

*SUCCÈS ET ÉCHECS EN RECHERCHE CLINIQUE*

Controverse

Les cohortes au temps des big data

Sigismond LASOCKI vs. Pierre INGRAND

Journée du GIRCI – HUGO

29 juin 2021

# Cohorte prospective = une entreprise risquée

- Niveaux de complexité inégaux, liés aux objectifs
  - Grandes cohortes populationnelles : *Constances, I-Share* => logistique +++
  - Cohortes pronostiques => données de suivi
  - Enquêtes d'évaluation => données de prise en charge
- Des difficultés opérationnelles => validité finale ?
  - Contraintes liées à la durée du suivi
  - Maintien du financement sur la durée (dépassé celle des AAP)
  - Ressources humaines (Sauvadet)
  - Dynamique des participants
  - Données manquantes

# Efficiency and validity

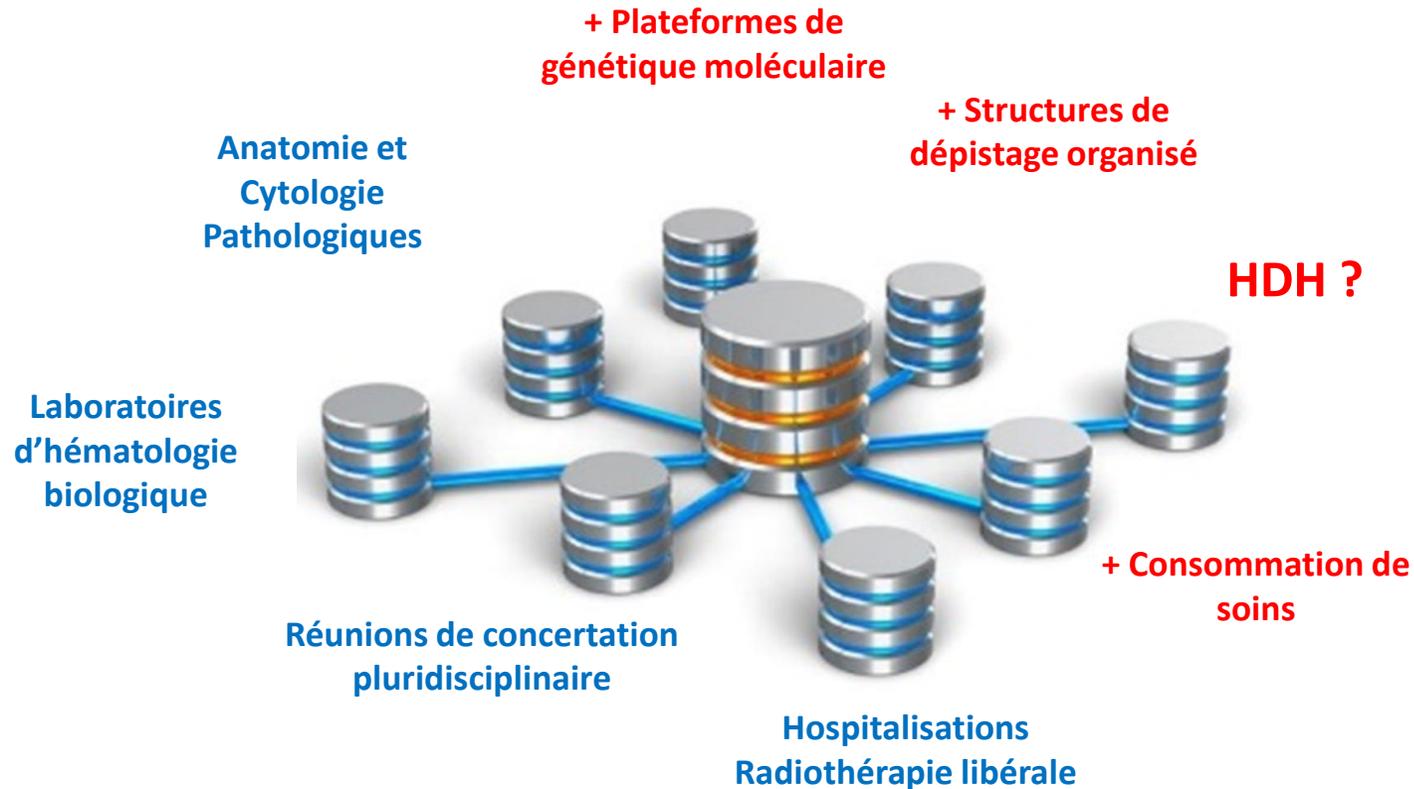
- Efficiency
  - Optimal precision => conditions of effectiveness
  - Minimization of costs => acquisition of information
- Validity : control of biases
  - Selection (representativeness, comparability)
  - Information => outcomes / exposures (classification bias)
  - Confusion
- Kenneth ROTHMAN : the best way to reduce costs is to use existing information sources
  - Surveillance systems (morbidity registers)
  - Historical cohorts
  - Reference population data

# Données de santé réutilisables pour la recherche

- Contexte législatif : caractère patrimonial commun des données de santé => disponibilité croissante
- Fort développement : SNDS, HDH, plateformes et entrepôts de données (Inca, hospitalières, imagerie, optique), registres
- Trois « V » des big data : Volume, Vitesse, Variété (un 4e : Véracité ?)
- Outils performants de fouille de données textuelles, et d'intelligence artificielle (machine learning)
- Appariements déterministes / probabilistes

# Registre des cancers de Poitou-Charentes (G Defossez, directeur)

*Un exemple de système d'information intégratif pour la recherche sur les parcours en vie réelle*



- **Nombreux domaines d'application**
  - Prévention
  - Dépistage
  - Diagnostic
  - Traitement
  - Suivi, seconds cancers
- **Evaluation des parcours en « vie réelle »**
- **Recherche sur les services de santé (Health Services Research)**

# Cadre méthodologique

- **Cohorte rétrospective** (historique) vs prospective (Rothman)
- Un concept ancien (santé travail) sujet à de récents développements
- Aspect essentiel : la qualité de l'information, et en premier lieu sa disponibilité
- De-contextualisation, re-contextualisation, et réutilisation (Leonelli)
- Réserver des ressources pour compléter les infos manquantes
- Approche ambidirectionnelle (Kleinbaum) : part rétrospective, part prospective
- Etudes de vie réelle, Evénements rares (effectifs)

# Contrepoints

- Préalable : disposer d'une connaissance approfondie des sources concernées (populations, règles d'enregistrement, nature des informations, fiabilité)
- Fausses cohortes (approches transversales)
- Axiome de temporalité « ***La cause précède toujours l'effet*** »
- Distinguer corrélation de causalité
- Distinguer exploratoire (génération d'hypothèses) vs confirmatoire
- Ne pas confondre qualité et quantité