DIABETES EPIDEMIOLOGY IN THE BIG DATA ERA:

SURVEILLANCE BASED ON HEALTH ADMINISTRATIVE DATABASES

Sonsoles Fuentes, PhD Student
13th WHO/IDF Cambridge Diabetes Seminar
31st March to 6th April 2019

- **Data sources for diabetes epidemiology:**

  - Cross-sectional surveys

  - Cohort studies

  - … **<u>BIG DATA</u>** : 3V (Volume, variety and velocity)

    → Health Administrative Databases

- ***Système National des Données de Santé* (SNDS)**

  - Data on ~ 66 million people (99% of French population)

  - Recorded individual and anonymous data from:

    − Reimbursement of out-of-hospital dispensed health care (drugs, physician visits,….)
    − Public and private hospitals
    − Death certificates

Objective: To improve diabetes surveillance based on SNDS in France

- **SO 1: to validate diabetes case definition algorithms already used in France**

  - 3 algorithms → Good performance:
    - ↑ Specificity, ↑ Sensibility and ↑ Kappa
    - To identify known and pharmacologically treated diabetes

- **SO 2: to assess trends on diabetes prevalence and incidence between 2010 and 2017 in France**

  - Type 2 diabetes among adults aged 45 years or more:
    - Prevalence→ increasing annual time trends
    - Incidence → decreasing annual time trends

- **SO 3: to develop an algorithm to identify type I diabetes cases**

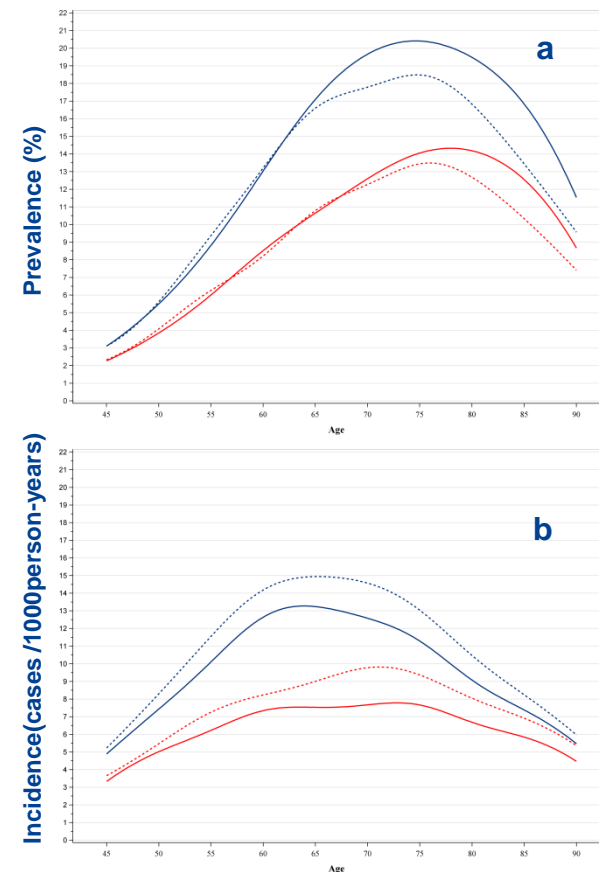 - Machine-learning methodology (without *a priori*)

Figure (a) Age-specific prevalence in 2010 (dot line) and in 2017 (solid line) and (b) aged-specific incidence in 2012 (dot line) and in 2017 (solid line) in France among adults aged 45 years or more, by sex (men in blue and women in red)

- **Opportunities**

  - To overcome limitations of "old sources":

    - Selection bias, memory bias ….

    - Lack of statistical power due to sample size

- **Challenges**

  - To develop new methodologies to exploit these sources

  - To review ethical standards in order to avoid misuse of information