

Place des « open data » et des « big data » dans les recherches en santé.

Suivi de la table ronde de Giens 2015

Gilles Chatellier

Département d'informatique, biostatistiques et recherche clinique
HEGP, AP-HP, Paris, France



Recommandations

Quel accès pour la recherche ? -1-

Constats

1. Délais
2. Institut National des Données de Santé
3. « Médicalisation » des données
4. Nomenclatures
5. Ressources méthodologiques insuffisantes

Améliorations

1. Respect des délais!
Méthodologies de référence
2. Le créer, et le faire fonctionner, moyens disponibles, valoriser les experts (valable pour la CNIL)
3. Côté de la pathologie, classification NYHA (modification de code) : groupe utilisateur (chercheurs, industriels)
4. Mise à jour rapide
5. Coursus « Big data » (recenser, créer ..)

Recommandations

Quel accès pour la recherche ? -2-

Constats

1. Modèle de données SNIRAM
2. Documentation SNIRAM
3. Appariement des bases
4. Open Data : cohortes, essais industriels, académiques, extraction big data
5. Intérêt public à définir

Améliorations

1. Améliorer le modèle
2. WIKI-SNIRAM
3. Simplification procédure /délais de mise en œuvre
4. Contractualiser l'accès ultérieur aux données et sa pérennisation (modèle INSERM)
5. Pratique décisionnelle élaborée par les acteurs

« Art. L. 1462-1. – Un groupement d'intérêt public, dénommé : "Institut national des données de santé", est constitué entre l'Etat, des organismes assurant une représentation des malades et des usagers du système de santé, des producteurs de données de santé et des utilisateurs publics et privés de données de santé, y compris des organismes de recherche en santé. Il est notamment chargé :

1° De veiller à la qualité des données de santé et aux conditions générales de leur mise à disposition, garantissant leur sécurité et facilitant leur utilisation dans le respect de la loi no 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés ;

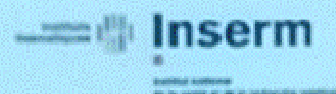
2° D'assurer le secrétariat unique mentionné à l'article 54 de la même loi ;

3° D'émettre un avis sur le caractère d'intérêt public que présente une recherche, une étude ou une évaluation, dans les conditions prévues au même article 54 ;

4° De faciliter la mise à disposition d'échantillons ou de jeux de données agrégées mentionnées au V dudit article 54, dans des conditions préalablement homologuées par la Commission nationale de l'informatique et des libertés ;

5° De contribuer à l'expression des besoins en matière de données anonymes et de résultats statistiques, en vue de leur mise à la disposition du public.

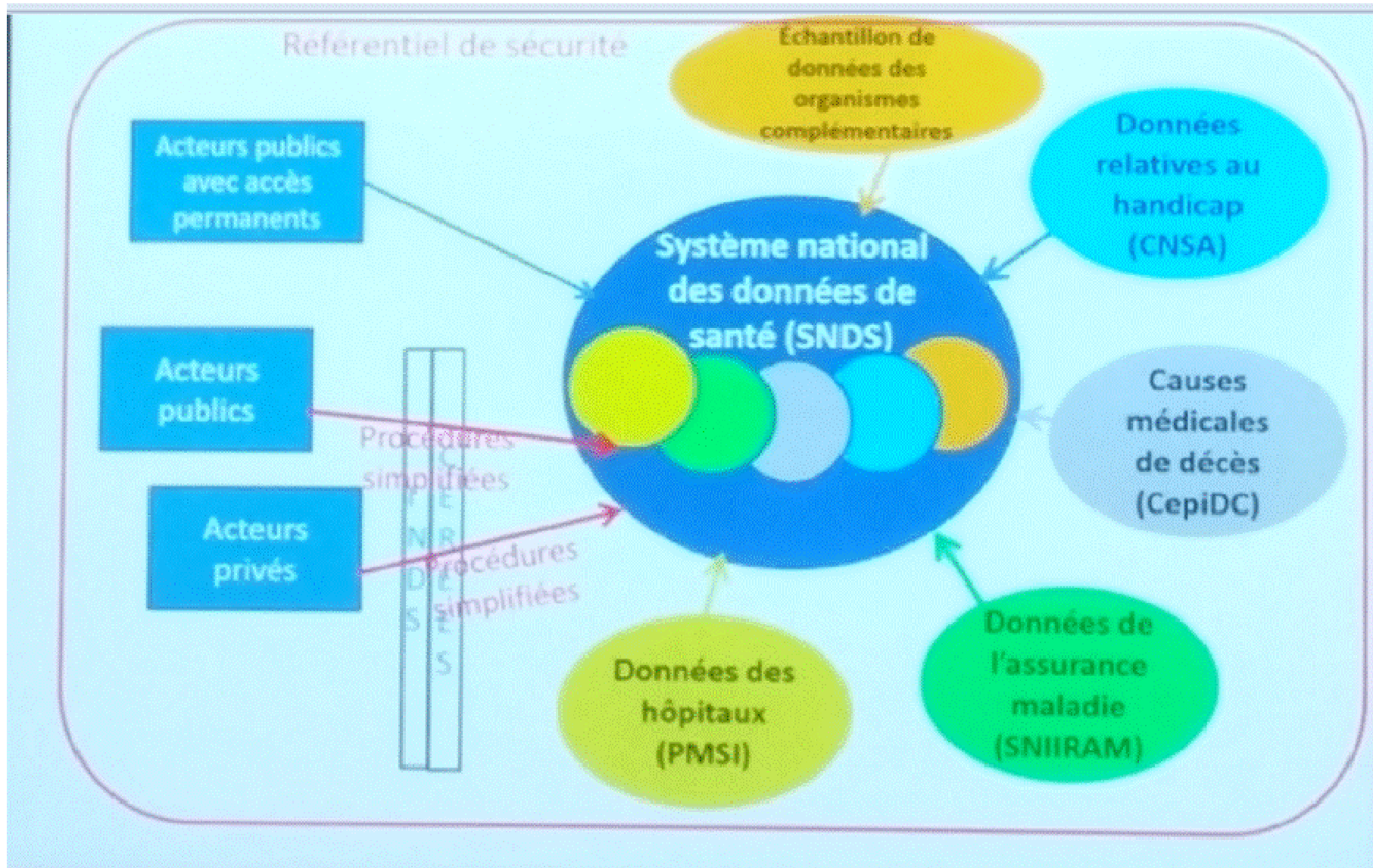
Il publie chaque année un rapport transmis au Parlement. »



Le SNDS met à disposition les données pour contribuer à :

- l'information sur la santé ainsi que sur l'offre de soins, la prise en charge médico-sociale et leur qualité ;
- la définition, à la mise en œuvre et à l'évaluation des politiques de santé et de protection sociale ;
- la connaissance des dépenses de santé, des dépenses d'assurance maladie et des dépenses médico-sociales ;
- l'information des professionnels, des structures et des établissements de santé ou médico-sociaux sur leur activité ;
- la surveillance, à la veille et à la sécurité sanitaires ;
- la recherche, aux études, à l'évaluation et à l'innovation dans les domaines de la santé et de la prise en charge médico-sociale.





EDS APHP: présentation générale

■ Objectif du projet Entrepôt de Données de Santé (EDS)

- ▶ Développer les possibilités d'**exploitation des données du Système d'information Clinique (SIC)** à d'autres finalités que les soins (**recherche, innovation diagnostique et thérapeutique**, enseignement, maîtrise des vigilances et risques, pilotage médico-économique et opérationnel)

■ Plateforme de partage et d'exploitation de « mégadonnées »

- ▶ Solutions sécurisées et performantes de traitement de données massives et complexes (séquences de gènes, images, signaux, textes, etc)
- ▶ Au sein de l'AP-HP et dans le cadre de partenariats régionaux/nationaux ou internationaux

■ Reposant sur trois « piliers »

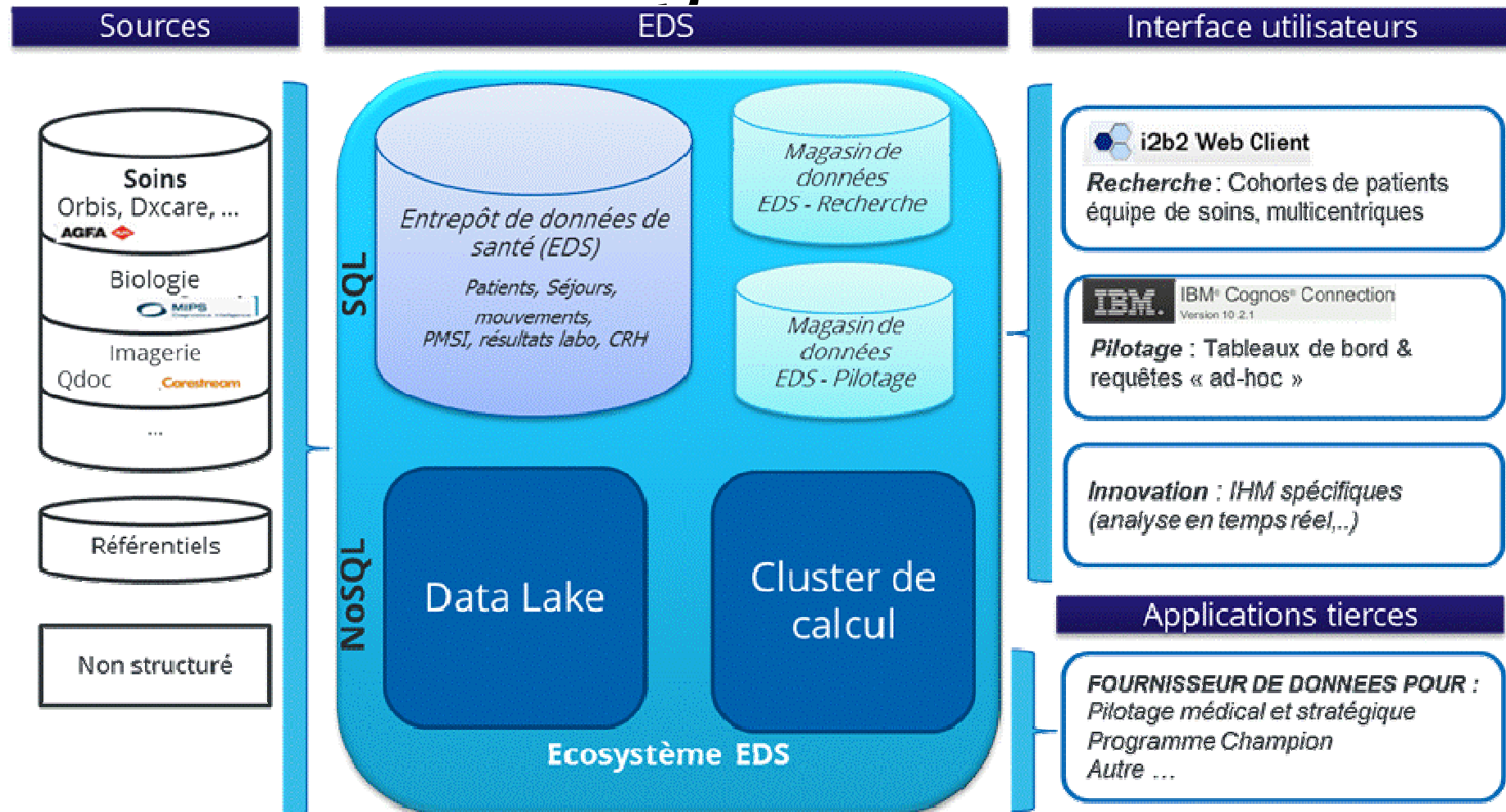
- ▶ Technique, Gouvernance des données, Réglementaire

EDS AP-HP

Plateforme technique méga-données développée en interne à la DSI de l'AP-HP (DSI WIND - Domaine Données)

- EDS – **Recherche** (solution i2b2) [en cours de déploiement]
 - Développer la recherche non interventionnelle sur données (épidémiologie, études de faisabilité, ...)
- EDS - **Innovation** - plateforme évolutive [en cours de conception]
 - Intégration de nouveaux types de données (**imagerie**, génétique, etc.)
 - Capacité d'organiser des calculs statistiques (R, Python, etc.) basés sur un cluster de calcul
- EDS – **Pilotage** (solution IBM Cognos) [en cours de qualification]
 - Support à l'évaluation et l'amélioration des pratiques grâce à la production d'indicateurs et de tableaux de bords

EDS AP-HP: Architecture globale



EDS AP-HP

Aspects réglementaires

- Déclaration normale effectuée pour l'usage équipe de soins / DIM (Février 2015)
- Demande d'autorisation CNIL déposée en juillet 2016 (n°1980120)
 - Constitution d'un entrepôt de données de santé
 - Finalités : Recherche « Equipe de Soins », Recherche Multicentrique et Analyse de l'activité et des pratiques (DIM)
 - Compléments d'information fournis à la CNIL fin 2016
 - Audition des équipes techniques et juridiques de l'AP-HP
 - Autorisation obtenue le **19 janvier 2017**
- Déclaration (n°1986127) de conformité à la méthodologie de référence MR003, afin d'éviter une démarche CNIL étude par étude pour celles qui se conforment à la méthodologie

Délibération n° 2016-263 du 21 juillet 2016 portant homologation d'une méthodologie de référence relative aux traitements de données à caractère personnel mis en œuvre dans le cadre des recherches dans le domaine de la santé ne nécessitant pas le recueil du consentement exprès ou écrit de la personne concernée (MR-003)

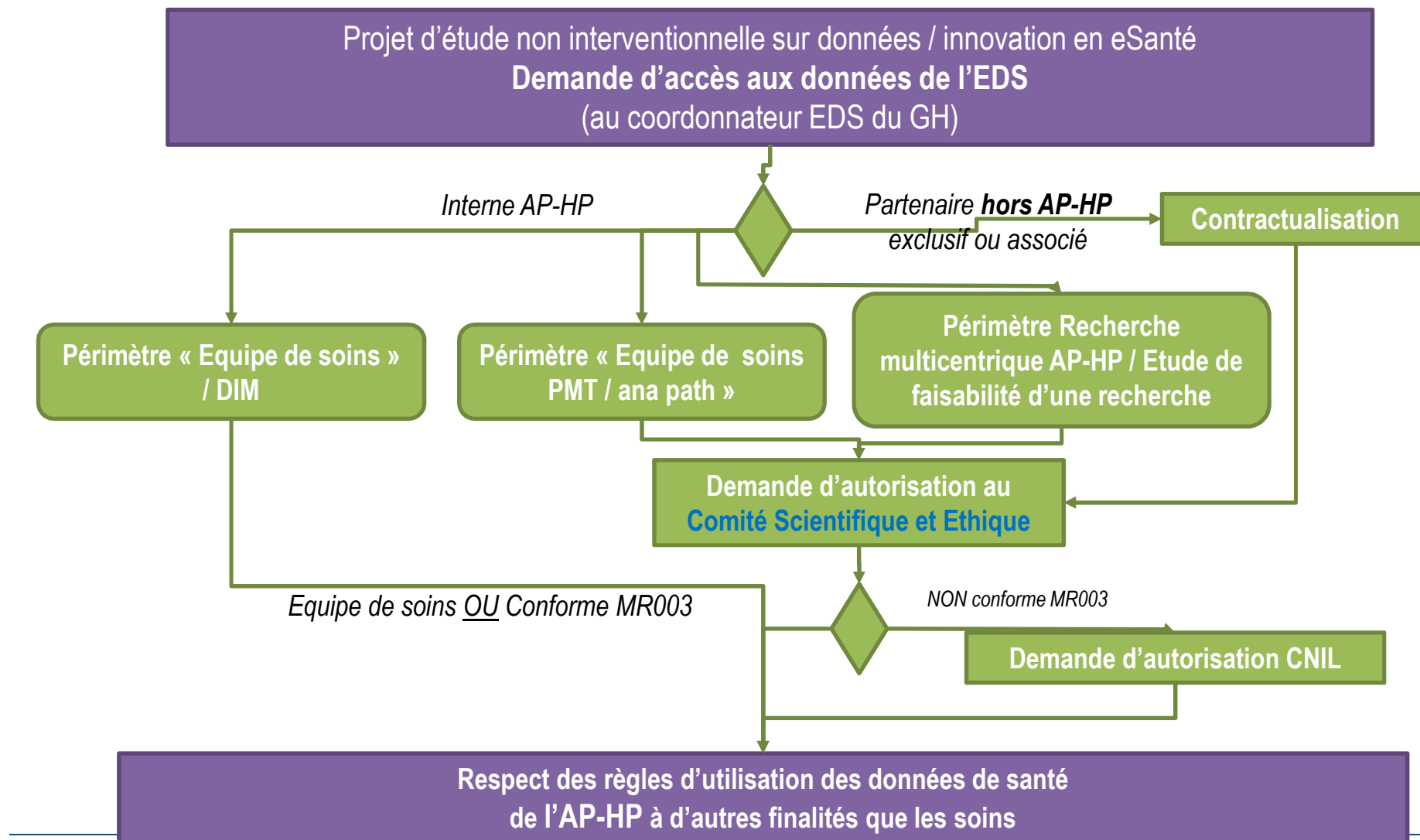
NOR: CNIX1622885X
ELI: Non disponible

EDS AP-HP

Aspects réglementaires

- **Information** du patient
 - Collective et individuelle
 - Multicanal : livret d'accueil, site web de transparence, documents remis à l'admission, documents de facturation, etc.
 - Facilitation des procédures d'opposition
 - Portail de transparence accessible par Internet
 - Formulaire **d'opposition en ligne** qui renvoie vers une boîte aux lettres mail générique (Accès CIL AP-HP et DSI-WIND)
 - En complément du circuit existant sur les sites
- Perspectives : convergence vers une procédure de **consentement express** avec traçabilité
- Information des professionnels de santé

Présentation EDS – Règles d'utilisation





Sachons raison garder...

Il y a devant nous beaucoup de travail, qui devra être effectué avec beaucoup de rigueur...

Clinical Studies That Have Evaluated Personalized Cancer Medicine

Table 1. Clinical Studies That Have Evaluated Personalized Cancer Medicine.*

Clinical Study	Design	Screened Sample	Patients with Genetic Profile	Patients with Mutation That Might Be Targeted by Drugs	Patients Receiving Matched Drug	Main Outcome Result
SHIVA trial ⁸	Randomized, controlled trial of matched molecular targeted agent or physician's choice	741 patients with metastatic solid tumors who were amenable to biopsy	496 (67%)	293 (40%), of whom 195 underwent randomization	96 (100% of experimental-therapy group)	No significant difference in progression-free survival (primary end point); hazard ratio for death or disease progression, 0.88 (95% CI, 0.65–1.19)
Lung Cancer Mutation Consortium	Testing for driver mutations in metastatic lung adenocarcinomas at multiple centers				Many treated as per guidelines for an approved biomarker	Longer overall survival in the subgroups with a mutation treated with directed therapy than in those without the mutation or those that do not receive directed therapy
Study I ⁵		1007 patients	733 (73%) tested for ≥10 genes	466 (46%)	260 (26%)	
Study II ⁶		1315 patients	919 (70%) tested for ≥8 genes	529 (40%) had mutations, with 187 (14%) of them that could be targeted by drugs and had follow-up	127 (10%)	
SAFIR-01 ⁹	Treatment chosen after genetic profiling by comparative genomic hybridization and gene sequencing	423 women with metastatic breast cancer	299 (71%)	195 (46%)	55 (13%)	4 patients had a partial response and 9 had stable disease for >16 wk (3% of screened sample)
M.D. Anderson Study ¹⁰	Treatment chosen after gene sequencing of patients with advanced cancer	2601 patients	2000 (77%)	789 (30%)	83 (3%) in genotype-matched trials; 116 (4%) with common mutations not in trial	Not stated
Princess Margaret IMPACT–COMPACT study ¹¹	Treatment chosen after gene sequencing of archival tissue	1893 patients with advanced solid tumors	1640 (87%)	938 (50%) had mutations, approximately 20% of which could be targeted by drugs	84 (4%) treated in genotype-matched trials	Response rate of 20% in genotype-matched trial vs. 11% in unmatched trials
Cleveland Clinic Study ¹²	Treatment chosen after gene sequencing	250 patients	223 (89%)	109 (44%)	24 (10%)	Not stated

* CI denotes confidence interval, COMPACT Community Oncology Molecular Profiling in Advanced Cancers Trial, and IMPACT Integrated Molecular Profiling in Advanced Cancers Trial.



REDSIAM

Le cœur de l'activité de REDSIAM est constitué par le développement, la discussion critique et la mise à disposition d'algorithmes spécifiques de pathologies basés sur les données du SNIIRAM par des combinaisons de diagnostics du PMSI et/ou d'ALD, de données sur les actes médicotéchniques, les consommations de médicaments ou d'autres produits de santé plus ou moins spécifiques de la pathologie concernée, les consultations de professionnels ...

<http://www.redsiam.org>

Outils statistiques: en plein développement

Performance of the high-dimensional propensity score in adjusting for unmeasured confounders

Jason R Guertin^{1,2,3} • Elham Rahme^{4,5} • Jacques LeLorier¹

Covariate adjustment of cumulative incidence functions for competing risks data using inverse probability of treatment weighting

Anke Neumann*, Cécile Billionnet¹

*French National Health Insurance Fund for Salaried Employees (CNAMTS), Department of Public Health Studies,
Avenue du Professeur André Lémierre 50, 75020 Paris, France*

Predicting the Future — Big Data, Machine Learning, and Clinical Medicine

Ziad Obermeyer, M.D., and Ezekiel J. Emanuel, M.D., Ph.D.

Machine learning has become ubiquitous and indispensable for solving complex problems in most sciences. The same methods will open up vast new possibilities in medicine.

NEJM, Sept 29, 2016

Aide à la décision: rétinopathie diabétique

A deep convolutional neural network was trained ...

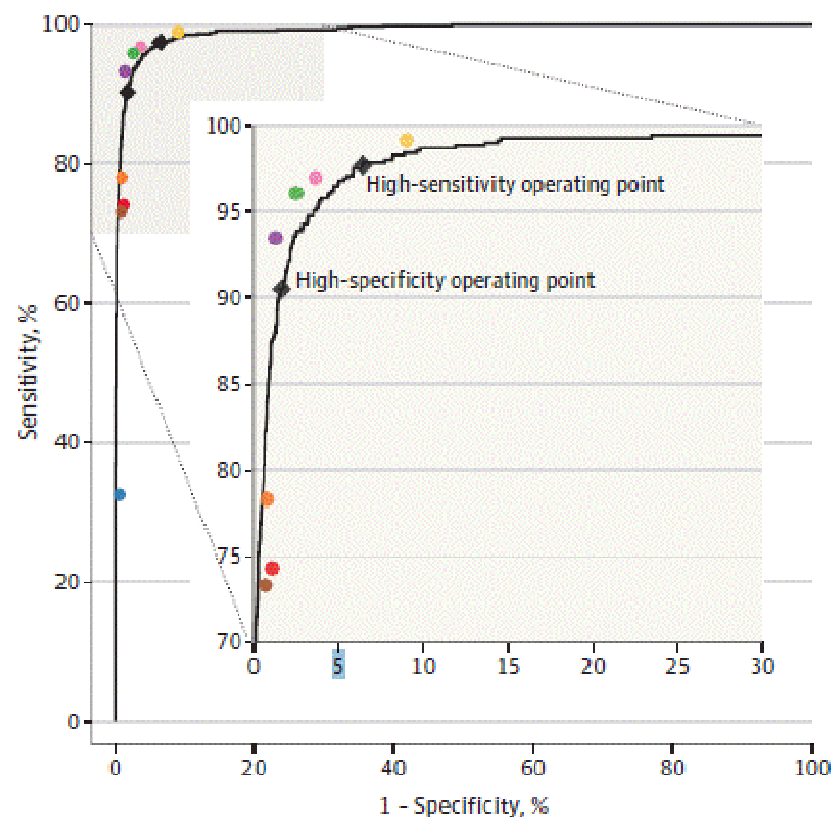
- using a retrospective development data set of **128 175 retinal images**
- graded 3 to 7 times for diabetic retinopathy, diabetic macular edema, and image gradability
- by a panel of 54 US licensed ophthalmologists and ophthalmology senior residents

Validation:

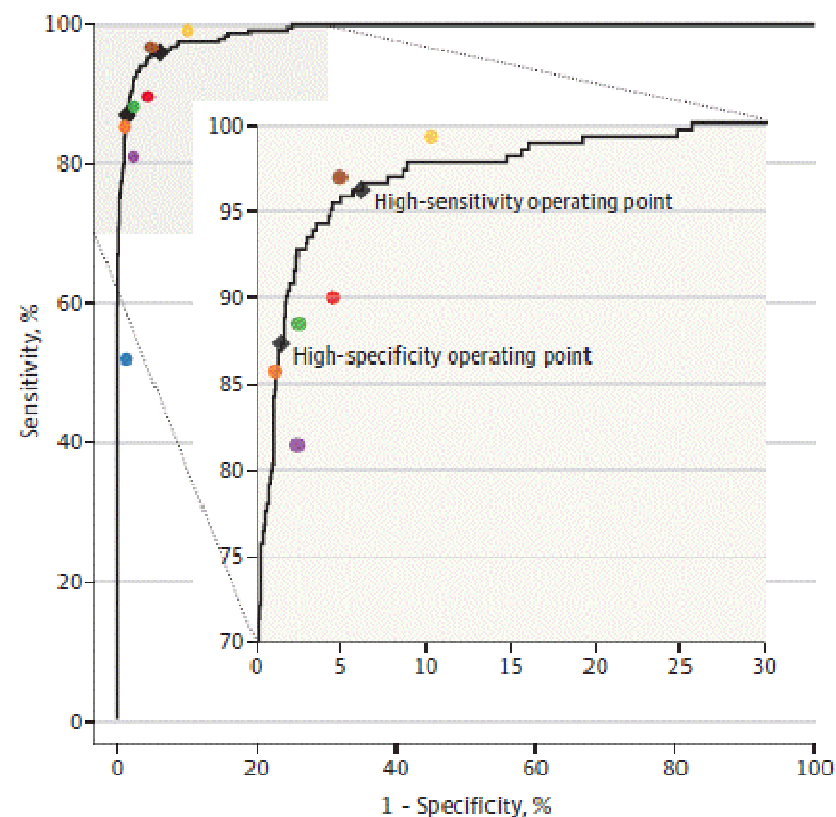
- **2 separate data sets, 9963 and 1748 images**
- both graded by at least 7 US board-certified ophthalmologists with high intragrader consistency.

Figure 2. Validation Set Performance for Referable Diabetic Retinopathy

A EyePACS-1: AUC, 99.1%; 95% CI, 98.8%-99.3%



B Messidor-2: AUC, 99.0%; 95% CI, 98.6%-99.5%



– 2nd data set – sensitivity : 96.1% ;specificity : 93.9%.

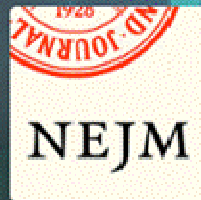


The Importance — and the Complexities — of Data Sharing

Jeffrey M. Drazen, M.D., Stephen Morrissey, Ph.D., Debra Malina, Ph.D.,
Mary Beth Hamel, M.D., and Edward W. Campion, M.D.

“As we work through these complex issues, we want to make it clear that the Journal is committed to making data sharing part of our everyday business. Just as we introduced the inclusion of clinical trial protocols with the publication of all clinical trial research reports, we are working, in the same spirit of transparency, toward the goal of making data sharing a reality”

- Warren E. Strengthening research through data sharing. N Engl J Med 2016;375:401-403
- Krumholz HM, Waldstreicher J. The Yale Open Data Access (YODA) Project — a mechanism for data sharing. N Engl J Med 2016;375:403-405
- The International Consortium of Investigators for Fairness in Trial Data Sharing. Toward fairness in data sharing. N Engl J Med 2016;375:405-407
- The Academic Research Organization Consortium for Continuing Evaluation of Scientific Studies — Cardiovascular (ACCESS CV). Sharing data from cardiovascular clinical trials — a proposal. N Engl J Med 2016;375:407-409



The SPRINT Data Analysis Challenge

To explore the potential of clinical trial data sharing, the New England Journal of Medicine (NEJM) is hosting a challenge: use the data underlying a recent NEJM article to identify a novel clinical finding that advances medical science.

 Share

Home

Winners

Submissions

About

How To Enter

Rules

FAQs

News

 Judges

[Aligning Incentives Web Event](#)

[NEJM Data Sharing Articles](#)

SPRINT Data Analysis Challenge

Want to stay updated?

 [Follow the Challenge](#)

<https://challenge.nejm.org/pages/home>

Pour conclure ...

