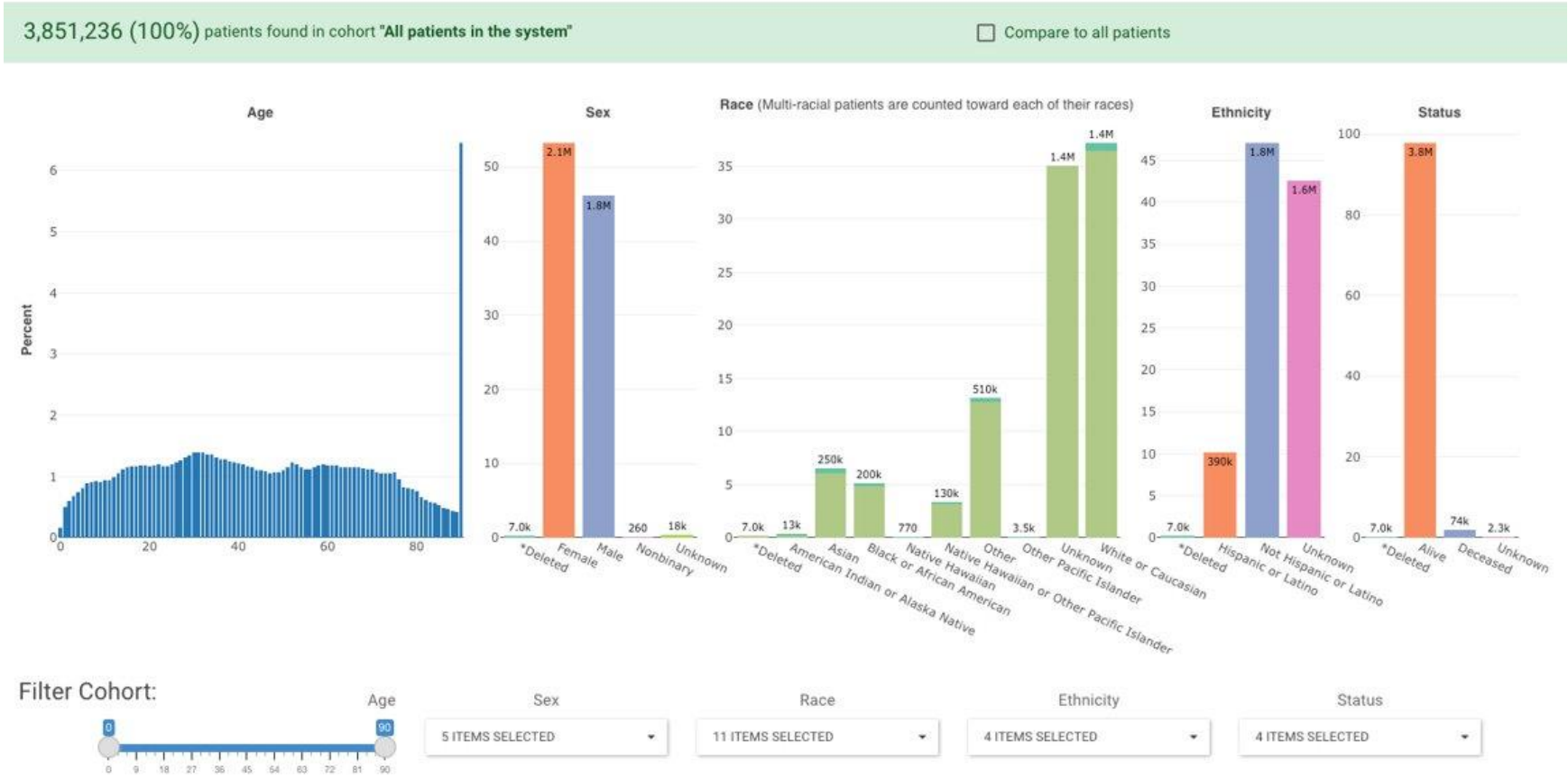
We can further explore associated diagnoses in these subtypes of preterm birth.

Identifying differences between conditions in each subtype may lead to further areas of investigation.

Cohort data - UCSF EHR

UCSF EHR

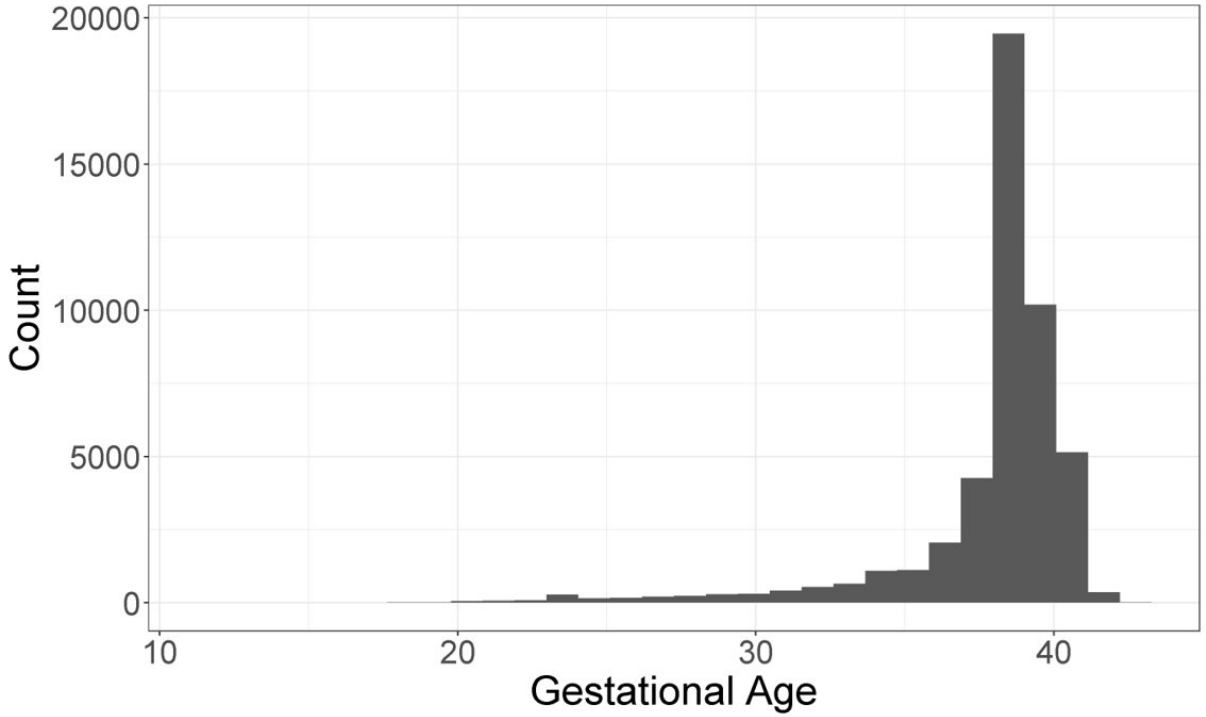
- ~4 million patients
- data types
  - **conditions**
  - medications
  - lab values
  - etc.



Cohort data - Perinatal Database

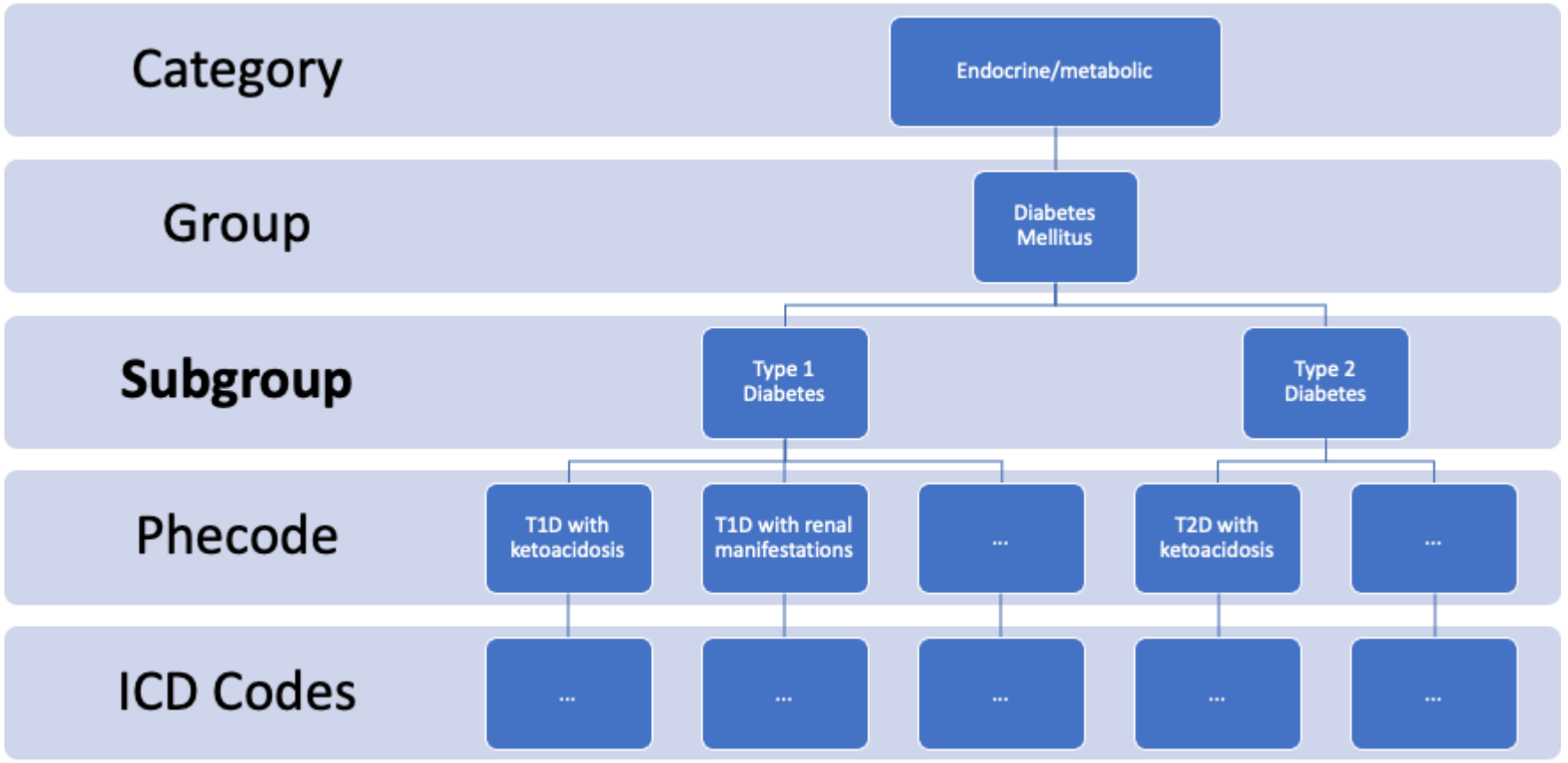
Perinatal Database (PDB)

- 47,350 deliveries
- 47,127 with linked maternal and infant data
- PTB rate: 16.36%
- PTB subtype rates
  - 40.6% indicated
  - 43.8% spontaneous



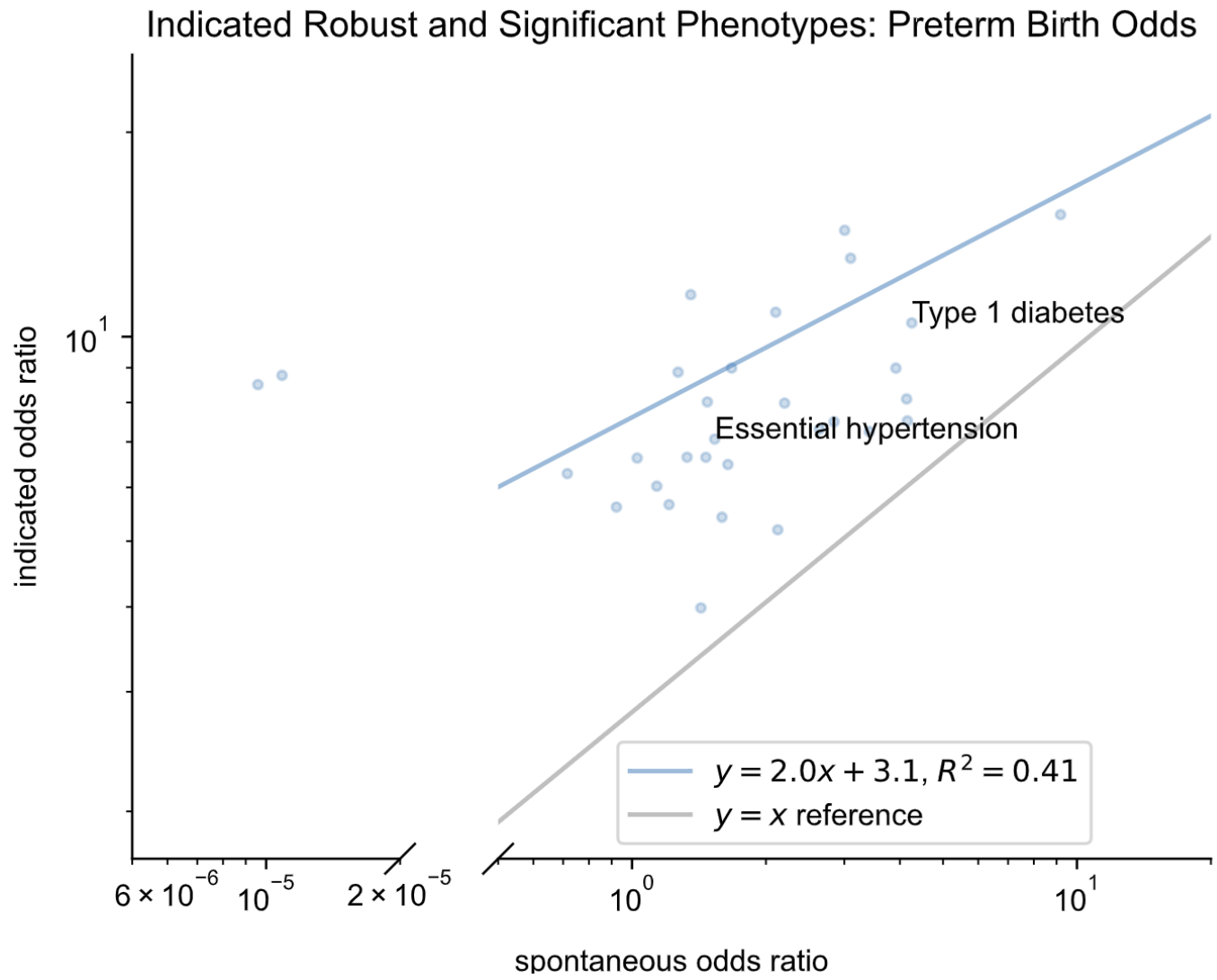
Step 1: Categorizing conditions

Phecodes are a hierarchical grouping of ICD codes

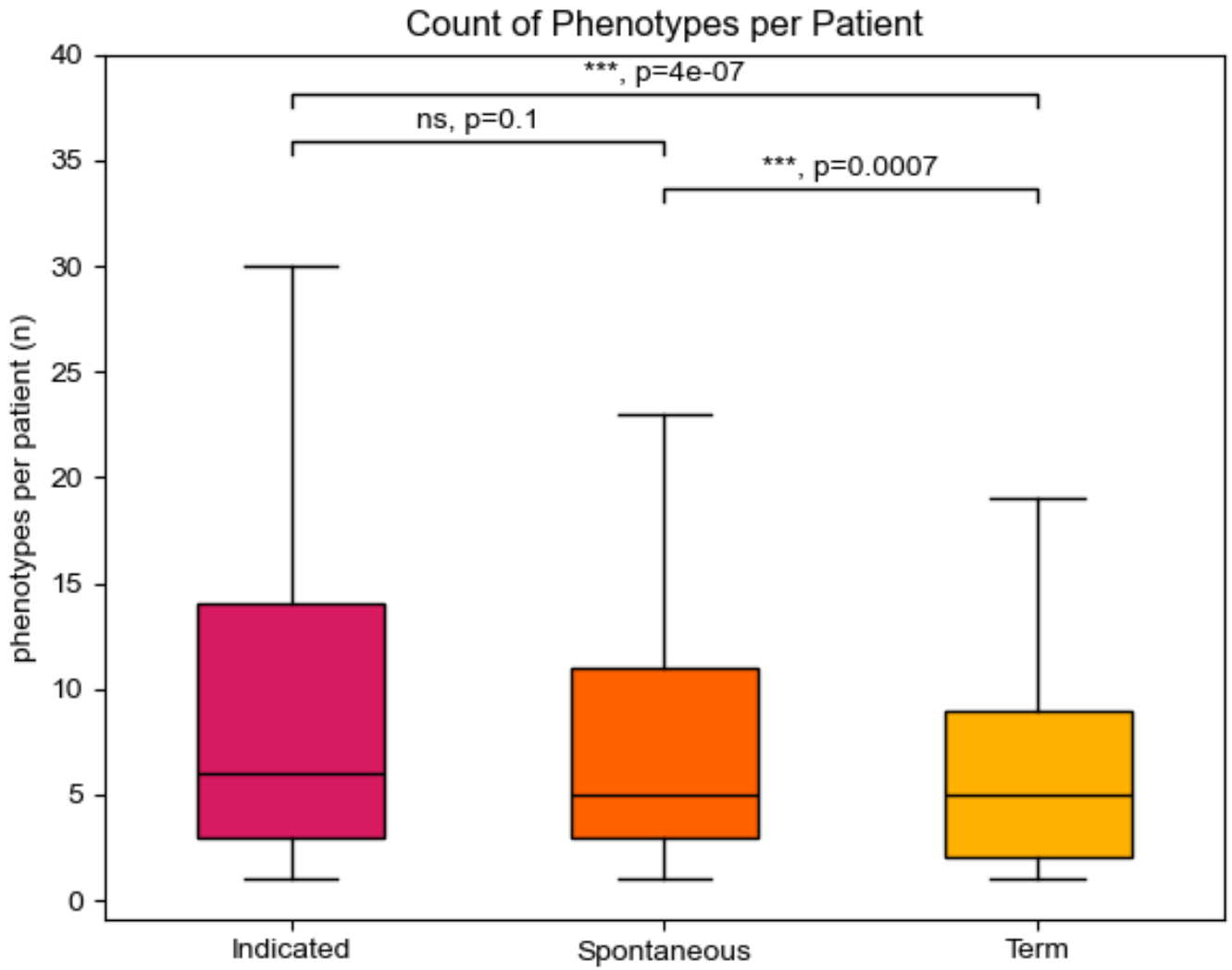


1,326 unique conditions present

Association trends are usually similar, but much weaker in spontaneous cohort

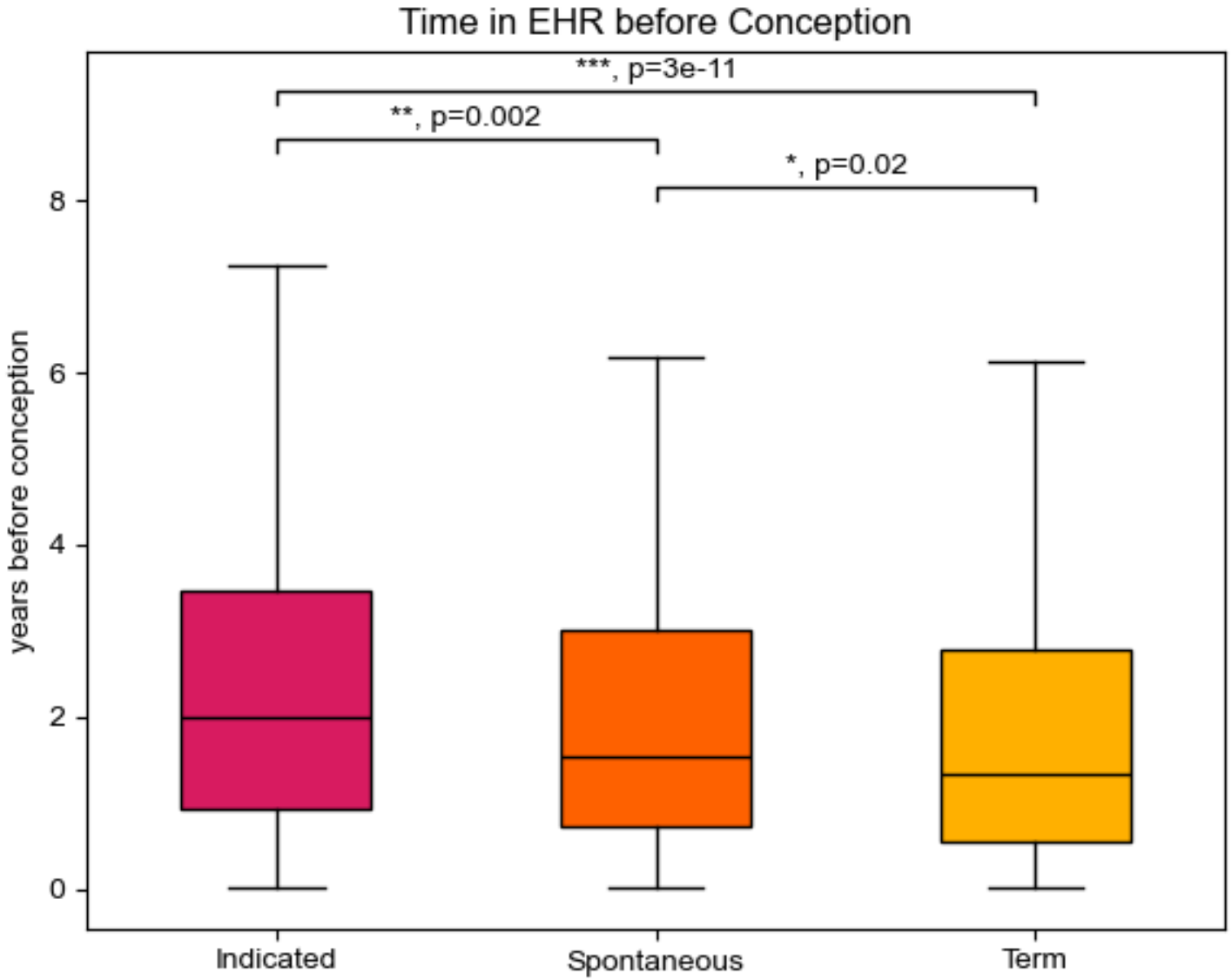


Spontaneous and indicated have no difference in # of conditions.



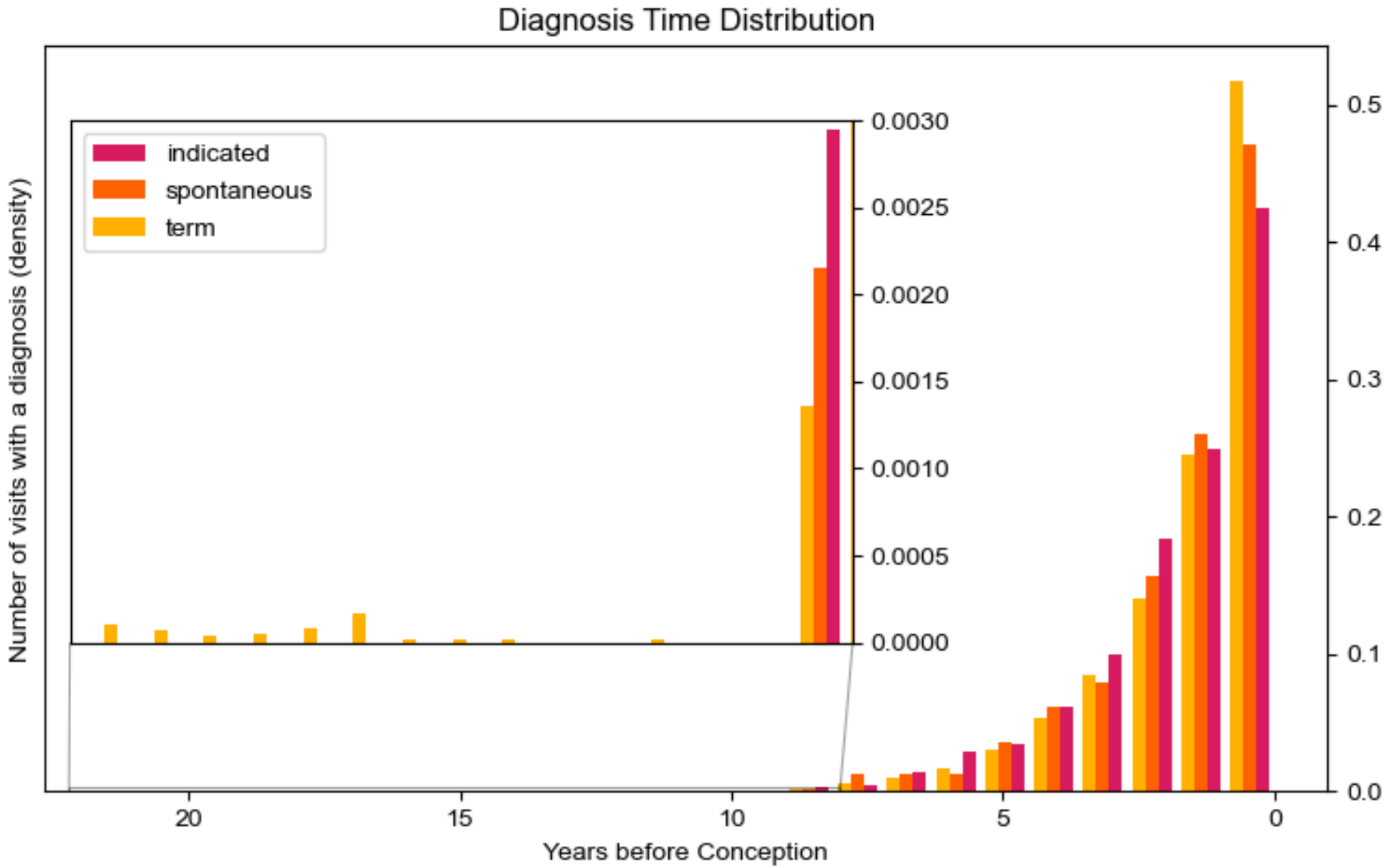
2-sided Mann-Whitney U rank test, outliers dropped

Spontaneous median length of time in EHR is 6 months less than indicated



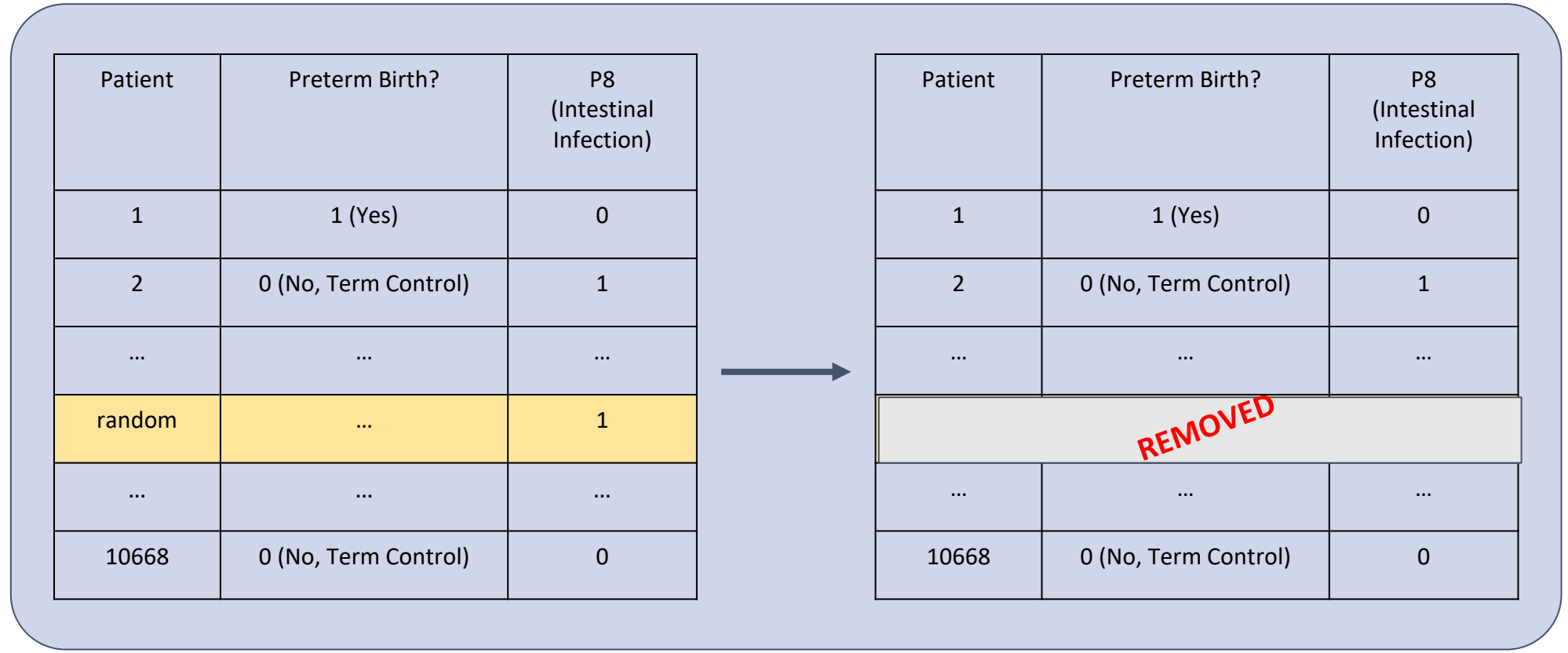
2-sided Mann-Whitney U rank test, outliers dropped

Visit distribution times are similar



Testing associations for robustness

Is this iteration significant ($p_{BH} < 0.05$)?



Patient	Preterm Birth?	P8 (Intestinal Infection)
1	1 (Yes)	0
2	0 (No, Term Control)	1
…	…	…
random	…	1
…	…	…
10668	0 (No, Term Control)	0

Patient	Preterm Birth?	P8 (Intestinal Infection)
1	1 (Yes)	0
2	0 (No, Term Control)	1
…	…	…
REMOVED		
…	…	…
10668	0 (No, Term Control)	0

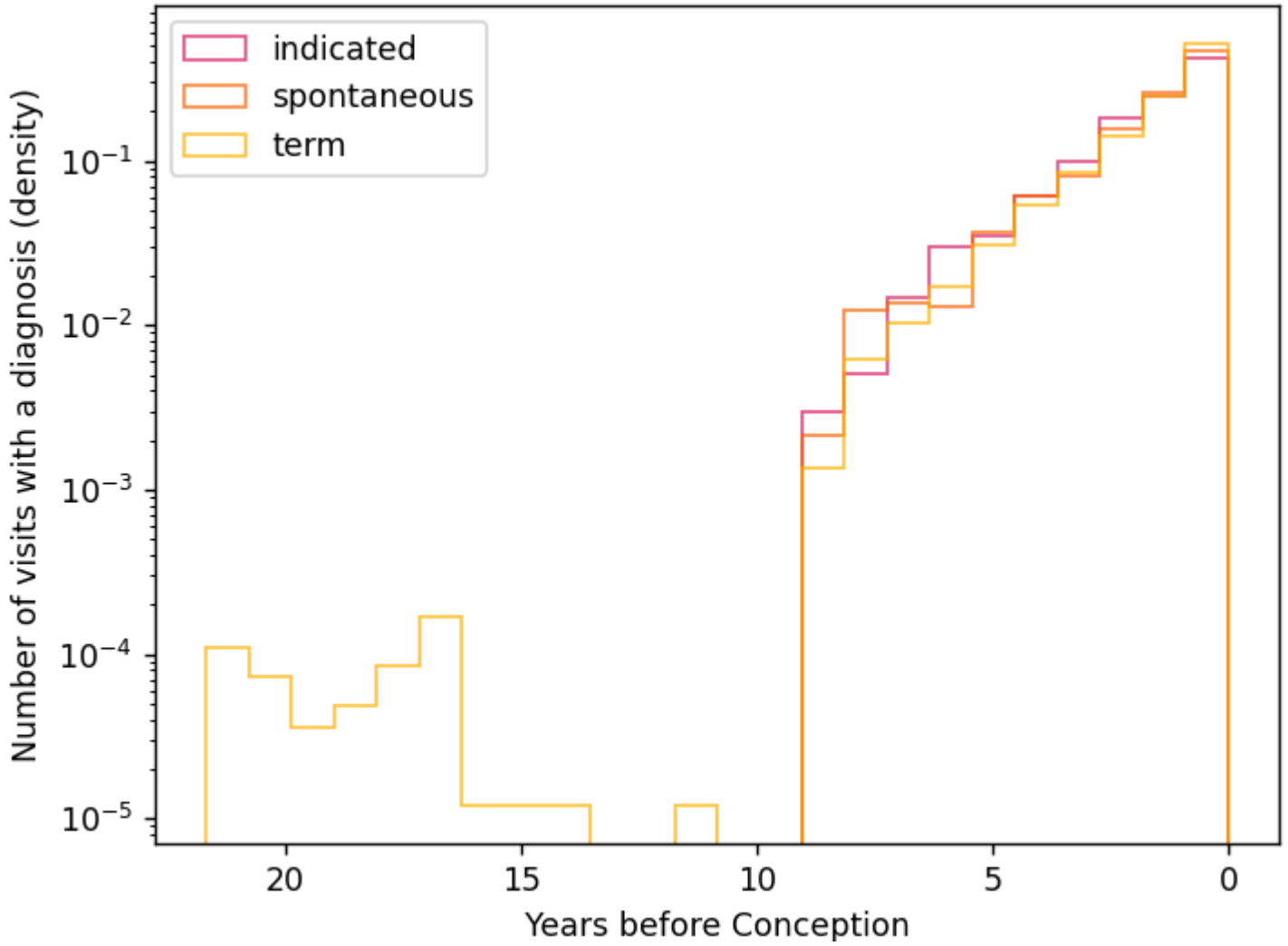
50 bootstrap iterations

robust phenotype? significant in 50/50 iterations

The absence of spontaneous associations is because of heterogeneity.

It is not caused by:

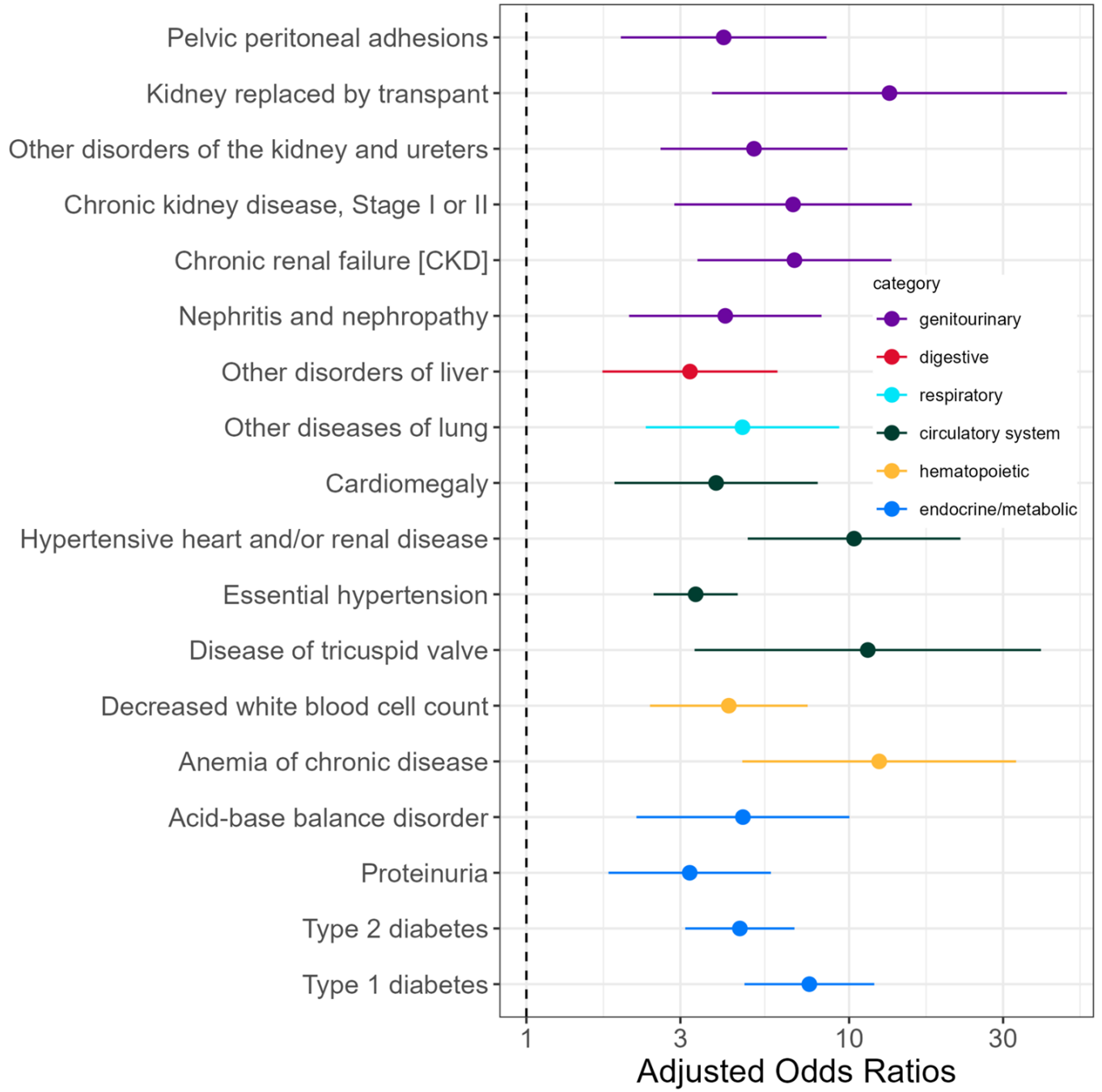
- fewer diagnoses in spontaneous than indicated
- a difference in time in EHR before conception
- a difference in visit distribution



Overall results

Many diagnoses were significantly associated (adjusted for multiple hypotheses) with PTB

- No diagnoses were protective
- Some diagnoses have fairly small sample sizes…



Joint work with Hannah Takasuka, Capra Lab

Interpretation (overall)

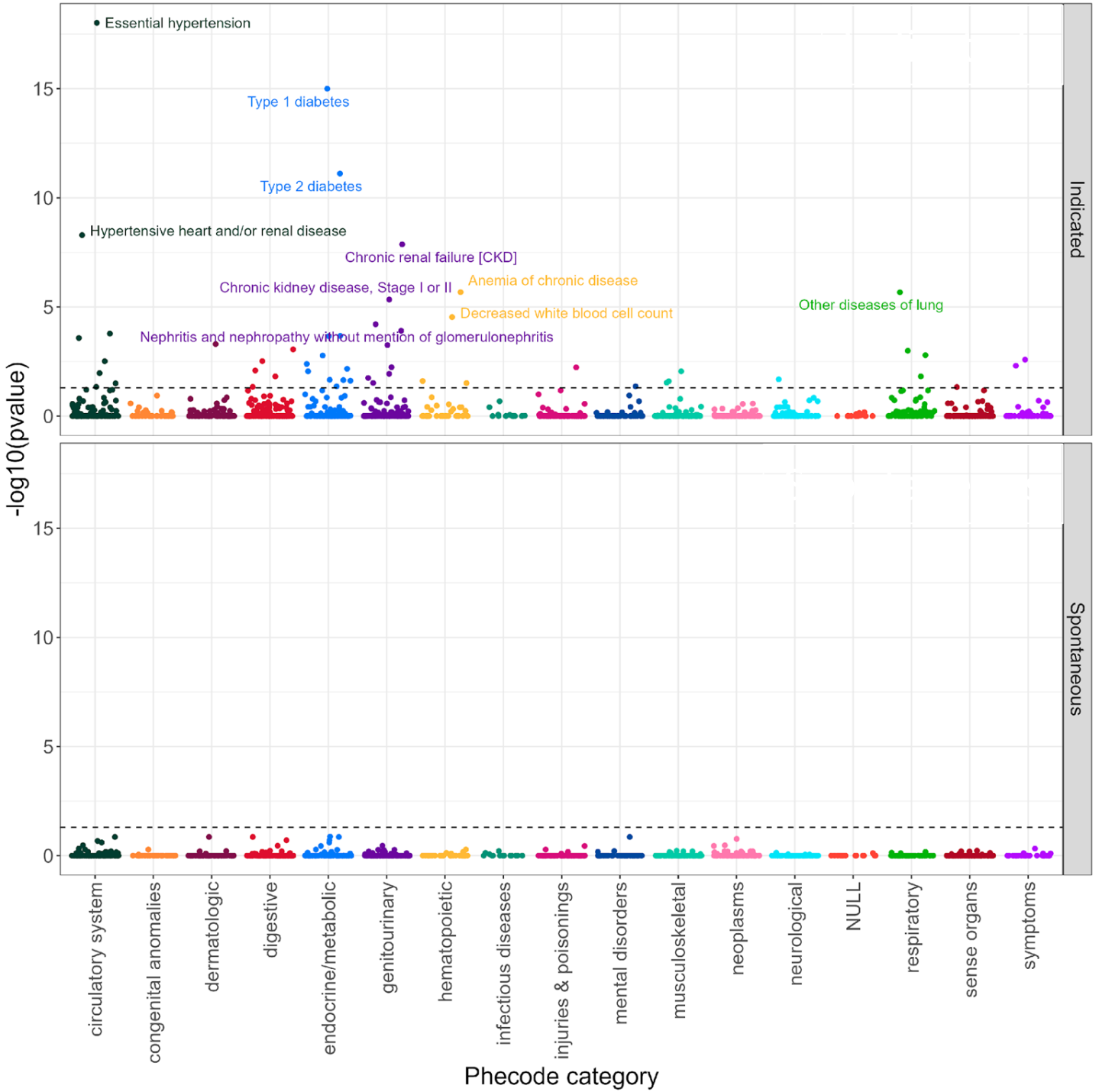
- We replicated known results
  - Hypertension
  - Diabetes
  - Kidney diseases
- Decreased white blood cell counts are likely indicative of serious health conditions
  - Cancer, HIV, immune conditions, etc.
  - **We will follow up with lab test data in the same cohort**
- Our most robust results are well-established

Joint work with Hannah Takasuka, Capra Lab

Subtype results

The diagnoses significant in the indicated group overlap with the overall group.

The spontaneous subtype had no significant diagnoses.



Joint work with Hannah Takasuka, Capra Lab