



Une filiale de DOCAPOSTE

# Application of the high-dimensional propensity score on medical administrative databases

## An example in assessing comparative mortality in COPD patients on triple therapy

Arnaud Panes<sup>1</sup>, Japheth Baloumbou-Nzamba<sup>2</sup>, Baptiste Pitel<sup>1</sup>, Stanislas Perrier<sup>3</sup>, Anne-Lise Vataire<sup>3</sup>, Nicolas Pages<sup>1</sup>, Nada Assi<sup>1</sup>, **Fanny Raguideau<sup>1</sup>**

1, Heva, Lyon, France,  
2, Université Paris Saclay, France  
3, Sanofi, 82 avenue Raspail 94250 Gentilly

ISPE 2025  
Monday, August 25, 2025



LA POSTE  
SANTÉ &  
AUTONOMIE

Arnaud Panes, Japheth Baloumbou-Nzamba, Baptiste Pitel, Nicolas Pages, Nada Assi, Fanny Raguideau are employees of Heva.

Stanislas Perrier and Anne-Lise Vataire are employees of Sanofi.

The work presented was funded by Heva.



## Confounding bias in observational studies

Despite using large databases



## Propensity score (PS)

A priori selection of variables by experts



## High-dimensional propensity score (HDPS)

Automatic selection of covariates without a priori selection

## Objective

To compare the performance of these 2 matching methods in order to assess mortality in COPD patients on triple therapy in France



Retrospective observational study using the French National Health Data System (SNDS)

68,7 million people lifetime



Population: COPD patients aged >40 years who had initiated triple therapy in 2015



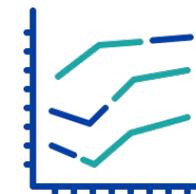
Application of two matching methods (PS and HDPS) to compare mortality of patients (with and without exacerbations)



Construction of PS and HDPS



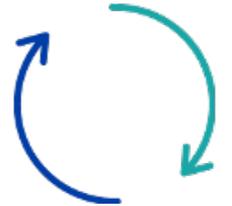
Sensitivity analyses varying the number of covariates in the HDPS (k=100, 200, 300, 400, 500)



Survival analyses (Kaplan-Meier curves and Cox models) on the matched cohorts

# Construction of HDPS score

## 1. Selection of data dimensions (6 dimensions)



Inpatient diagnosis

Drugs dispensing  
(outpatient  
and inpatient  
expensive drugs)

Lab tests  
reimbursements

Medical devices  
dispensing

Medical  
consultations

Long term disease  
registration  
for full  
reimbursement

ICD 10 codes  
3 digits

ATC codes

NABM codes

Chapters of LPP  
codes list

Consultation  
codes

ICD 10 codes  
3 digits

3 years  
look back period

5 years  
look back period

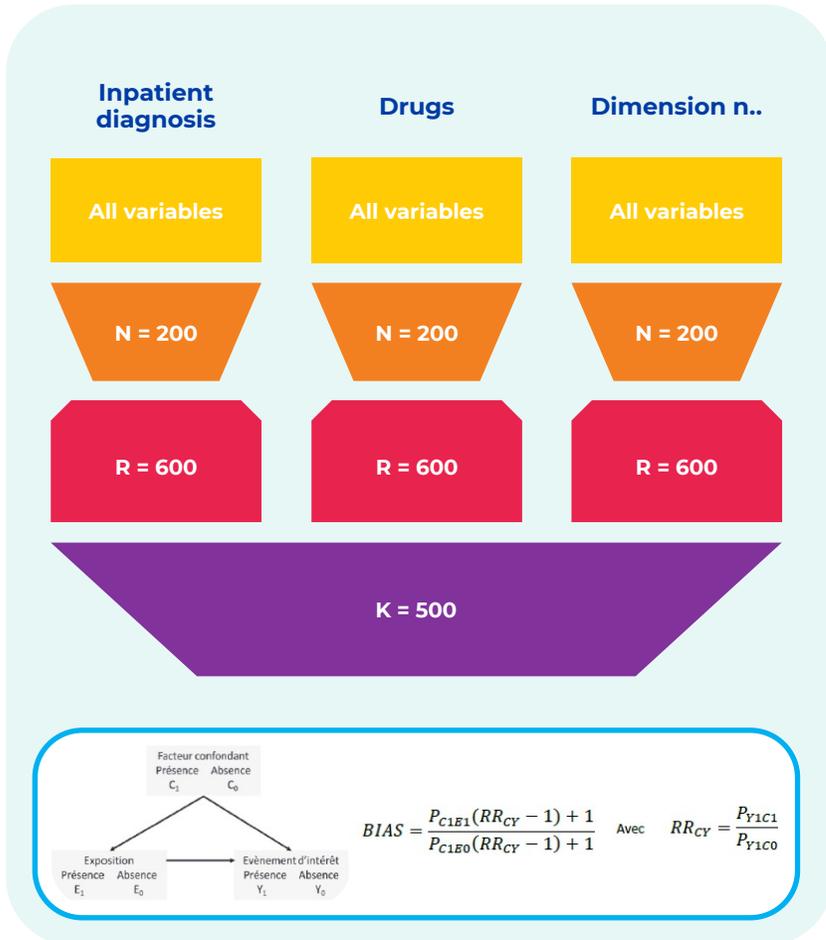
3 years  
look back period

5 years  
look back period

3 years  
look back period

3 years  
look back period

**Addition of covariables used in PS** (and not present in selected dimensions)



## 1. Selection of data dimension (6 dimensions)



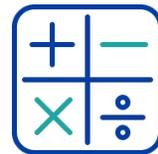
## 2. Identification of empirical variables

Selection of the 200 most prevalent covariates in each dimension



## 3. Assessment of occurrence of covariates

Creation of 3 sub covariates per covariate based on occurrence (once, sporadic or frequent) (200\*3= 600 covariates)



## 4. Assessment of potential adjustment

Calculation of the multiplicative bias introduced by each covariate using Bross formula.



## 5. Selection of final covariates

Ranking of the 3600 covariates by potential bias and selection of top k (K=100, 200, 300, 400, 500)

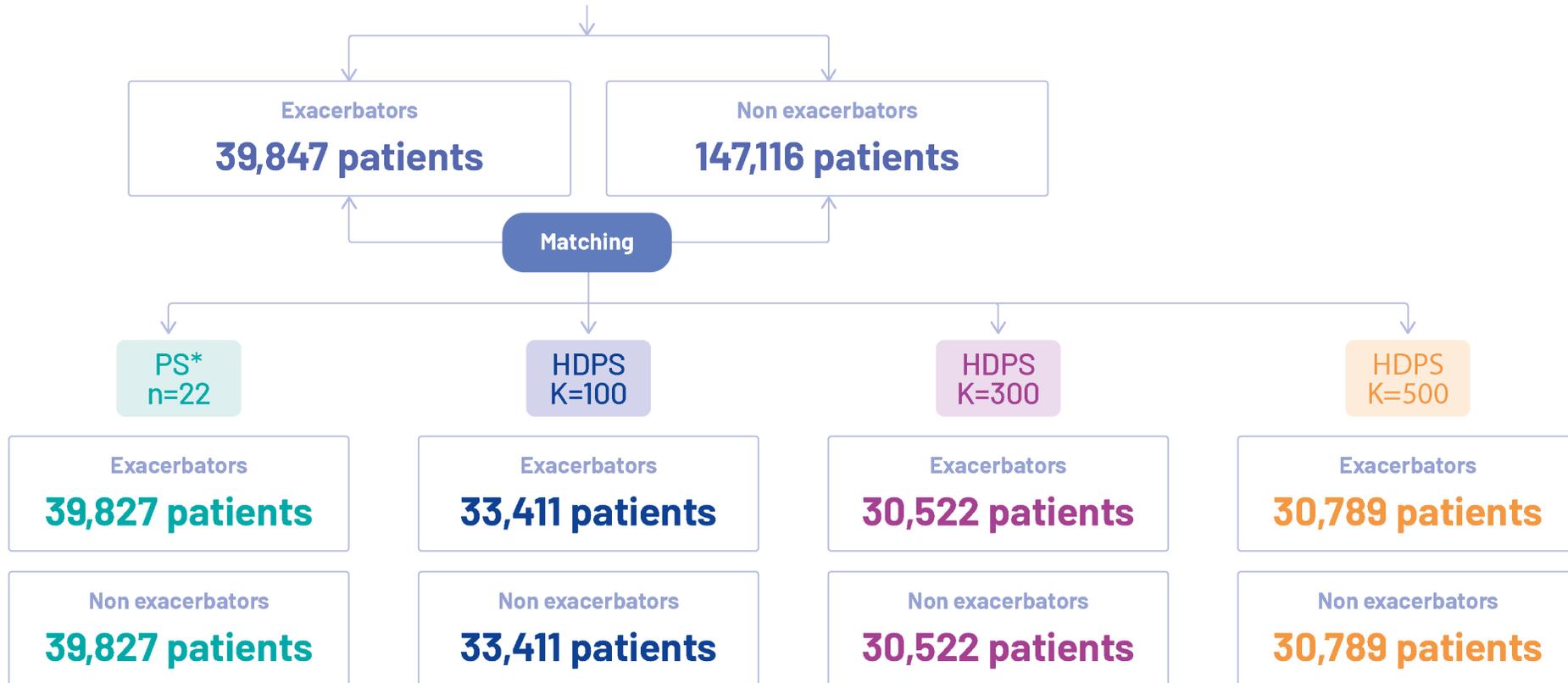


## 6. Use of HDPS as a classical PS (matching)

1:1 matching (Greedy matching) with caliper = 0.2 × SD (log[P(A=1 | Xi)]).

### BPCO patients aged over 40 years initiating triple therapy in 2015

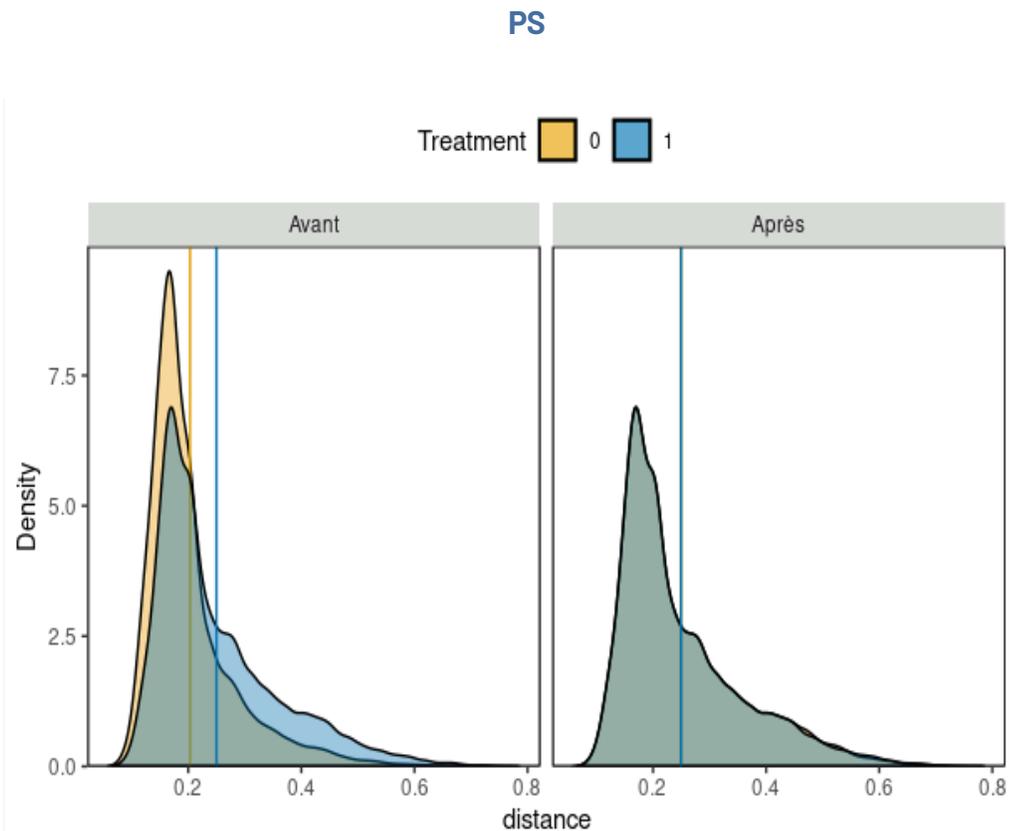
**186,963 patients**



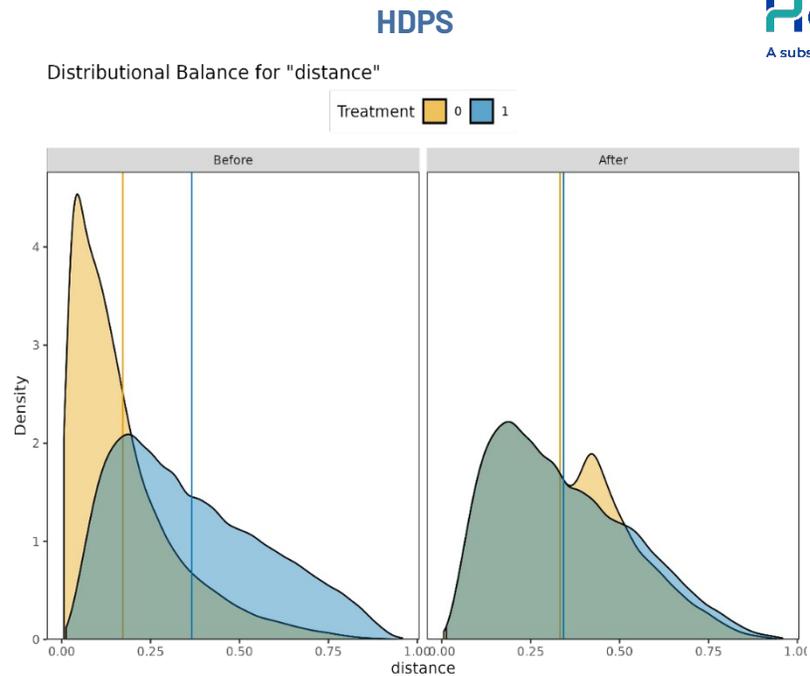
\*PS variables: age, sex, geographical density of pneumologists, anxiety, dementia, hemiplegia, asthma, pneumonia, sleep apnea, myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular diseases, HIV/AIDS, diabetes, osteoporosis, undernutrition, moderate or severe kidney diseases, tumors, connective tissue diseases, ulcerative diseases and liver diseases.

# Performance of covariates balance

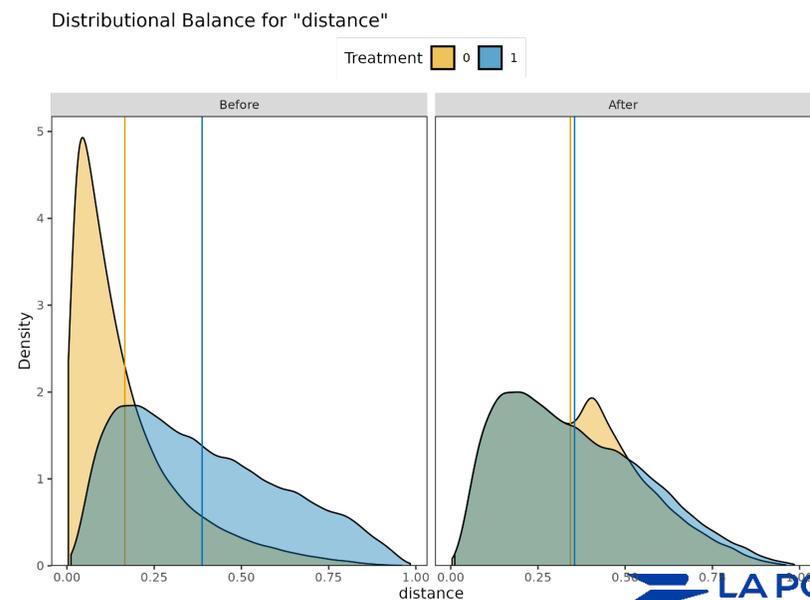
## Plots of the PS distribution



K = 100

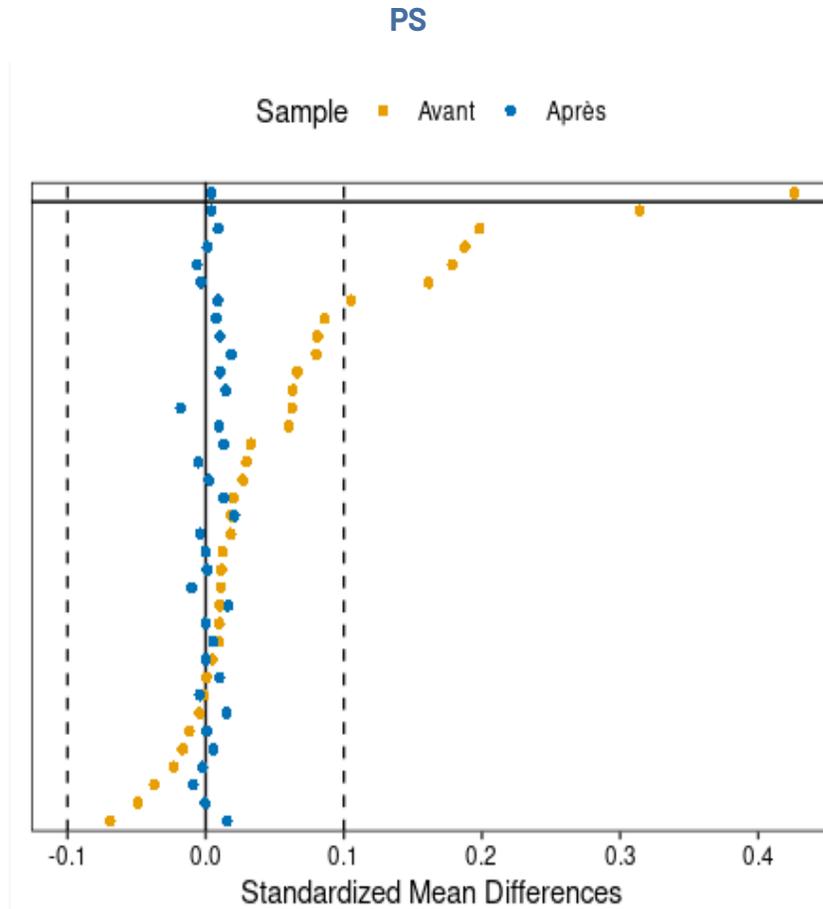


K = 500

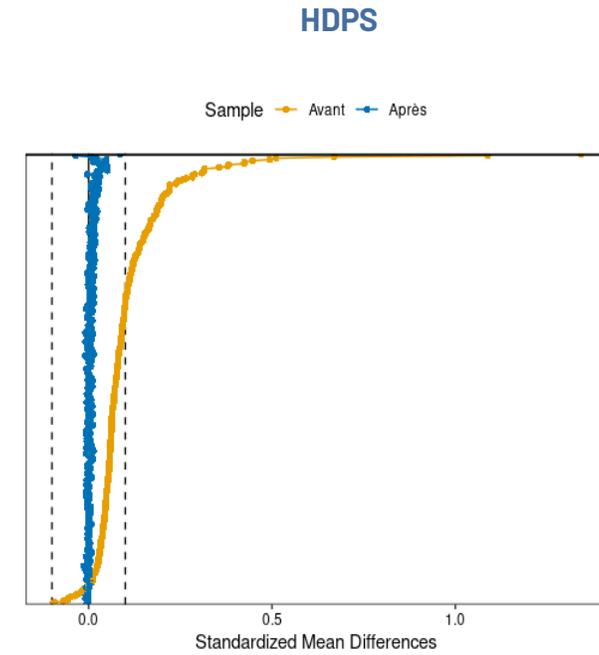


# Performance of covariates balance

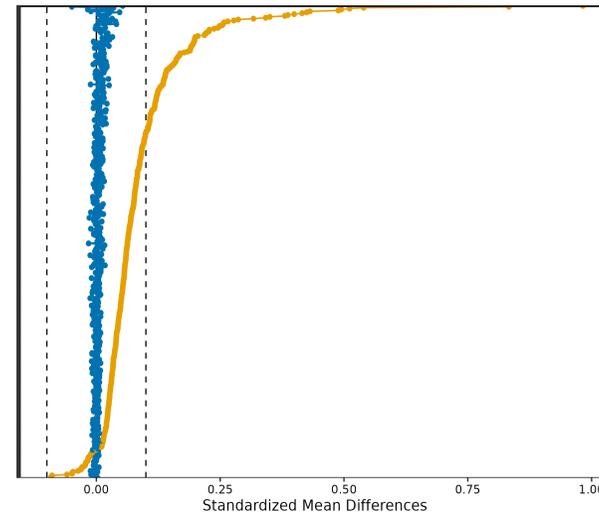
## Plots of standardized differences



K = 100



K = 500



## Variation of number of covariates (k)

Beyond 100 covariables the addition of covariates in the HDPS does not modify the HR

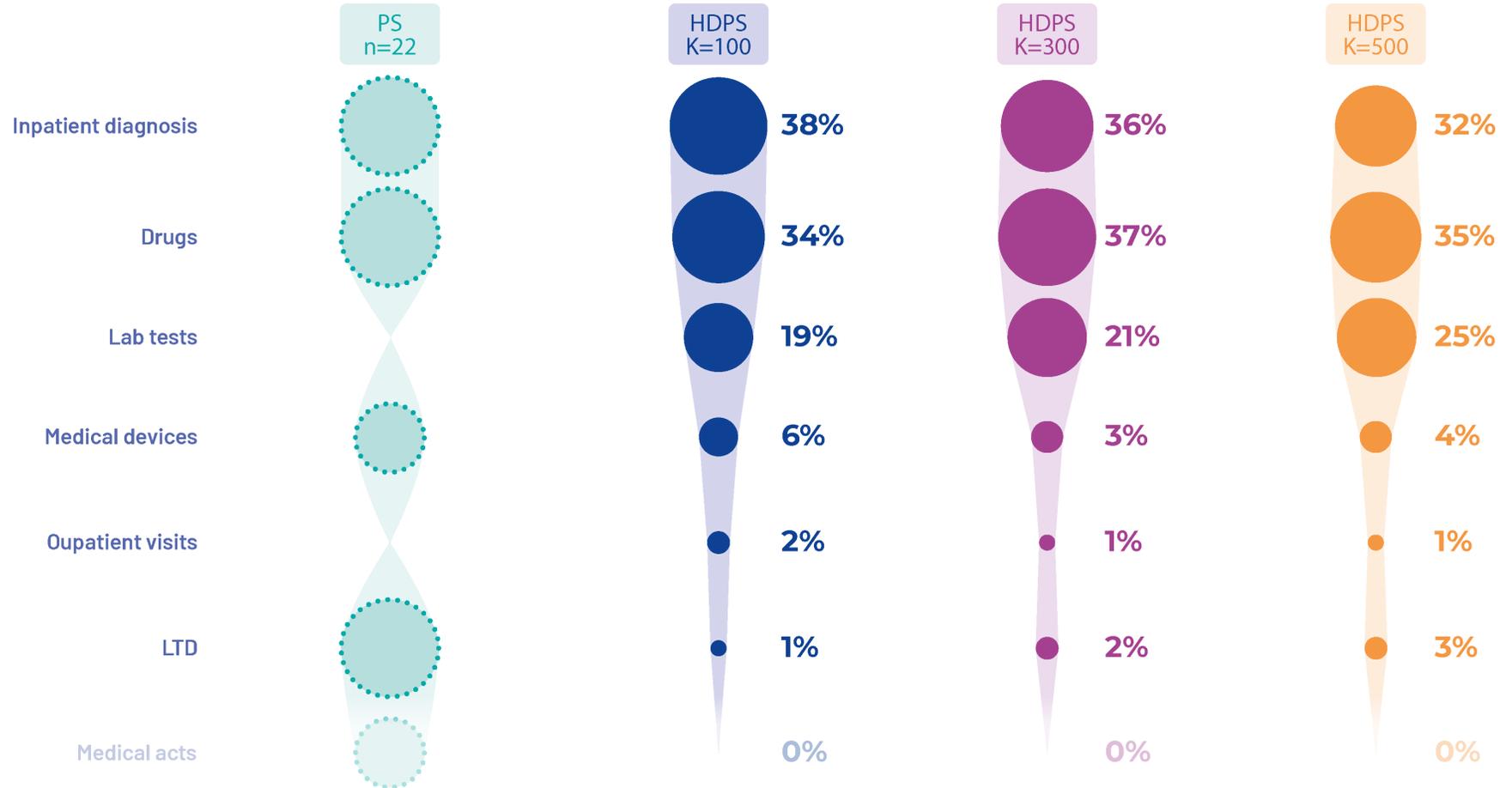


# Results

## Repartition of selected covariates

Dimensions of covariables selected in PS and HDPS are different

Addition of covariables marginally modify the repartition of covariables



- **Mitigates unmeasured confounding in real-world studies comparing the effectiveness or safety of interventions**
- **Automatic selection of variables (as per algorithm)**
- **Optimizes the possibility of taking into account all the information available in the database (even indirect variables)**
- **Choice in construction of HDPS may impact the results**
  - Choice of number of dimensions – 5-7 dimensions<sup>1</sup>
  - Duration of look back period – 12 months<sup>2</sup>
  - Number of prevalent variables – 200<sup>3</sup>
  - Number of selected covariables – 500<sup>4</sup>
- **Multicollinearity and, consequently, increase in variance and risk of overfitting can't be excluded<sup>4</sup>**
- **More transparency in reporting is needed in particular pertaining to methodological choices and sensitivity analyses to increase reproducibility of Hdps analyses**

1- Rassen JA, Blin P, Kloss S, Neugebauer RS, Platt RW, Pottegård A, Schneeweiss S, Toh S. High-dimensional propensity scores for empirical covariate selection in secondary database studies: Planning, implementation, and reporting. *Pharmacoepidemiol Drug Saf.* 2023 Feb;32(2):93-106. doi: 10.1002/pds.5566. Epub 2022 Nov 22. PMID: 36349471; PMCID: PMC10099872.

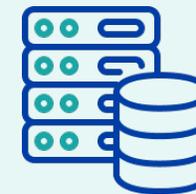
2 - Toh S, Garcia Rodriguez LA, Hernán MA. Confounding adjustment via a semi-automated high-dimensional propensity score algorithm: an application to electronic medical records. *Pharmacoepidemiol Drug Saf.* 2011 Aug;20(8):849-57.

3 Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology.* 2009 Jul;20(4):512-22.

4- Tazare J, Wyss R, Franklin JM, Smeeth L, Evans SJW, Wang SV, et al. Transparency of high-dimensional propensity score analyses: Guidance for diagnostics and reporting. *Pharmacoepidemiol Drug Saf.* 2022 Apr;31(4):411-23.



**HDPS matching produced results comparable to those obtained by PS, despite a different selection of covariates (including an increased importance of biological tests).**



**The HDPS allowed an adjustment on additional covariates not retained by the committee of experts in PS, suggesting that the HDPS could be a valuable tool for sensitivity analyses, to confirm results obtained with more conventional methods.**

**Thank you for your attention !**