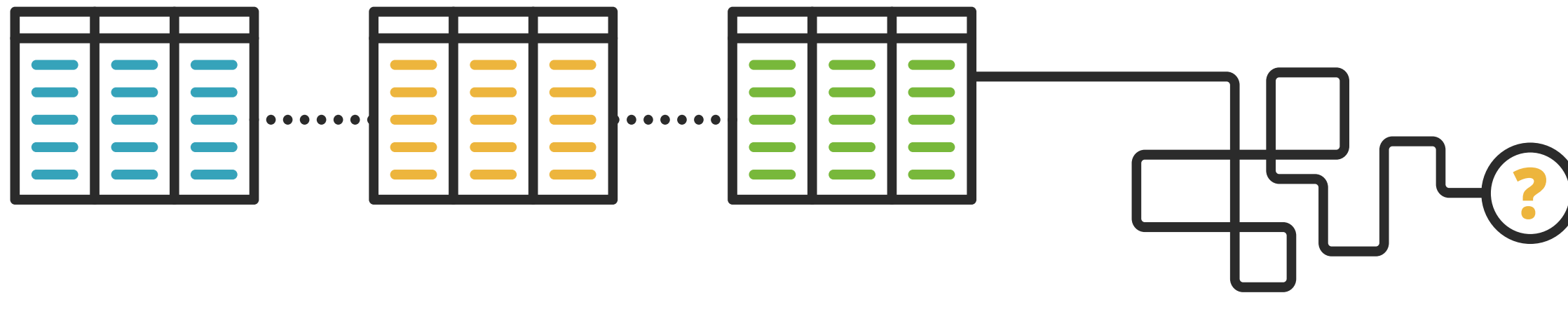


# dxpr:

## An R package for generating analysis-ready data from electronic health records—diagnoses and procedures

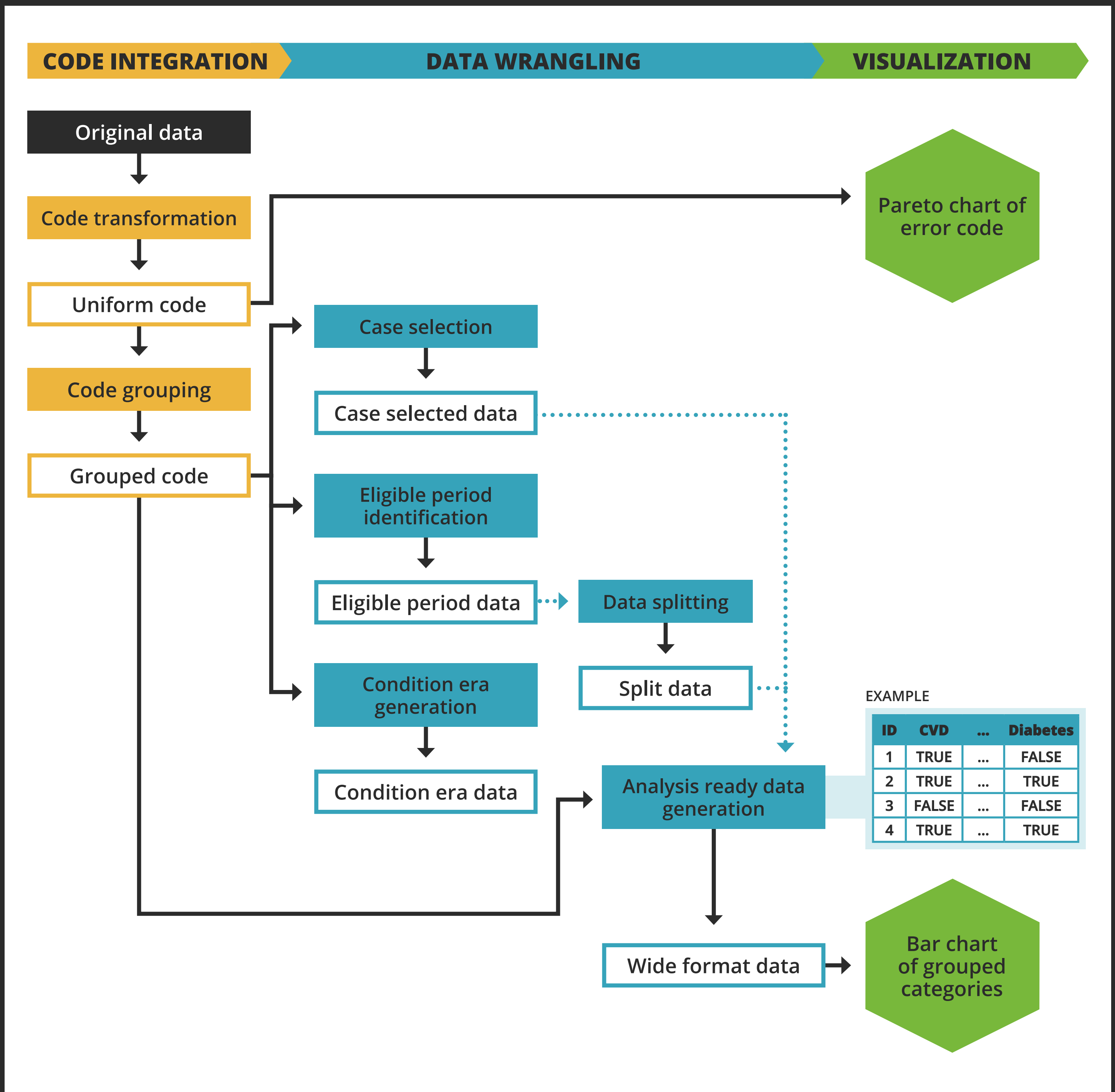
### BACKGROUND

Healthcare data analytics is an essential process that helps accelerate clinical research. However, **processing and analyzing EHR data commonly bottlenecks healthcare data analytics.**



### METHODS

The *dxpr* R package provides mechanisms for integration, wrangling, and visualization of clinical data, including diagnosis and procedure records.



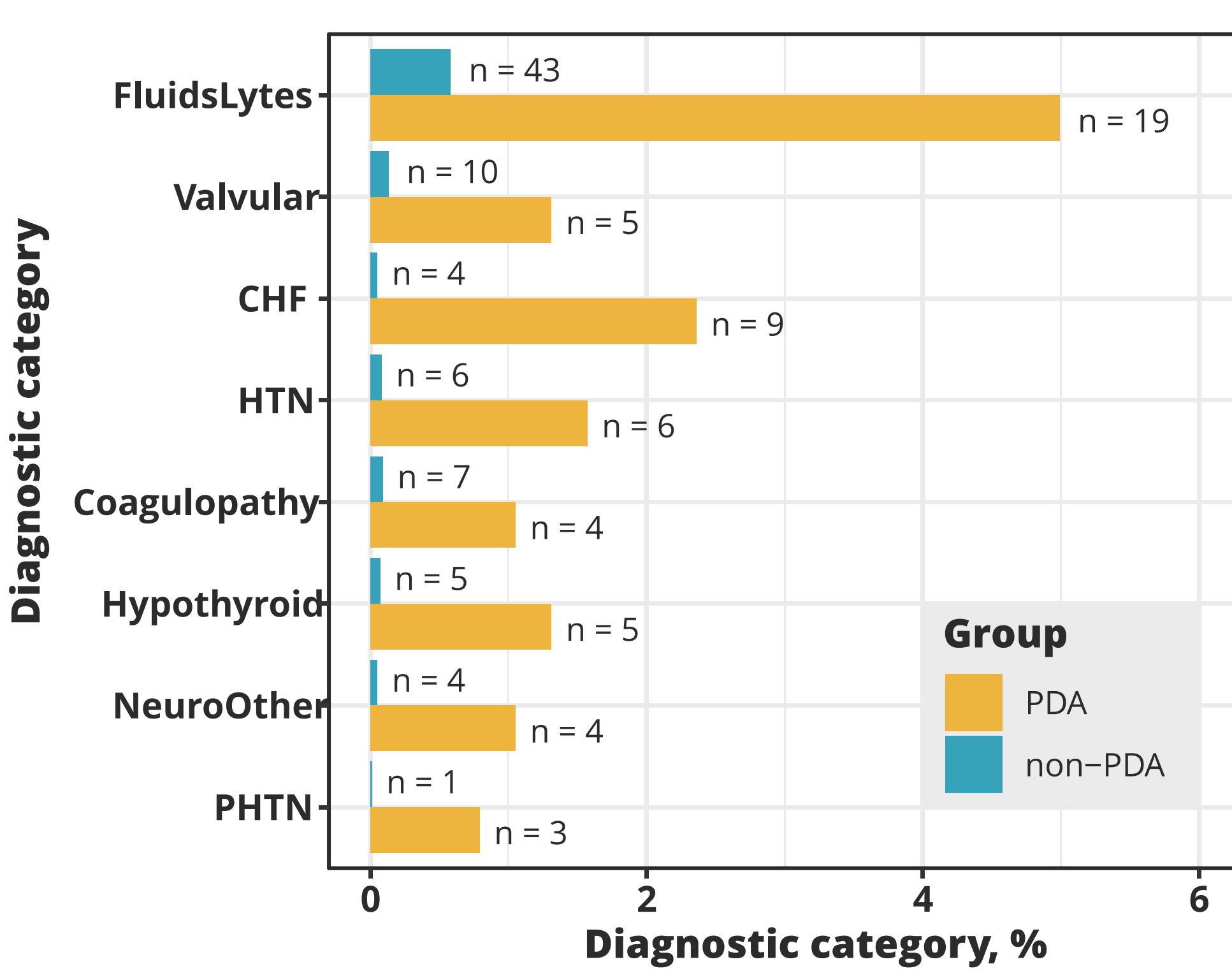
First, the *dxpr* package helps users uniformly reformat International Classification of diagnosis codes (ICD). Then, the *dxpr* package supports **four strategies for grouping clinical diagnostic data.** After EHRs are integrated, users can employ a set of flexible built-in querying functions for dividing data into case and control groups. Subsequently, **the structure of integrated long data can be converted into wide, analysis-ready data that are suitable for statistical analysis and visualization.**

### RESULTS

We conducted comorbidity data processes based on a cohort of newborns from MIMIC-III (n = 7,833) by using the *dxpr* package. We first defined patent ductus arteriosus (PDA) cases as patients who had at least one PDA diagnosis (ICD-9 7470\*). Controls were defined as patients who never had PDA diagnosis. Then, we grouped the diagnoses into defined comorbidities.



Diagnostic category: Top 10



Finally, we observed a statistically significant difference in 8 of the 16 comorbidities, including fluid and electrolyte disorders and valvular disease, among patients with and without PDA.

### CONCLUSION



Find *dxpr* here:  
<https://github.com/DHLab-TSENG/dxpr>

This *dxpr* package helps clinical data analysts address common bottlenecks caused by clinical data characteristics, such as heterogeneity and sparseness.

