

L'accès aux données et les politiques publiques

Guy Lacroix
Université Laval
HEC Montréal

27 mai 2019

ACFAS - Gatineau

Préoccupation ici et aux USA

**David Card, Raj Chetty, Martin Feldstein, Emmanuel Saez,
2011**

We argue that the development and expansion of direct, secure access to administrative micro-data should be a top priority for the NSF...

... The United States would be much better positioned in this scientific race if access to US administrative data was the norm rather than the exception in limited pilot programs.

Liran Einav, Jonathan Levin, 2014

Government administrative data are almost certainly under-utilized, both by government agencies and, because of limited and restricted access, by researchers . . . In cases where the government has allowed access to administrative datasets, there often have been profound consequences for economic policy discussions.

Données administratives vs données d'enquête

Avantages des données administratives

1. Meilleure qualité de l'information
 - Peu de valeurs manquantes
 - Problème d'attrition négligeable (ELIC vs BDIM)
 - Enquête : Taux de non-réponse sur les revenus dans le CPS : 31%
 - Taux d'attrition dans le SIPP sur 3 ans : 30%
2. Échantillons énormes : permet l'utilisation de méthodes d'analyse puissantes
 - Expériences naturelles
 - Méthodes non-paramétriques
 - Analyse causale fondée sur l'intelligence artificielle

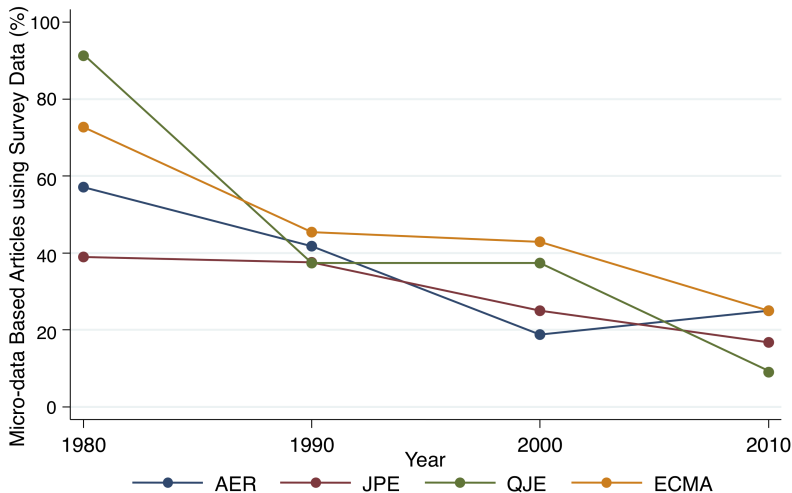
Données administratives vs données d'enquête

Désavantages des données d'enquête

1. Information moins riche
2. Pas colligées pour des fins de recherche
3. Peu documentées
4. Exploitation dépend de ressources internes (taux de roulement)
5. Aucune information quand hors programme (AS, AE, Santé, *etc.*).

Trends in Research : Raj Chetty, 2012

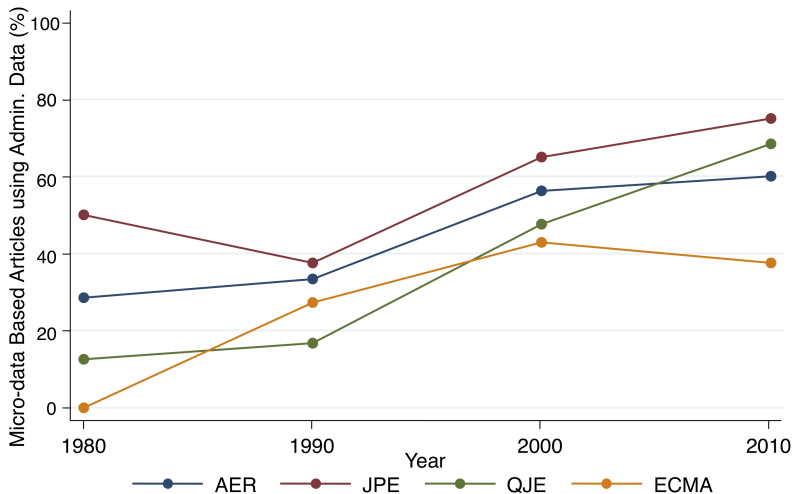
Use of Pre-Existing Survey Data in Publications in Leading Journals, 1980-2010



Note: "Pre-existing survey" datasets refer to micro surveys such as the CPS or SIPP and do not include surveys designed by researchers for their study. Sample excludes studies whose primary data source is from developing countries.

Trends in Research : Raj Chetty, 2012

Use of Administrative Data in Publications in Leading Journals, 1980-2010



Note: "Administrative" datasets refer to any dataset that was collected without directly surveying individuals (e.g., scanner data, stock prices, school district records, social security records). Sample excludes studies whose primary data source is from developing countries.

Quelques constats

1. CAI : règles très sévères et démarches administratives lourdes, délais souvent déraisonnables
2. Contraintes sur le jumelage de données entre ministères
3. Jumelage avec données fédérales très difficile, voire impossible.
4. CADRISQ vs CDR
 - Seulement deux sites
 - Dispose essentiellement de données de l'ISQ
 - Pas ou peu de données administratives

Études passées

1. Évaluation sommative des mesures actives offertes aux individus par Emploi-Québec, 2003 et 2013
 - Données d'enquête fusionnées avec fichiers administratifs de l'aide sociale, fichiers d'impôts T4, Profil vectoriel, RE

Études passées

1. Évaluation sommative des mesures actives offertes aux individus par Emploi-Québec, 2003 et 2013
 - Données d'enquête fusionnées avec fichiers administratifs de l'aide sociale, fichiers d'impôts T4, Profil vectoriel, RE
2. Évaluation économique du RQAP, 2016
 - Fichiers administratifs CGAP, Relevés I, Enquêtes sur la couverture de l'assurance-emploi

Études passées

1. Évaluation sommative des mesures actives offertes aux individus par Emploi-Québec, 2003 et 2013
 - Données d'enquête fusionnées avec fichiers administratifs de l'aide sociale, fichiers d'impôts T4, Profil vectoriel, RE
2. Évaluation économique du RQAP, 2016
 - Fichiers administratifs CGAP, Relevés I, Enquêtes sur la couverture de l'assurance-emploi
3. L'intégration en emploi des immigrants de la catégorie des travailleurs qualifiés au Québec, 2013
 - Enquête basée sur les dossiers de demande de résidence permanente.

Études en cours

1. Évaluation des interventions axées sur la réinsertion sociale des détenus québécois, 2019
 - Fichiers administratifs du Ministère de la Sécurité publique + Données des établissements de Montréal, Québec et St-Jérôme.

Études en cours

1. Évaluation des interventions axées sur la réinsertion sociale des détenus québécois, 2019
 - Fichiers administratifs du Ministère de la Sécurité publique + Données des établissements de Montréal, Québec et St-Jérôme.
2. Évaluation des mesures offertes par Emploi-Québec aux adultes prestataires du programme de solidarité sociale présentant une contrainte sévère à l'emploi, 2019
 - Fichiers administratifs 1998–2018

Études en cours

1. Évaluation des interventions axées sur la réinsertion sociale des détenus québécois, 2019
 - Fichiers administratifs du Ministère de la Sécurité publique + Données des établissements de Montréal, Québec et St-Jérôme.
2. Évaluation des mesures offertes par Emploi-Québec aux adultes prestataires du programme de solidarité sociale présentant une contrainte sévère à l'emploi, 2019
 - Fichiers administratifs 1998–2018
3. Action concertée : Mobilité intragénérationnelle, une comparaison interprovinciale
 - Données ELIA, enquête + données administratives

Conclusion

1. La recherche en science sociale, du moins en science économique, va reposer de plus en plus sur des données administratives. Cela pose plusieurs défis :
 - Formation adéquate en méthodes statistiques et techniques informatiques
 - L'accès aux données : Où ?
 - Faciliter le jumelage de banques de données : Lesquelles ? Fédérales ? Provinciales ?
 - Puissance de calcul : les ordinateurs standards sont incapables de gérer les modèles de type ML avec de grands échantillons

Conclusion

1. La recherche en science sociale, du moins en science économique, va reposer de plus en plus sur des données administratives. Cela pose plusieurs défis :
 - Formation adéquate en méthodes statistiques et techniques informatiques
 - L'accès aux données : Où ?
 - Faciliter le jumelage de banques de données : Lesquelles ? Fédérales ? Provinciales ?
 - Puissance de calcul : les ordinateurs standards sont incapables de gérer les modèles de type ML avec de grands échantillons
2. On apprend en se comparant . . .
 - Les chercheurs québécois sont désavantagés
3. Pas très optimiste. . .

Merci !