

Big Data et Cardiopathies Congénitales

40^{ème} Séminaire de Cardiologie Pédiatrique

Sarah Cohen



Marie Lannelongue

« Big Data is like teenage sex:
everyone talks about it,
nobody really knows how to do it,
everyone thinks everyone else is doing it,
so everyone claims they are doing it... »

Dan Ariely, Duke University, 2013



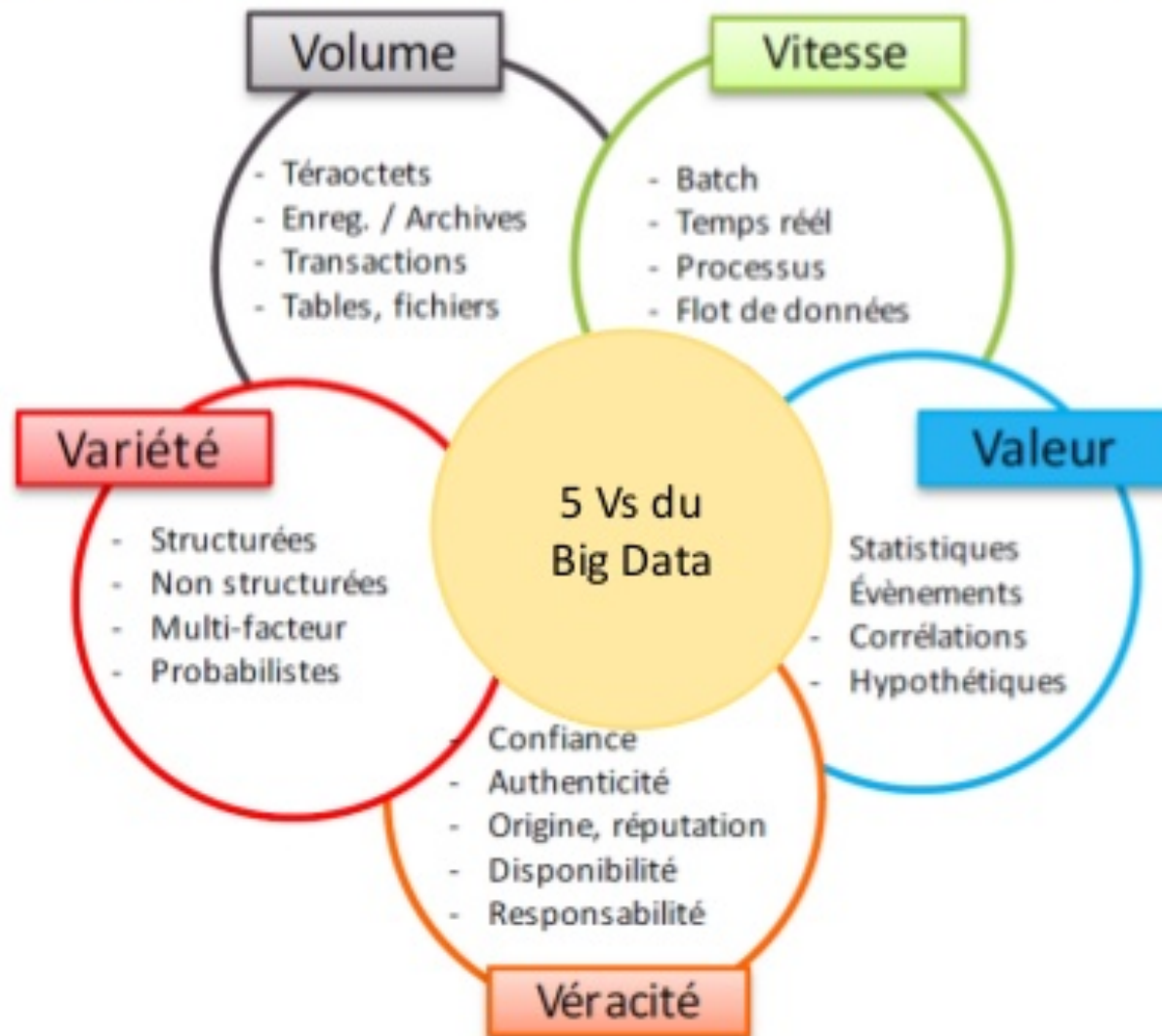
Marie Lannelongue



Qu'est ce que le « big data »?

Le big data, parfois appelées **données massives**, désigne des ensembles de données devenus si volumineux qu'ils **dépassent l'intuition et les capacités humaines d'analyse et même celles des outils informatiques classiques** de gestion de base de données ou de l'information.





Data never sleeps!

Diversité et volume des sources de données

2018 *This Is What Happens In An Internet Minute*



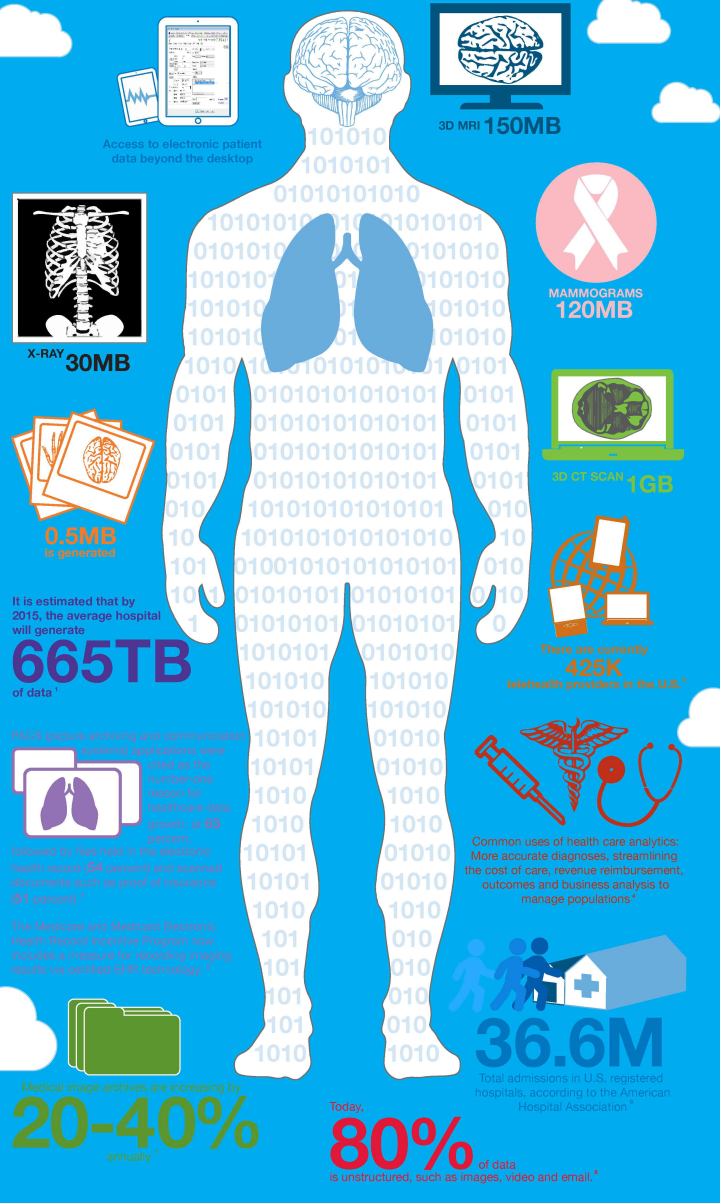
Marie Lannelongue



Created By:
@LoriLewis
@OfficiallyChadd

The Body as a Source of Big Data

Today data storage is essential for healthcare providers to see a patient's complete story of care, make the most informed decisions and enhance treatment and outcomes.



Complexité et abondance croissantes des données de santé

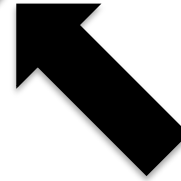
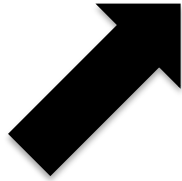
- Généralisation des **dossiers médicaux électroniques**
- Développement de **nouvelles technologies** (applications mobiles etc)
- Collecte d'**informations génomiques** devenue moins chère
- **Communications sociales** des patients sous forme numérique en augmentation
- **Accumulation de connaissances médicales**

Données cliniques

- Données structurées des DME
- Données non structurées des DME
- Imageries médicales



Données de Santé



Données génomiques

- Séquences d'ADN

Données comportementales

- Données des réseaux sociaux
- Données de localisation
- Données de consommation



Données cliniques

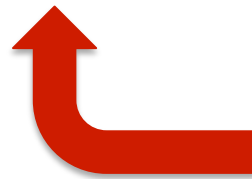
- Données structurées des DME
- Données non structurées des DME
- Imageries médicales

Données structurées (codées)

saisies dans un format structuré (informations démographiques, codes diagnostiques de facturation, actes médicaux)

Données non structurées (narratives)

stockées sous forme de texte libre dans les rapports médicaux



Méthodes de **traitement automatique
du langage naturel (NLP)**



Bases de données médico-administratives (BDMA)

- Stockage de données à des fins de **déclaration, d'enregistrement, de facturation ou de remboursement**
- **Collecte des données** : Informations recueillies durant les contacts des patients avec le système de soins



Analyse secondaire des données médico-administratives à des fins de recherche

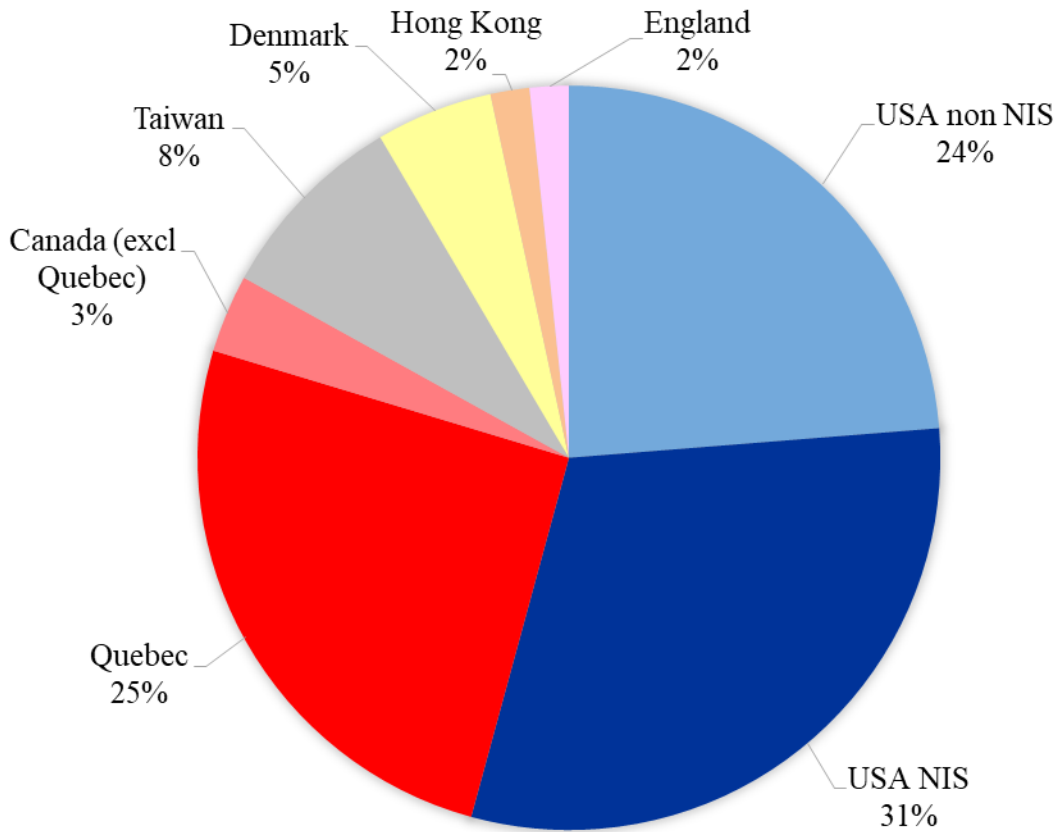
- Définition : analyse secondaire de données à des fins allant **au-delà de l'intention initiale**
- **Domaines de recherche** : pharmacoépidémiologie¹, surveillance des maladies², enquêtes épidémiologiques, analyse économique des dépenses et de la consommation de soins³



Marie Lannelongue

1. Weill A, et al. *Pharmacoepidemiol Drug Saf.* 2010
2. Gershon AS, et al. *Am J Epidemiol.* 2010;
3. Mackie AS, et al, *Can J Cardiol.* 2017

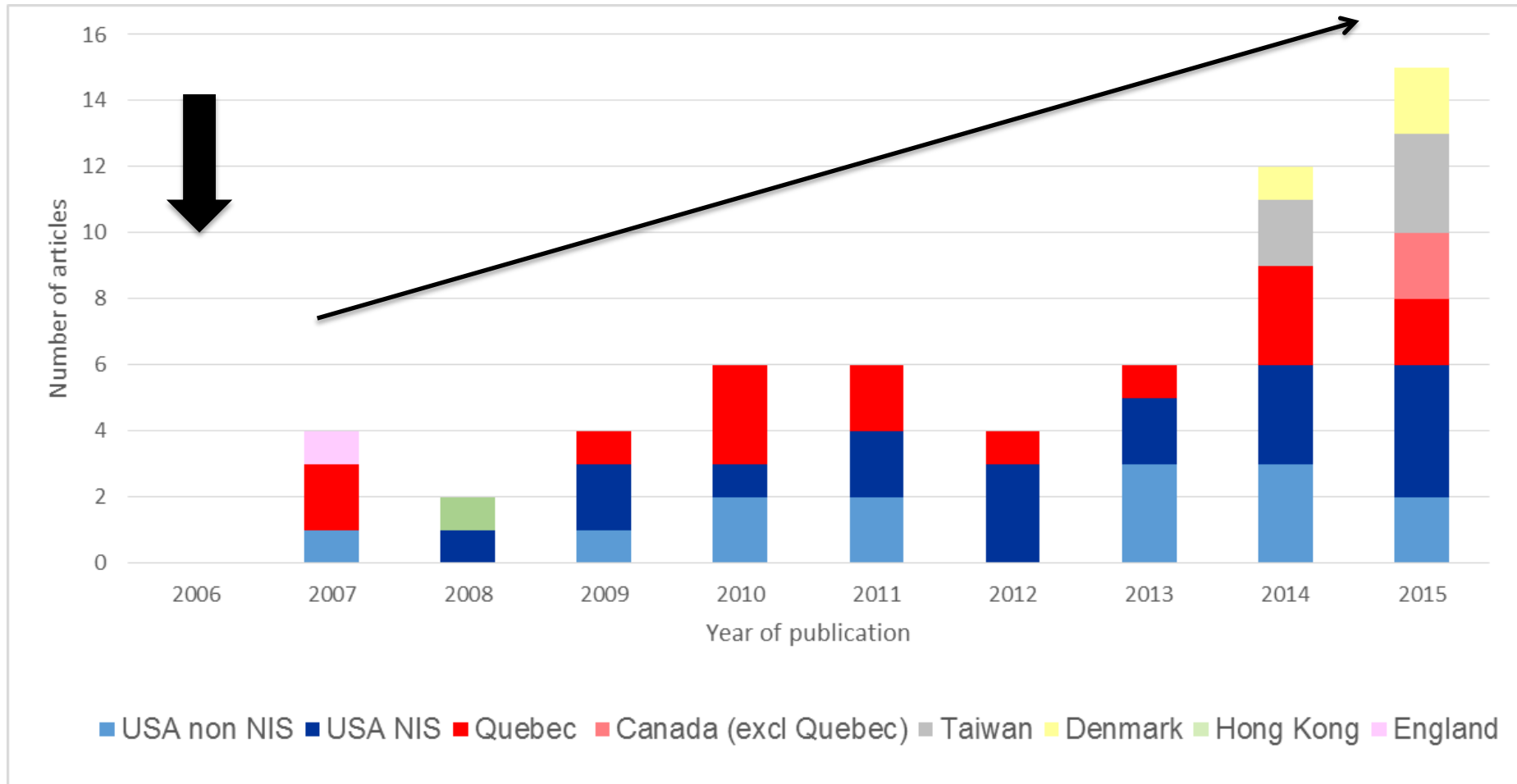
Généralisation de l'utilisation des BDMA en Cardiologie Congénitale Adulte



- 59 études
- 12 sources de données
- 6 pays

*NIS =Nationwide Inpatient Sample

Utilisation croissante des BDMA en cardiologie congénitale adulte



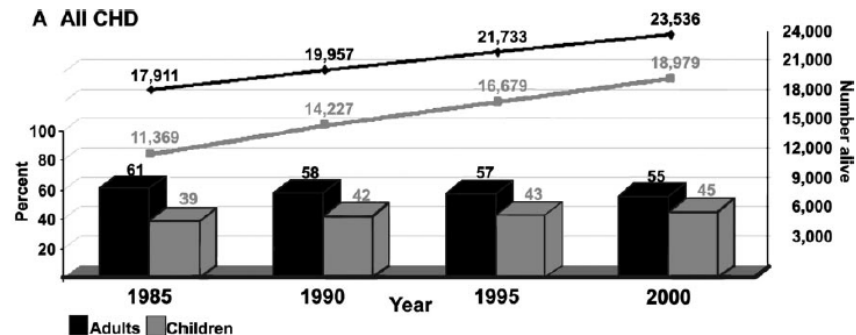
Méthodologie commune, objectifs divers

Prévalence

Congenital Heart Disease

Congenital Heart Disease in the General Population Changing Prevalence and Age Distribution

Ariane J. Marelli, MD; Andrew S. Mackie, MD, SM; Raluca Ionescu-Ittu, MSc;
Elham Rahme, PhD; Louise Pilote, MD, MPH, PhD



Méthodologie commune, objectifs divers

Mortalité

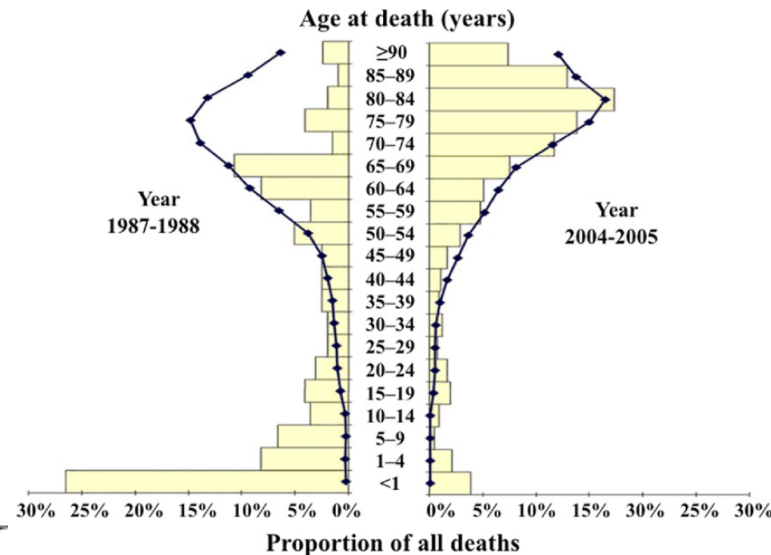
Journal of the American College of Cardiology
© 2010 by the American College of Cardiology Foundation
Published by Elsevier Inc.

Vol. 56, No. 14, 2010
ISSN 0735-1097/\$36.00
doi:10.1016/j.jacc.2010.03.085

CONGENITAL HEART DISEASE

Changing Mortality in Congenital Heart Disease

Paul Khairy, MD, PhD,* Raluca Ionescu-Ittu, MSC,†§ Andrew S. Mackie, MD, SM,†
Michal Abrahamowicz, PhD,§ Louise Pilote, MD, MPH, PhD,‡§ Ariane J. Marelli, MD†



Marie Lannelongue

Méthodologie commune, objectifs divers

Coût

Inpatient admissions and costs of congenital heart disease from adolescence to young adulthood



Yang Lu, PhD,^a Garima Agrawal, MPH,^b Chia-Wei Lin, MS,^b and Roberta G. Williams, MD^c *Torrance, and Los Angeles, CA*



Marie Lannelongue



Méthodologie commune, objectifs divers

Association

Circulation

ORIGINAL RESEARCH ARTICLE

Exposure to Low-Dose Ionizing Radiation From Cardiac Procedures and Malignancy Risk in Adults With Congenital Heart Disease

Editorial, see p 1346

BACKGROUND: Adults with congenital heart disease (CHD) are exposed to increasing amounts of low-dose ionizing radiation (LDIR) from cardiac procedures. Cancer prevalence in this population is higher than in the general population. This study estimates the association between LDIR exposure from cardiac procedures and incident cancer in adult patients with CHD.

METHODS: The study population derived from the Quebec Congenital Heart Disease Database. We measured cumulative numbers of LDIR-related cardiac procedures for each patient until 1 year before the time of cancer diagnosis or

Sarah Cohen, MD, MSc
Aihua Liu, PhD
Michelle Gurvitz, MD, MSc
Liming Guo, MSc
Judith Therrien, MD
Claudie Laprise, PhD
Jay S. Kaufman, PhD
Michal Abrahamowicz,
PhD
Ariane J. Marelli, MD, MPH



Méthodologie commune, objectifs divers

Outcomes

Perioperative Outcomes of Major Noncardiac Surgery in Adults with Congenital Heart Disease

Bryan G. Maxwell, M.D., M.P.H.,* Jim K. Wong, M.D.,† Cindy Kin, M.D.,‡
Robert L. Lobato, M.D., M.S.§

Device complications in adult congenital heart disease

Robert M. Hayward, MD,* Thomas A. Dewland, MD,* Brian Moyers, MD,*
Eric Vittinghoff, PhD,† Ronn E. Tanel, MD, FHRS,‡ Gregory M. Marcus, MD, MAS, FHRS,*
Zian H. Tseng, MD, MAS, FHRS*



Marie Lannelongue



Méthodologie commune, objectifs divers

Utilisation des ressources de santé

Decentralization of Care for Adults with Congenital Heart Disease in the United States: A Geographic Analysis of Outpatient Surgery

Bryan G. Maxwell^{1*}, Thane G. Maxwell², Jim K. Wong³

¹ Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States of America, ² Department of Social Science, Metropolitan State University, Saint Paul, MN, United States of America, ³ Department of Anesthesiology, Perioperative, and Pain Medicine, Stanford University, Stanford, CA, United States of America

Adults or Big Kids: What Is the Ideal Clinical Environment for Management of Grown-Up Patients With Congenital Heart Disease?

Tara Karamlou, MD, Brian S. Diggs, PhD, Ross M. Ungerleider, MD, MBA, and Karl F. Welke, MD, MS

Divisions of Cardiothoracic Surgery and Surgery, Oregon Health and Science University, Portland, Oregon; Division of Pediatric Cardiothoracic Surgery, Department of Surgery, Case Western Reserve University, Cleveland, Ohio; and Mary Bridge/Swedish Pediatric Cardiothoracic Surgery Program, Mary Bridge Children's Hospital and Health Center, Multicare Health System, Tacoma, Washington

Specialized Adult Congenital Heart Disease Care The Impact of Policy on Mortality

Darren Mylotte, MD; Louise Pilote, MD, MPH, PhD; Raluca Ionescu-Ittu, PhD; Michal Abrahamowicz, PhD; Paul Khairy, MD, PhD; Judith Therrien, MD; Andrew S. Mackie, MD, SM; Ariane Marelli MD, MPH

Trends in Hospitalizations for Adults With Congenital Heart Disease in the U.S.

Alexander R. Opatowsky, MD, MPH, Omar K. Siddiqi, MD, Gary D. Webb, MD

Intérêts des BDMA en Cardiologie Congénitale

- Volume et effectifs importants permettant l'étude de maladies relativement rares
- Exhaustivité des données (données en population)
- Suivi longitudinal avec longues périodes d'observation autorisant l'étude de certaines comorbidités ou complications (extra)cardiaques, de survenue tardive
- Recherche à moindre coût

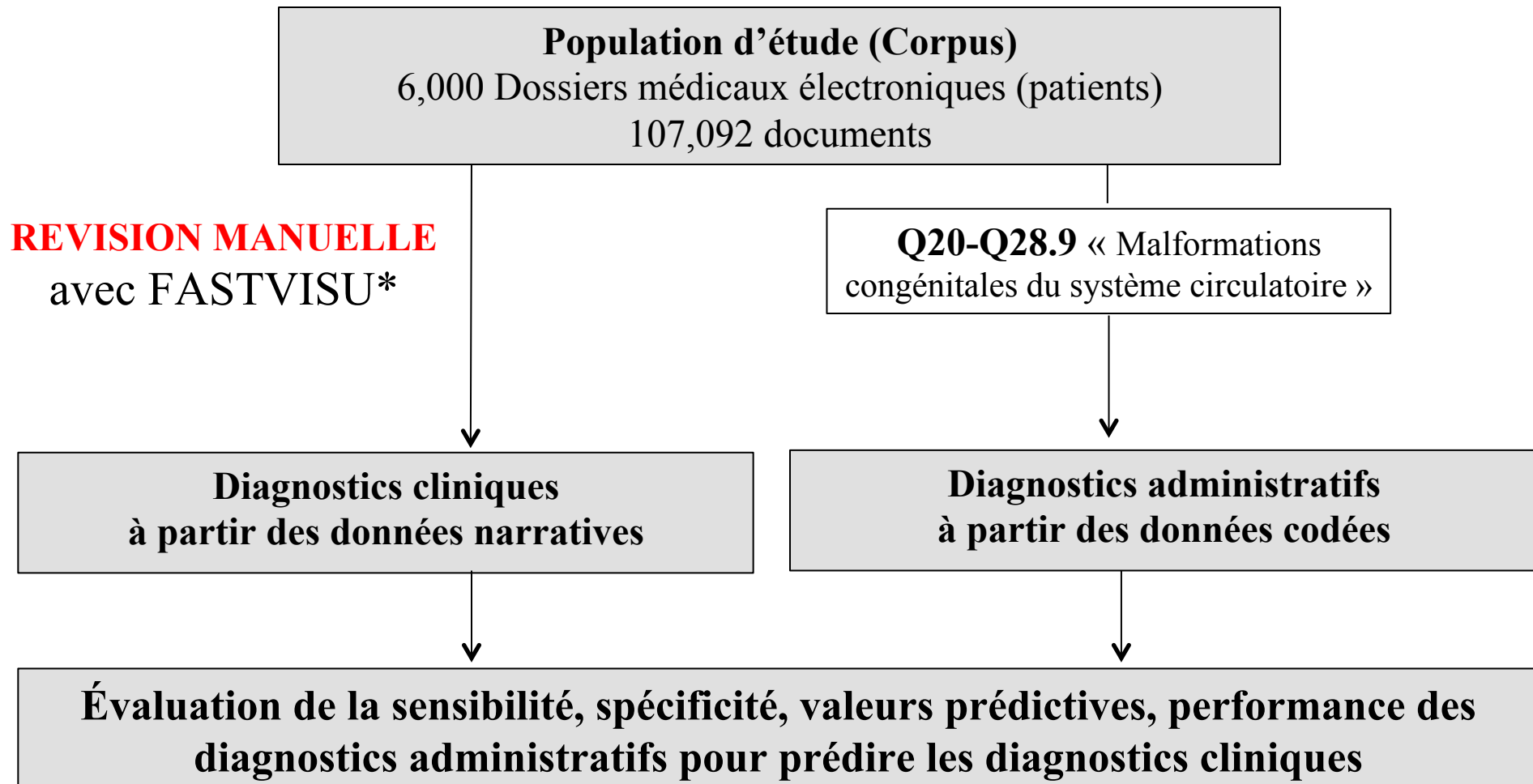


Avoir à l'esprit quelques limites

- Manque de données cliniques
- Facteurs de confusion non mesurés (tabac, alcool, surpoids, données sociales etc.)
- Biais de sélection : Ceux en contact avec le système de soins
- Limites inhérentes au système de santé
- Précision du diagnostic : algorithmes et procédures de validation



Performance de la CIM-10 pour détecter les adultes avec cardiopathies congénitale



Performance de la CIM-10 pour détecter les adultes avec cardiopathies congénitale

Diagnostics administratifs	Diagnostics cliniques			
	ACC	Non ACC	Total	
ACC	629 (VP)	48 (FP)	677	VPP = 0.93 (0.91-0.95)
Non ACC	151 (FN)	5172 (VN)	5323	VPN = 0.97 (0.96-0.98)
Total	780	5220	6000	
	Se = 0.81 (0.78-0.83)	Spe = 0.99 (0.99-1)		Acc = 0.97 (0.96-0.98)

Extraction d'informations : FASTVISU*

Select a serie SC_CC_CC_tranche003 - 2.cardiopathi Select a patient 11) 2004886071 (14) Select a regex catalog 2.cardiopathies congénitales v(R
 Select a poll SC_CC Select allowed vote values OUI/NON Select a voter DG1 (2522) Show categories without match
 << R >>
 - VOTE -

COMPTE RENDU D'USIC

NOM : NOM
 PRENOM : PRENOM
 Née le : DATENAI
 Hospitalisation du 16/06/2004 au 18/06/2004 en USIC puis jusqu'au 23/06/2004 en Cardiol

Fait le 16 juin 2004 Interne : Dr XXXXXX, Dr XXXXXX
 REF : NS CCA/PH : Dr XXXXXX, Dr XXXXXX

Destinataires :
 Dr XXXXXX. Cardiologue. Cardiologie pédiatrique. Necker

MOTIF D'HOSPITALISATION

Patiente âgée de 62 ans hospitalisée pour malaise et dyspnée

ANTECEDENTS

Cardiovasculaires

CAV avec HTAP fixée (syndrome d'Eisenmenger)

Médicaux
 HTA paroxystique
 Maladie d'Alzheimer
 Episodes d'hémoptysie en septembre 2001 et avril 2003
 Oxygène à domicile
 Cyanose chronique

Chirurgicaux
 Kystectomie ovarienne à 18 ans

FACTEURS DE RISQUES CARDIO-VASCULAIRES

Hypertension Artérielle, sous bithérapie, mal équilibrée (Sectral et amlor)
 Hérité Coronarienne (mère hypertendue, IDM à 55 ans)
 Pas de Tabagisme
 Pas de Surcharge pondérale (Poids 50Kg Taille 164cm BMI=18)

MODE DE VIE

Q218 - syndrome d.eisenmenger ☒ Oui ☐ Non

P fixée (syndrome d'Eisenmenger) Médicaux HTA paroxystique Maladie d'alzhei
 ixée et syndrome d'Eisenmenger, traitée depuis quelques années par tildiem 18
 n 2004 Maladie d'Eisenmenger Ventricule droit très hypertrophié, hypocontract
 P fixée (syndrome d'Eisenmenger) Médicaux HTA paroxystique Maladie d'alzhei
 ixée et syndrome d'Eisenmenger, traitée depuis quelques années par tildiem 18
 n 2004 Maladie d'Eisenmenger Ventricule droit très hypertrophié, hypocontract
 P fixée (syndrome d'Eisenmenger) Médicaux HTA paroxystique Maladie d'alzhei
 ixée et syndrome d'Eisenmenger, traitée depuis quelques années par tildiem 18
 n 2004 Maladie d'Eisenmenger Ventricule droit très hypertrophié, hypocontract
 canal artériel, un Eisenmenger et une HTA. Elle a une pression artérielle à 1
 canal artériel, un Eisenmenger et une HTA. Je fais les mêmes constatations qu
 e NOM, qui a un Eisenmenger sur un canal artériel. Elle est sous 125 mg deux
 patiente au stade d'Eisenmenger sur canal artériel. ANTECEDENT et HISTOIRE DE
 patiente au stade d'Eisenmenger sur canal artériel. ANTECEDENT et HISTOIRE DE
 ATION : Syndrome d'Eisenmenger ANTECEDENTS ET CONTEXTE CLINIQUE : Allergies
 tale, Syndrome d'Eisenmenger (canal artériel perméable). Suivie par le Dr Ise
 ATION : Syndrome d'Eisenmenger ANTECEDENTS ET CONTEXTE CLINIQUE : Allergies
 tale, Syndrome d'Eisenmenger (canal artériel perméable). Suivie par le Dr Ise
 ire secondaire -- Eisenmenger - SIGNES FONCTION
 SATION : Syndrome d'Eisenmenger ANTECEDENTS ET CONTEXTE CLINIQUE : Allergies
 tale, Syndrome d'Eisenmenger (canal artériel perméable). Suivie par le Docteu
 SATION : Syndrome d'Eisenmenger ANTECEDENTS ET CONTEXTE CLINIQUE : Allergies
 tale, Syndrome d'Eisenmenger (canal artériel perméable). Suivie par le Docteu
 ur un syndrome d'Eisenmenger. Le patient présente -il des signes d'antécédents
 ely. She had Eisenmenger syndrome le pulmonary hyper
 itale (syndrome d'Eisenmenger). ORIGINE : USIP - HEG
 e d'un syndrome d'Eisenmenger. Décès de la patiente le
 d'Eisenmenger. c'est à dire une hypertension
 d'Eisenmenger. c'est à dire une hypertension

Q221 - stenose congenitale de la valve pulmonaire ☒ Oui ☐ Non

(sible. Absence de sténose pulmonaire. Pas d'épanchement péricardique. EVOLUTION :
 (sible. Absence de sténose pulmonaire. Pas d'épanchement péricardique. EVOLUTION :
 (sible. Absence de sténose pulmonaire. Pas d'épanchement péricardique. EVOLUTION :
 (sible. Absence de sténose pulmonaire. Pas d'épanchement péricardique. EVOLUTION :

Q222 - Insuffisance congénitale de la valve pulmonaire ☐ Oui ☒ Non

e systolique 2/6 d'Insuffisance pulmonaire. Auscultation pulmonaire libre Absence de
 signe
 e systolique 2/6 d'Insuffisance pulmonaire. Auscultation pulmonaire libre Absence de
 signe
 e systolique 2/6 d'Insuffisance pulmonaire. Auscultation pulmonaire libre Absence de
 signe

Niveau d'agrégation pertinent

Filtrage par le traitement de
 texte et mise en évidence des
 regex

Validation du diagnostic par
 système de vote (examen du
 dossier médical)



European Society
of Cardiology

European Heart Journal (2019) 0, 1–9
doi:10.1093/eurheartj/ehy915

CLINICAL RESEARCH

Congenital heart disease

Machine learning algorithms estimating prognosis and guiding therapy in adult congenital heart disease: data from a single tertiary centre including 10 019 patients

**Gerhard-Paul Diller^{1,2,3,4*}, Aleksander Kempny^{1,2}, Sonya V. Babu-Narayan^{1,2},
Marthe Henrichs³, Margarita Brida^{1,5}, Anselm Uebing^{1,6}, Astrid E. Lammers⁶,
Helmut Baumgartner^{3,4}, Wei Li^{1,2}, Stephen J. Wort^{1,2}, Konstantinos Dimopoulos^{1,2},
and Michael A. Gatzoulis^{1,2}**



Marie Lannelongue



„Tetralogy of Fallot. Tetralogy repair 1968. Tricuspid valve replacement (33mm carpentier-edwards valve) 2007. Moderate tricuspid regurgitation. Complete heart block post-operatively requiring DDD pacemaker 2007. Pacemaker generator box change august 2013.“

„Lisinopril 10mg, Thyroxine 150 mcg, Aspirin 75mg, Frusemide 40mg“

“Was delighted to see this gentleman with his father at my clinic today. He remains well and asymptomatic”

*Diagnosis field
(raw text)*

*Medication
(raw text)*

*Symptoms / clinical
status (raw text)*

Convolutional
network

Dense
network

Convolutional
network

Merging layer

**Deep learning
network**

Disease severity
score

Multivariate Cox
survival
model

Predicted survival



Demographic
parameters

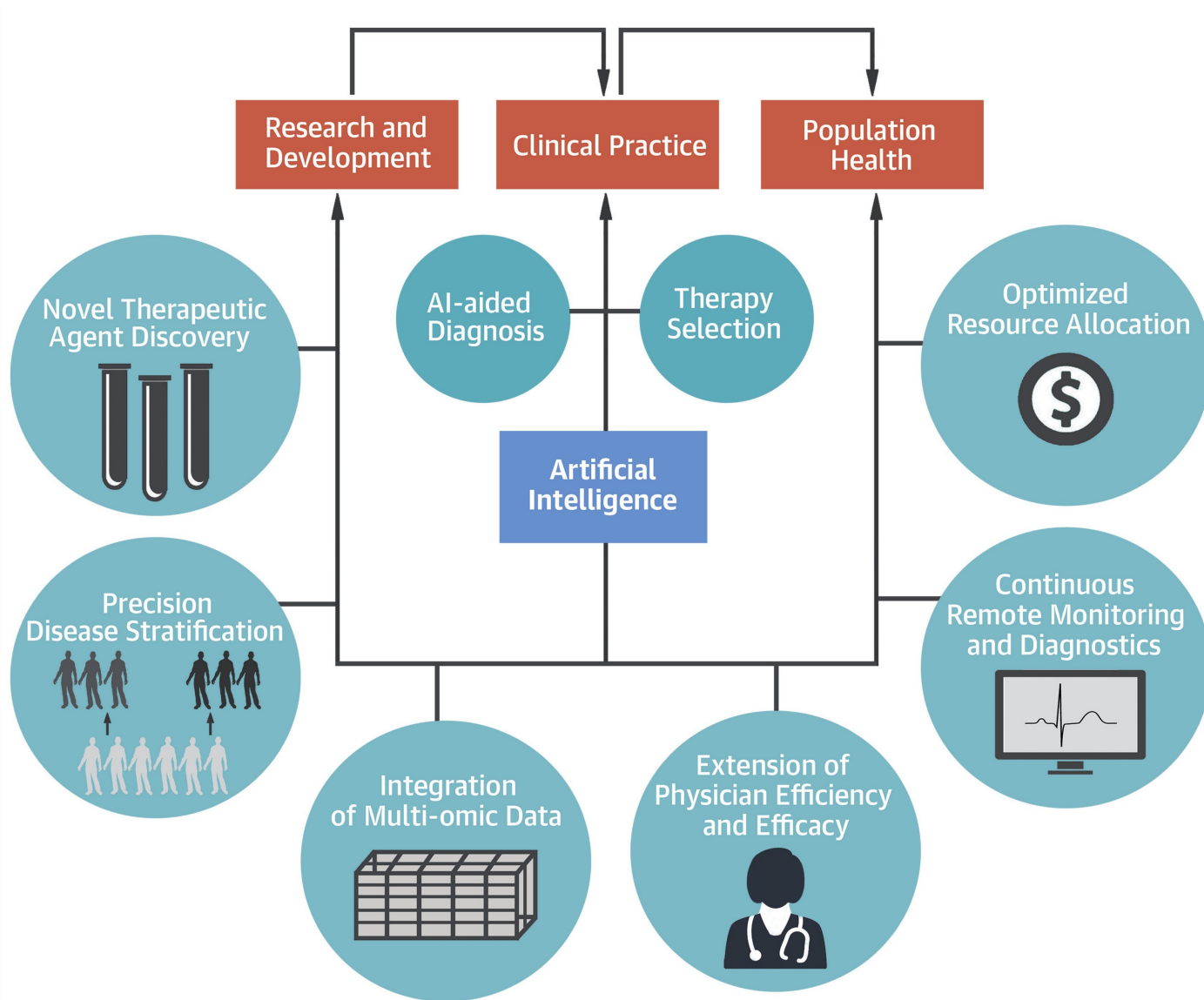
Laboratory
parameters

ECG
parameters

Exercise
parameters



Imaginer demain : la promesse de l'Intelligence Artificielle



With Great Power Comes Great Responsibility

- Nous sommes à l'ère de la **production massive de données**
- L'**utilisation secondaire** de ces données permet et devrait permettre de répondre à certaines questions cliniques et épidémiologiques et faire avancer *l'evidence based medicine*
- A condition de **travailler avec rigueur**, de **connaitre les limites** de nos données, de nos conclusions et de **valider les nouveaux outils**



Marie Lannelongue



Challenge des Big Data en Cardiologie Congénitale

- **Big Data on small numbers** : Déduire des connaissances à partir d'une population hétérogène de patients
- Traiter les **données d'imagerie médicale**
- Capturer les **données comportementales** des patients à travers différents capteurs :
 - Outils numériques et connectés : ordinateurs, smartphones (« téléphones intelligents »)
 - Télésurveillance
 - Réseaux sociaux



L'analyse et la modélisation de ces données fait appel à des technologies d'**Intelligence Artificielle** (Natural Language Processing, Text Mining, Machine Learning, Deep Learning)

MAIS

L'ingéniosité du cardiopédiatre est de faire face à une **cible mouvante**



« Une accumulation de faits n'est pas plus
une science qu'un tas de briques n'est une
maison.[...]
Le scientifique doit ordonner »

H. Poincaré (1908)



Marie Lannelongue

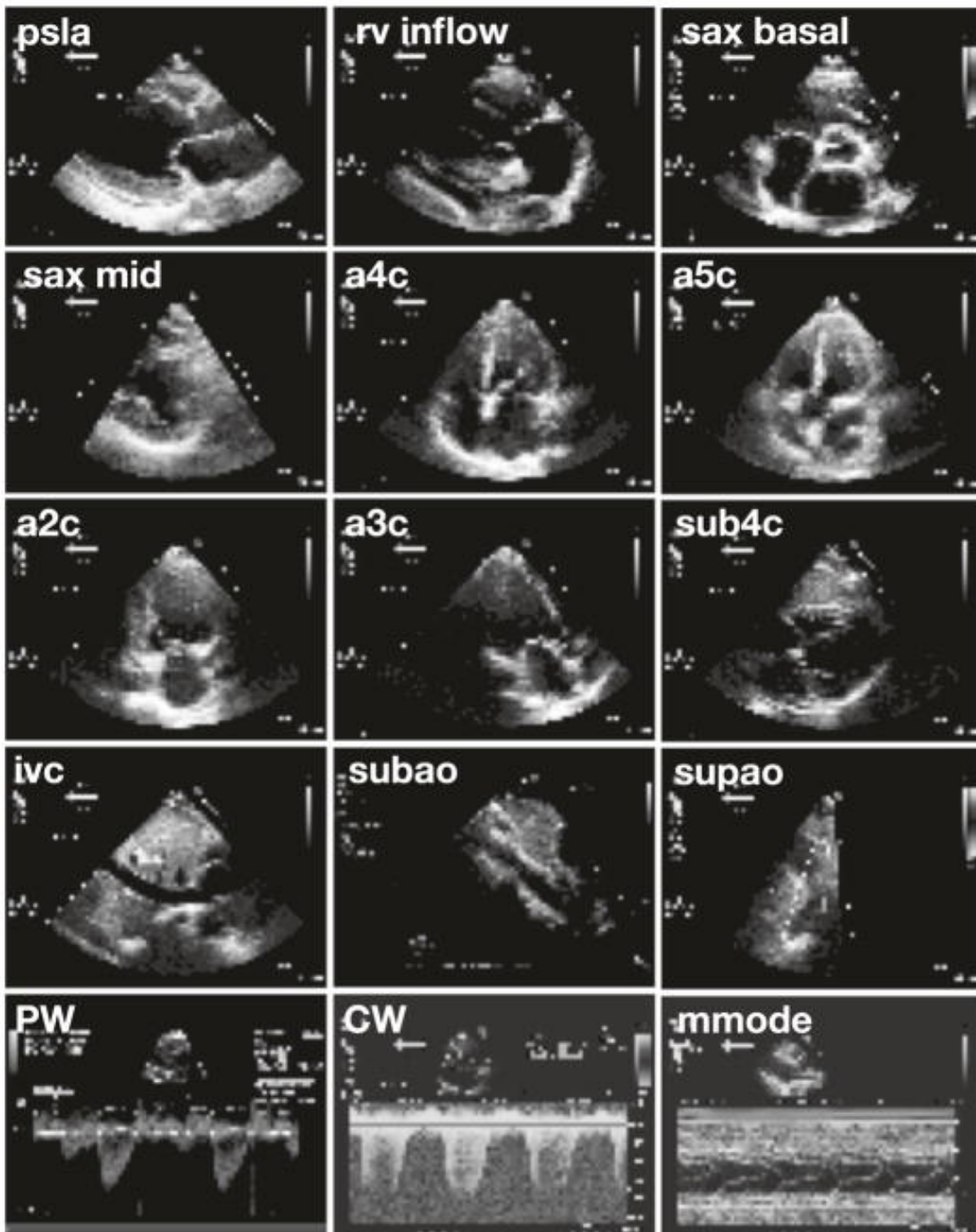


Marie Lannelongue



- Stocker et exploiter les données (informatique)
- pour les analyser ou modéliser (statistique)
- afin de les restituer et éclairer les décisions (épidémiologie)





Fast and accurate view classification of echocardiograms using deep learning

Classification en 15 vues standard
à partir de 267 ETT
Accuracy 91.7%

Big data on small numbers

Cliquez pour modifier les styles du texte du masque

Deuxième niveau

- Troisième niveau
 - Quatrième niveau
 - » Cinquième niveau



Marie Lannelongue

Modalités de traitement des données

Stat

Deep learning

Donnes labellisées

Les algorithmes sont les nouveaux outils des scientifiques
et qui

sont utilisés tous les jours !



Marie Lannelongue



Importance de définir les variables
how to avoid big data = big noise?



Marie Lannelongue



Algorithme de modélisation

All models are wrong but some are useful ! (Georges BOX)



Marie Lannelongue



Big data becomes a big deal

Big data becomes big noise

Est ce que la data compense le manque de connaissance ? On stocke actuellement avant même de savoir quoi faire des données



	ICD	CPT	Lab	Medication	Clinical notes
Availability	High	High	High	Medium	Medium
Recall	Medium	Poor	Medium	Inpatient: High Outpatient: Variable	Medium
Precision	Medium	High	High	Inpatient: High Outpatient: Variable	Medium high
Format	Structured	Structured	Mostly structured	Structured and unstructured	Unstructured
Pros	Easy to work with, a good approximation of disease status	Easy to work with, high precision	High data validity	High data validity	More details about doctors' thoughts
Cons	Disease code often used for screening, therefore disease might not be there	Missing data	Data normalization and ranges	Prescribed not necessary taken	Difficult to process



Possibilité de chainage : suivi longitudinal

Données disponibles par patient

Identifiant unique

permettant de suivre le patient à **travers plusieurs hospitalisations** (Canadian Institute for Health Information)

ou à **travers différentes sources de données** : hôpital, décès, prescriptions (Québec, Danemark)

Concerne n=34 études (58%) provenant de 8 sources de données

Données disponibles par hospitalisation

Pas de chainage entre les différentes hospitalisations d'un même patient

Pas de suivi longitudinal possible (Nationwide Inpatient Sample)

Concerne n=25 études (42%) provenant de 2 sources de données



Marie Lannelongue

Utile seulement une fois raffinée!

"DATA IS THE NEW OIL."

From the beginning of recorded time until 2003, we created **5 exabytes** of data.

In 2011 the same amount was created every two days.

By 2013, it's expected that the time will shrink to 10 minutes.

Every hour, we create enough Internet traffic to fill **7 billion DVDs**.

Side by side, that's that's seven times the height of Everest.

Coined in 2006 by Clive Humby, a British data commercialization entrepreneur, this now famous phrase was embraced by the World Economic Forum in a 2011 report, which considered data to be an economic asset, like oil.

