# Contenu et qualité des données des projets de recueil automatisé des dossiers médicaux de soins primaires : Revue systématique de la littérature.

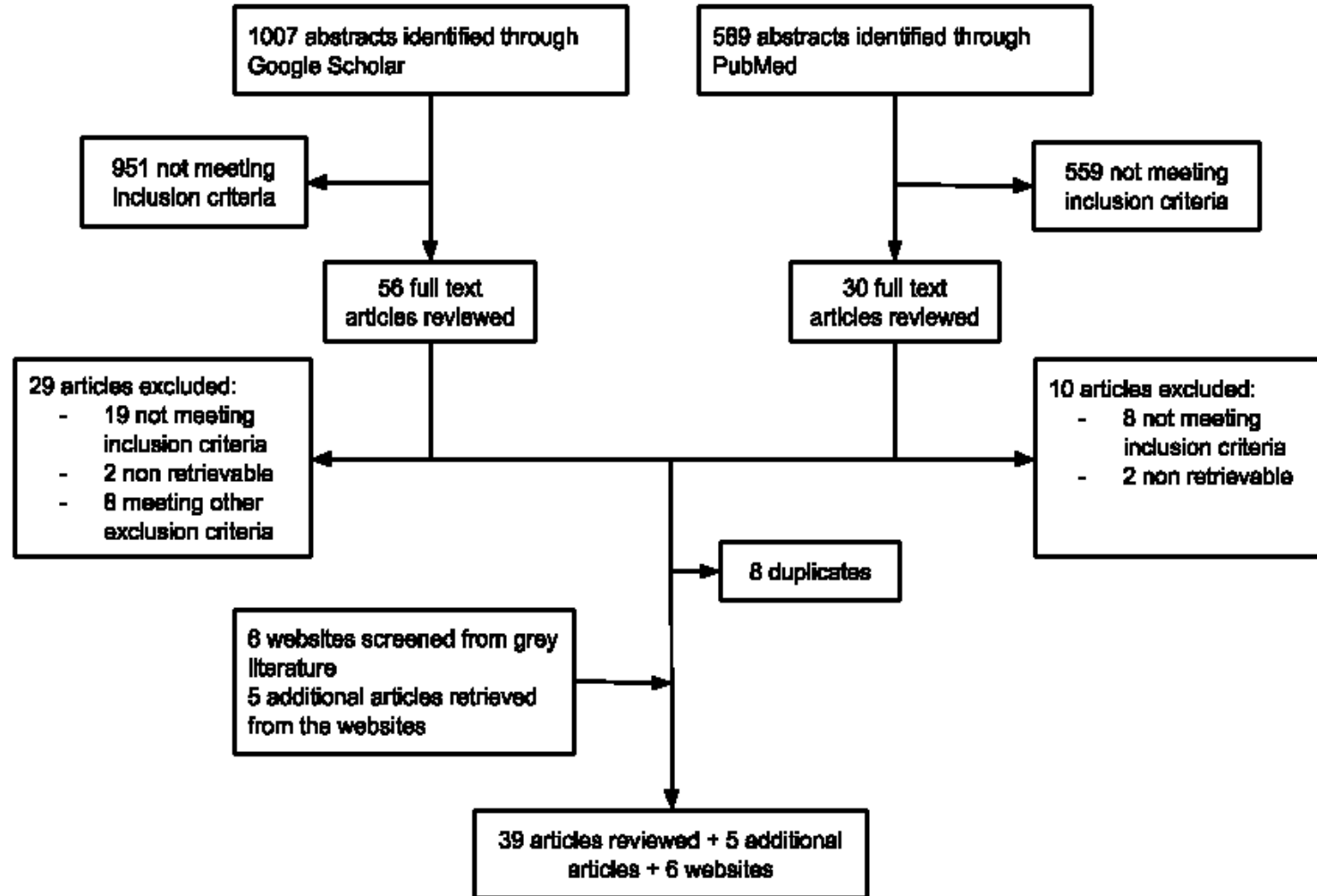Dr Camille Hagenbourger

Dr Marie-Line Gentil

# Introduction

- Dossiers médicaux informatisés
  - Synthèse parcours du patient
  - Population représentative soins primaires
- Factors influencing the development of primary care data collection projects from electronic health records: a systematic review of the literature. *BMC Medical Informatics and Decision Making*
  - 2 axes à approfondir
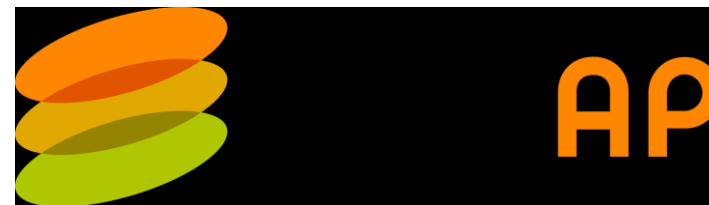    - Confidentialité
    - Contenu et qualité

# Méthode

- Revue systématique de la littérature
  - PRISMA, grille AMSTAR2
  - PubMed et Google Scholar
  - Double sélection, double extraction sur formulaire
- Sources complémentaires
  - Contact des projets
  - Sites internets des projets identifiés

18e CONGRÈS NATIONAL 2018
CNGE Collège Académique

Tours
Vinci

# Résultats : flow chart



1007 abstracts identified through Google Scholar

589 abstracts identified through PubMed

951 not meeting inclusion criteria

559 not meeting inclusion criteria

56 full text articles reviewed

30 full text articles reviewed

29 articles excluded:
- 19 not meeting inclusion criteria
- 2 non retrievable
- 8 meeting other exclusion criteria

10 articles excluded:
- 8 not meeting inclusion criteria
- 2 non retrievable

8 duplicates

6 websites screened from grey literature
5 additional articles retrieved from the websites

39 articles reviewed + 5 additional articles + 6 websites

# Résultats : 6 projets

18e CONGRÈS NATIONAL 2018
CNGE Collège Académique
Tours Vinci

# Résultats : Fond

| | CPRD | FIRE | CPCSSN | SIDIAP | Intego | NIVEL |
|---|---|---|---|---|---|---|
| **Date** | system date duration of consultation date of clinical event date of data entry | contact date | encounter date | Date of diagnosis Expiry date | Data registration date Diagnosis date | Unknown |
| **Vital signs** | ✔ | Systolic blood pressure diastolic blood pressure pulse height weight waist circumference | Blood pressure Height Weight BMI | ✔ | blood pressure pulse weight height | ✘ |
| **Reason for consultation** | entity type specific clinical details related to that entity (clinical history, diagnosis, signs and symptoms) | ✘ | ✔ | ✘ | ✘ | ✔ |
| **Diagnosis** | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ |
| **Clinical history** | ✔ | ✘ | ✔ | ✔ | not specified | ✔ |
| **Lifestyle factors** | Smoking and alcohol status | ✘ | Smoking frequency, duration, end | Smoking and alcohol status | Smoking | ✘ |

18ᵉ CONGRÈS NATIONAL 2018
CNGE Collège Académique
Tours Vinci

# Résultats : Forme

- Liens avec thésaurus
- Classifications communes
  - ATC
  - ICPC-2

| | CPRD | FIRE | CPCSSN | SIDIAP | Intego | NIVEL |
|---|---|---|---|---|---|---|
| ICPC-2 | ✓ | ✓ | | | ✓ | (ICPC-1) |
| ICD 9 | | | ✓ | | | |
| ICD 10 | | | | ✓ | ✓ | |
| Thesaurus linked with ICPC or ICD | Read codes | ✗ | ✗ | ✓ | ✓ | ✗ |

- Emergence du texte libre

18é CONGRÈS NATIONAL 2018
CNGE Collège Académique
Tours Vinci

# Résultats : Qualité des données

- Définition complexe et variable
  - Exhaustivité et exactitude
- Contrôles internes
- Validation externe

18e CONGRÈS NATIONAL 2018
CNGE Collège Académique
Tours Vinci

# Discussion

- Force méthodologique
  - Double lecture et double analyse
  - Sélection des projets
  - Collaboration avant publication

- Nombreux items communs
  - Transposition
  - Collaboration

- Données majoritairement codées
  - Limites
  - Emergence du texte libre

18e CONGRÈS NATIONAL 2018
CNGE Collège Académique
Tours Vinci

# Discussion

- Qualité des données
  - Difficile à définir
  - Difficile à évaluer

- Amélioration
  - Formation des médecins
    - Data reuse
    - Feed-backs
  - Alternative
    - Progrès du TAL
    - Fouille de données massives

18e CONGRÈS NATIONAL 2018
CNGE Collège Académique

Tours Vinci

# Conclusion

- Connaissance approfondie des projets internationaux
- Points communs sur le fond
  - Collaboration
- Place du texte libre
  - TAL développé par techniques IA
- Perspectives

18e CONGRÈS NATIONAL 2018
CNGE Collège Académique
Tours Vinci

# Merci de votre attention

# Annexes - équations de recherche

**Pubmed query**

("Electronic Health Records"[Mesh] OR "Medical records systems, computerized"[Mesh] OR "Electronic Medical Record"[All Fields])  →  EHR synonyms

**AND**

("Health Information Management"[Mesh] OR "Information Storage and Retrieval"[Mesh] OR "Natural Language Processing"[Mesh] OR "Automatic Data Processing"[Mesh] OR "Algorithms"[Mesh] OR "Clinical Coding"[Mesh] OR "International Classification of Diseases"[Mesh] OR "icpc" or "read code" OR "Data Accuracy"[Mesh] OR "Quality Control"[Mesh])  →  Database and data quality synonyms

**AND**

("General Practice"[Mesh] OR "General Practitioners"[Mesh] OR "Primary Health Care"[Mesh] OR "Physicians, Primary Care"[Mesh] OR "Ambulatory care"[Mesh] OR "Ambulatory Care Facilities" [Mesh] OR "Primary care"[All Fields])  →  Primary care synonyms

**AND**

("2010/01/01"[PDAT]: "2018/12/31"[PDAT])

**Google Scholar queries**

"data mining" OR "text mining" OR "text analysis" OR "text notes" OR "clinical notes" OR "Natural Language Processing" OR "Clinical Coding" OR "structured data" OR "code list" OR "diagnostic codes" "primary care database"  →  database content and form of the data

"quality assessment" OR "data quality" OR "data accuracy" OR "data reliability" OR "data validity" OR "data consistency" OR "data completeness" OR "data correctness" OR "data availability" "primary care database"  →  data quality

16

# Annexes - demographics (1)

| Country | Project name | Number of patients (% population)[a] | Number of GPs (other healthcare measurement) |
|---|---|---|---|
| UK | CPRD | 13 million (20%) | 600 GPs |
| | THIN | 11.1 million (17%) | 562 GPs |
| | Qresearch | 18 million (28%) | 1000 GPs |
| USA | Veterans database | 17 million (5%) | GPs unknown (53,000 healthcare professionals) |
| | Geisinger© Health System project | 400,000 (0.1%) | GPs unknown (41 community practice clinics) |
| | General Electric Centricity© | 8.9 million (2.8%) | 7259 GPs |
| | OCHIN, Inc | 1 million patients annually (0.3%) | GPs unknown (200 clinic sites) |
| | Baylor Health Care System (BHCS) and Christiana Care Health System (CCHS) | 680,000 patients (0.2%) | GPs unknown (33 primary care practices + 35 office practices) |
| | University of California Davis Health System's electronic medical record system | unknown | GPs unknown (13 clinics in this analysis) |
| | CHCI – Connecticut | 130,000 (0.04%) | GPs unknown (13 primary health centers) |
| | MGPO | 87,568 patients in this study (0.02%) | 148 GPs in this study |
| | OpenDoor | 40,000 patients annually (0.01%) | GPs unknown (4-site federally qualified community health center in New York) |
| | New York City Primary Care Information Project | unknown | GPs unknown (82 practices) |
| | IFH | 90,000 patients (0.02%) | GPs unknown (17 community health centers) |
| | Department of Family Medicine – Wisconsin | 110,000 for these studies (0.03%) | GPs unknown (28 primary care clinics) |

#CNGE

018
ue

Tours
Vinci

# Annexes - demographics (2)

| | | | |
|---|---|---|---|
| Canada | CPCSSN | 1.28 million (3.6%) | 1031 GPs |
| | EMRALD | 0.5 million (1.4%) | 350 GPs |
| | DELPHI | 30,000 (0.08%) | 30 GPs |
| Netherlands | IPCI | 1.5 million (8.8%) | 600 GPs |
| | Nivel-PCD | 1.6 million (10%) | GPs unknown (500 practices) |
| | LINH | 350,000 (2%) | 85–150 GPs |
| | Julius | 156,176 (0.9%) | 60 GPs |
| | RNH | 51,700 (0.3%) | 65 GPs (22 practices) |
| Spain | SIDIAP | 6 million (12.8%) | 274 GPs |
| | ECRPC | unknown | GPs unknown |
| | BIFAP | 3.2–4.8 million (6.8–10%) | 1183 GPs |
| Australia | GRAPHC | 32,000 (0.1%) | GPs unknown (1 GP practice) |
| | Canning Division of General Practice | unknown | GPs unknown (8 GP sites) |
| | MEGPN | 505,600 (2.1%) | 578 GPs (141 practices) |
| Belgium | Intego | 285,000 (2.5%) | GPs unknown (95 GP practices) |
| | NIHDI | 29,000 (0.25%) | GPs unknown (5000 practices) |
| France | LPD – Cegedim | 1.5 million (2%) | 1200 GPs (+ 750 various medical specialists) |
| Italy | Health Search Database / CSD© | 2.5 million (4.1%) | 900 GPs |
| Sweden | Swedestar project | unknown | GPs unknown |
| Switzerland | FIRE | 113,335–293,000 (14–36%) | 60–120 GPs |

[a]Percentage denominator (i.e., the country population was extracted from national official websites)

# Méthode : formulaire

- Content of the database
  - Administrative data
  - Clinical data
  - Prescription
  - External data
- Form
  - Form of the data : coded, structured, free-text
  - Coding system
- Quality
  - Méthods used to be sure that data are fit for research
    - Internal checks
    - External validation