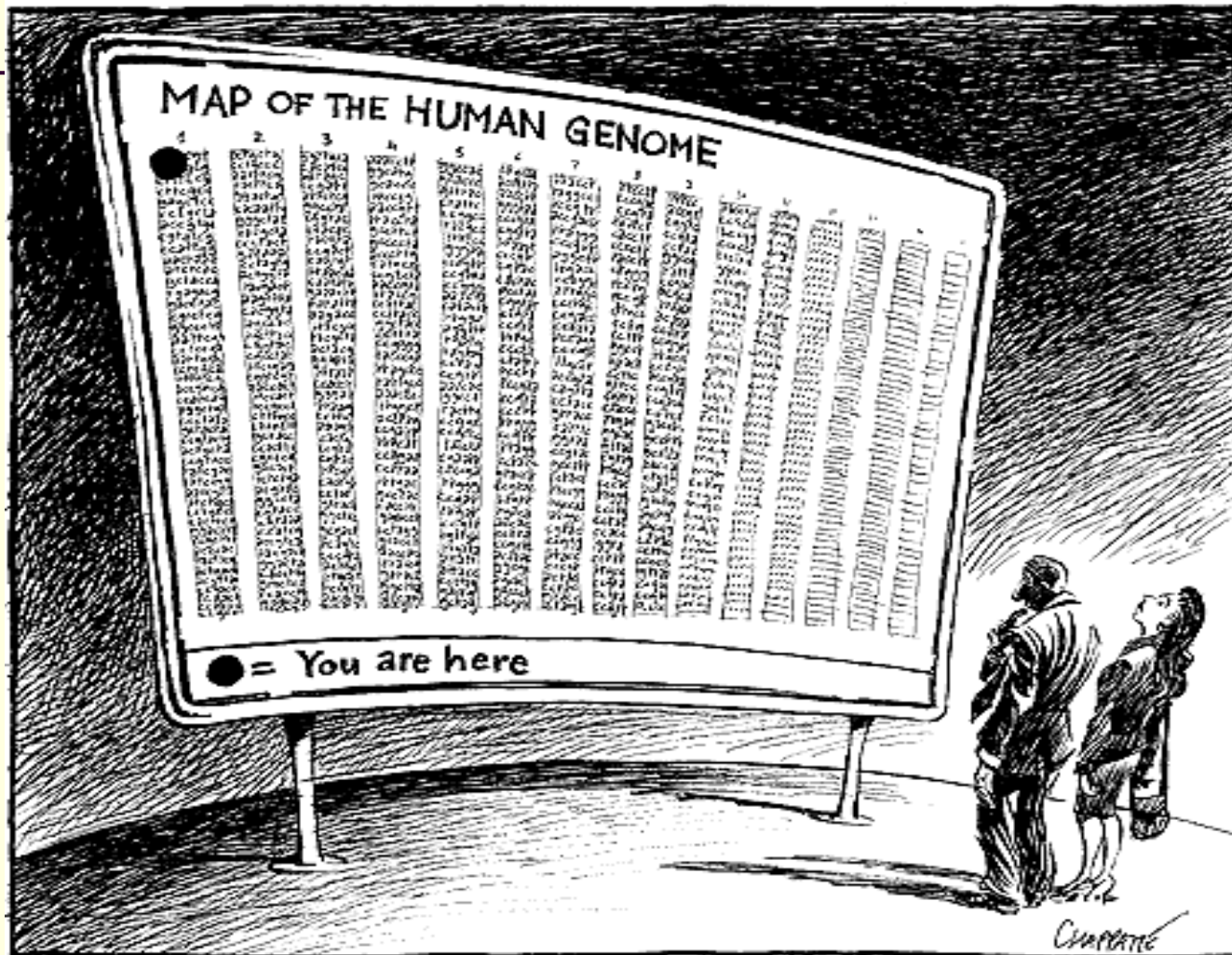


Bases de données électroniques : de la variable au système

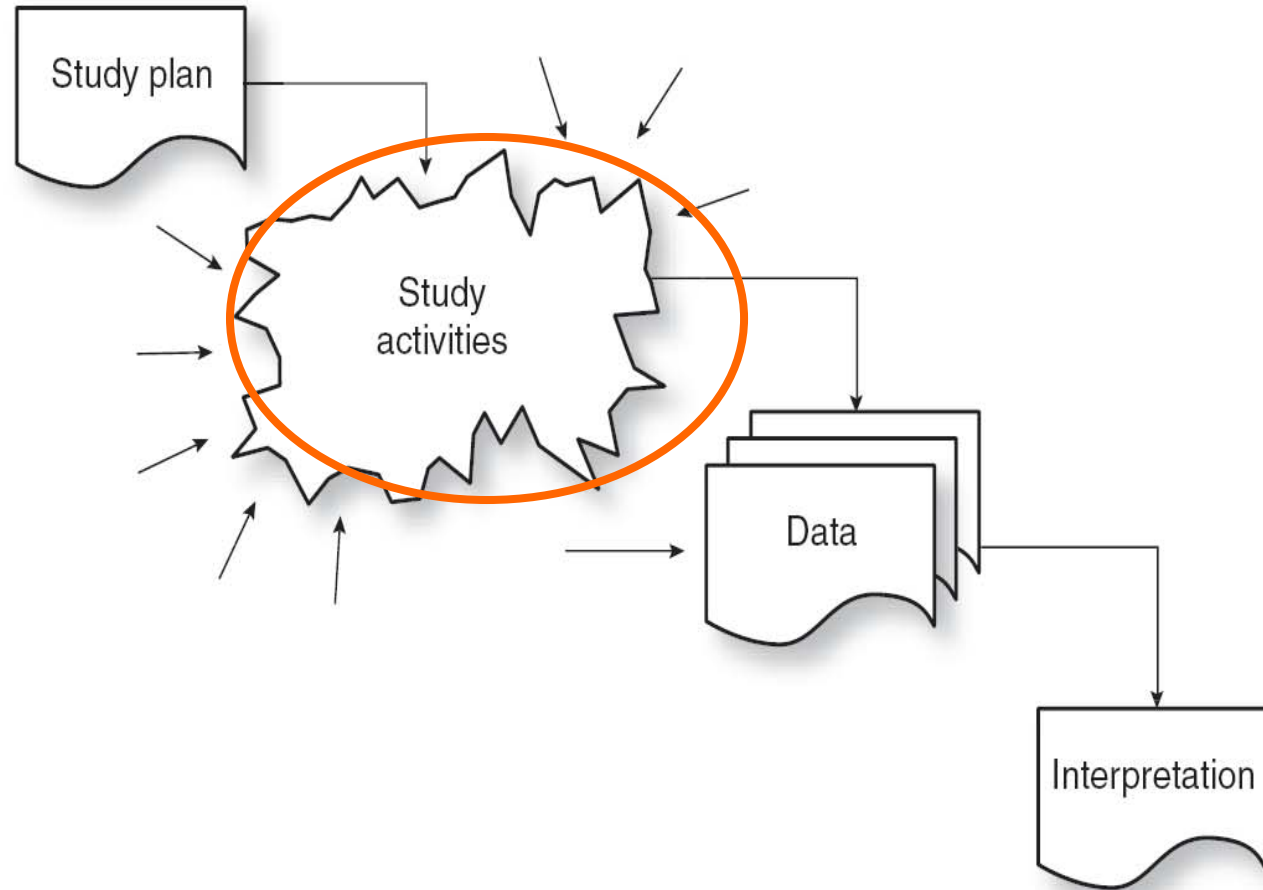


Guy Haller
Chef de clinique scientifique

Unité d'Epidémiologie Clinique-Centre de Recherche Clinique
Département APSI



STRUCTURE GENERALE



OBJECTIF

- Rappel méthodes quantitatives
- Comprendre les exigences d'une bonne récolte de données
 - Connaitre les principales bases de données
 - Etre familier avec les étapes du contrôle qualité

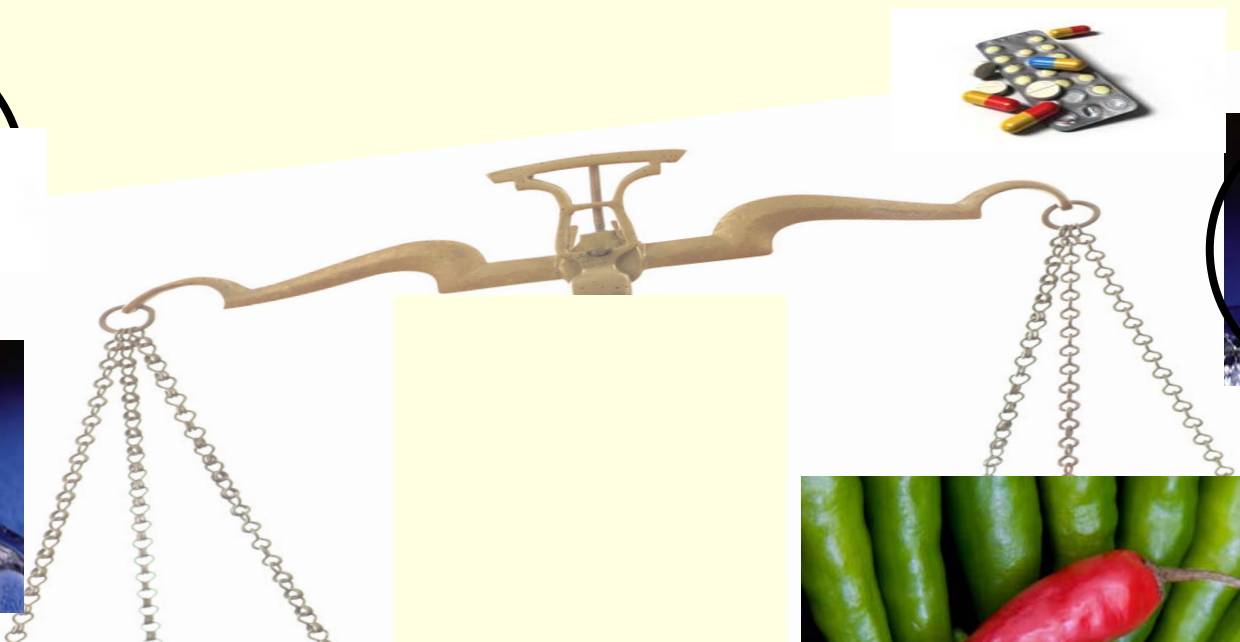
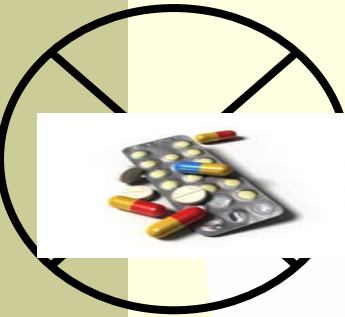
RAPPEL MAXIME



« All is about risk
and disease! »

Risk/treatment → **Disease**

LA RECHERCHE QUANTITATIVE



VARIABLES



Risque

Facteurs de confusion



Traitement



Maladie/Santé

Variables à mesurer:



Les 5 Principes

- **Pertinence**
- **Performance**
- **Parcimonie**
- **Plagiat**
- **Pré-test**

Pertinence

- Variables en relation avec la question de recherche
 - **Maladie** (*≈ outcome, issue, variable dépendante*)
 - **Traitement/exposition** (*≈ facteurs de risque, variables indépendante*)
 - **Autres variables** (*≈ facteurs de confusion*)
 - Tant qu'on y est, ce sera intéressant
 - de savoir

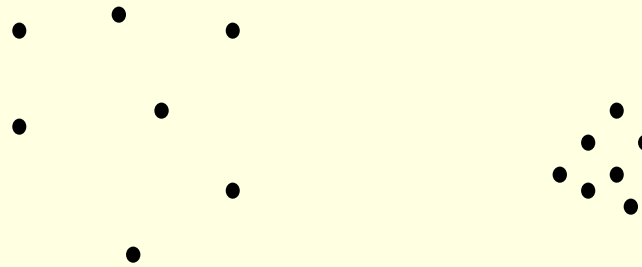
Performance

- ***La mesure de la variable doit être...***
 - *Précise* (mesure reproductible »precision «)
 - *Valide* (mesure représentative « accuracy «)

Précision

Métaphore du tir sur la cible

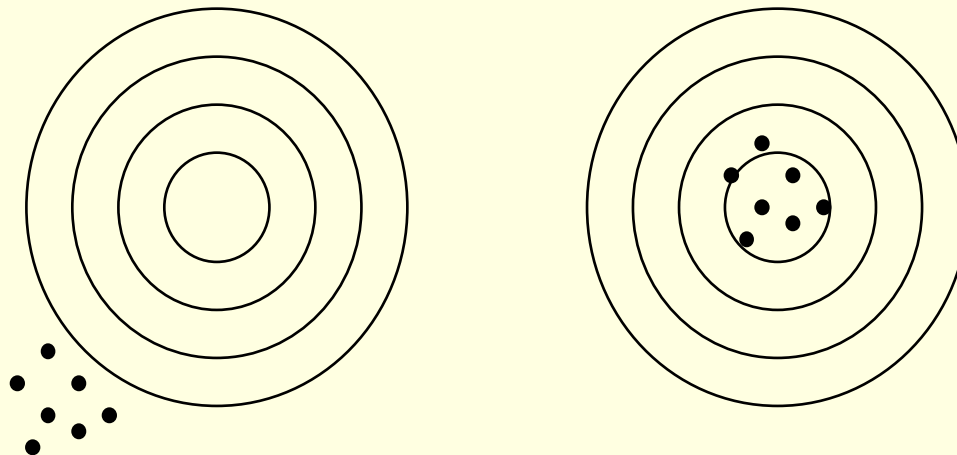
- tir groupé?



Validité

Métaphore du tir sur la cible

- dans le mille?



Exemples questionnaires

- Est-ce qu'une infirmière vous a donné des conseils au sujet de votre traitement à domicile et de votre suivi médical après votre sortie ?
- Avez-vous reçu au cours de votre séjour des informations sur la douleur et son traitement ?
- Que pensez-vous de la disponibilité des infirmières et des médecins qui se sont occupés de vous ?

***Double
contenu***

Exemples questionnaires

L'accueil par le personnel des urgences était chaleureux et courtois

- Pas du tout d'accord.....*
- Pas d'accord.....*
- Pas sûr.....*
- D'accord.....*
- Tout à fait d'accord.....*

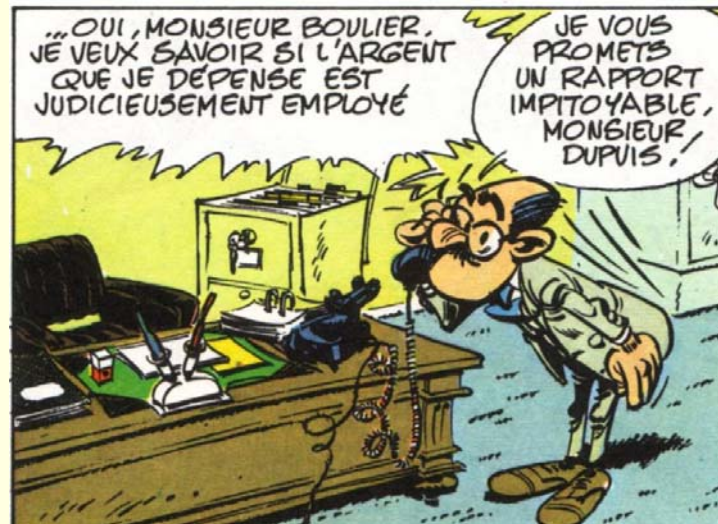
L'accueil par le personnel des urgences était chaleureux et courtois

- Pas du tout d'accord.....*
- Pas vraiment d'accord.....*
- D'accord.....*
- Tout à fait d'accord.....*
- Ne sait pas.....*

**échelle de
réponse
ambiguë**

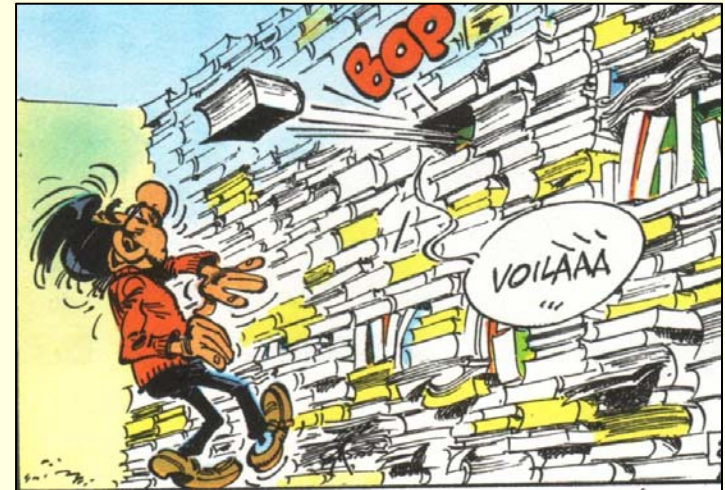
Parcimonie

- Tout recueil de donnée a un coût (*temps* \approx \$\$\$)
- Chaque donnée doit être récoltée dans un but précis, spécifié dans le protocole de recherche
- Prévoir « a priori » comment chaque variable sera analysée



Plagiat

- Toujours faire une revue de la littérature !
- Utiliser le plus possible des instruments pré-existants et reconnus (avec l'accord des auteurs...)
- Parler aux experts dans le domaine
 - Avantages:
 - moins de travail
 - performances connues de l'instrument
 - comparabilité
- *Attention aux traductions si on utilise des questionnaires !!!*

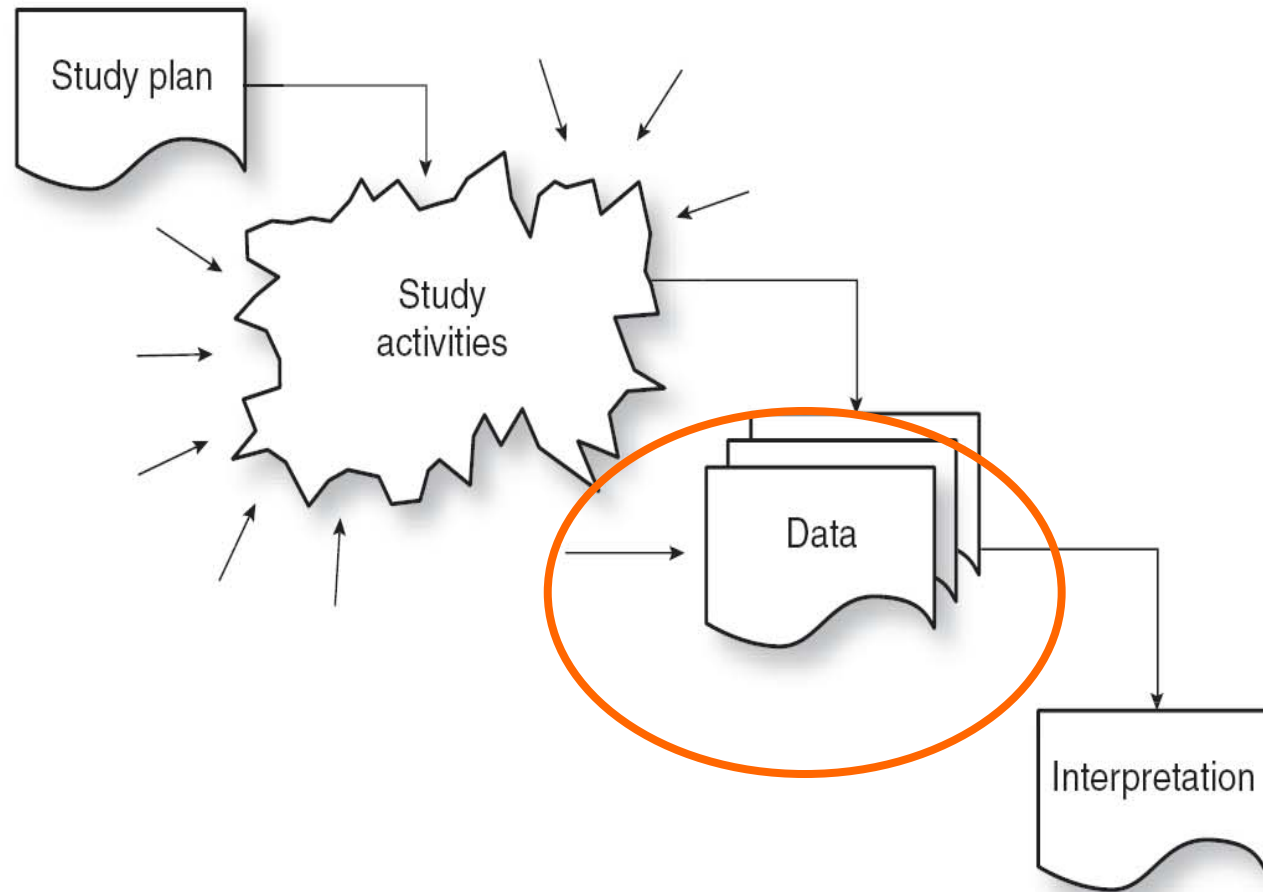


Eviter les pièges (étude pilote)

- **ESSENTIEL ! ! !**
- Différents objectifs:
 - tester la formulation des variables
 - tester la précision et fiabilité de la récolte des variables
 - estimer temps nécessaire pour récolter les variables
 - tester les procédures de recrutement
 - tester l'acceptation de votre questionnaire auprès de la population-cible
 - ...

Prendre son temps et recommencer jusqu'à ce que l'on soit satisfait ! ! !

STRUCTURE GENERALE



Etapes de collections de données

- Acquisition des données
- Relevé des données
- Enregistrement des données
- Consolidation des données

Acquisition des données

Processus de mesure et utilisation d'instruments

- Exemples
 - Sphygmomanomètre
 - Electrocardiogramme
 - Examen clinique et dossier médical
 - Questionnaire

Relevé des données

Transcription de l'information sur un formulaire patient individualisé

- Dossier papier
- Dossiers semi-computérisés
- Dossier computérisé

Exemples semi/computérisés

- Formulaires scannables (NCR)
- Formulaires téléfax (Teleform)
- MySQL server
- Handy-base

1	How old are you?	Please Select...
2	What is your sex?	<input type="radio"/> Female <input type="radio"/> Male
3	In what grade are you?	<input type="radio"/> 9th <input type="radio"/> 10th <input type="radio"/> 11th <input type="radio"/> 12th
4	Where were you born?	<input type="radio"/> United States <input type="radio"/> Mexico <input type="radio"/> Other Country
5	How many years have you lived in the United States?	Please Select...
6	What language(s) do you feel most comfortable speaking?	Please Select...
7	How would you describe yourself?	<input type="radio"/> Hispanic <input type="radio"/> Mexican <input type="radio"/> Mexican-American <input type="radio"/> Other
8	With what hand do you usually do things like comb your hair or throw a ball?	<input type="radio"/> Left <input type="radio"/> Right
9	Has your family ever moved around Texas or the United States to do farm work ? (For example, picking fruits or vegetables for work)	<input type="radio"/> Yes <input type="radio"/> No

Enregistrement des données

Se réfère aux différentes méthodes de utilisées pour transformer et enregistrer les données dans un format électronique

Principales méthodes

- Lecture optique formulaire (scanning)
- Reconnaissance optique (DATAFAX)
- Entrée manuelle des données dans une base de données

Types de base de données

- Locale
- Centrale +/- web-based

Locale versus centrale

Données enregistrées localement par l'équipe de recherche

→ Résolution rapide des omissions, erreurs de transcription

ou

Formulaire envoyés à un sponsor ou à un centre de coordination

→ Données enregistrées par du personnel expérimenté, mais peu de moyens de corriger les erreurs "sur place"

Type de Base de données

- Feuilles de calcul (Excell, SPSS, STATA)
- Bases simples (Epidata)
- Bases relationnelles (Access, File-maker Pro, Oracle)
- Bases relationnelles objet (Business object)
- Bases hiérarchiques (Light Weight Directory Access Protocol)
- Bases web (mySQL, cold fusion, ortavista, sinatras)

Base de données

- Représentation symbolique du monde réel
- Collection structurée de données

Mais pas.....

Une représentation exacte du monde

Un système de stockage des données

Une solution à tous les problèmes

Infailible

Consolidation des données

- Nettoyage
- Remise à jour
- Addition de données manquantes
- Addition de nouvelles données (autres sources)

Récolte des données: Qualité-un problème crucial

“No study is better than the quality of its data.”

Types d'erreurs

- Acquisition des données (3 -10%)
- Relevé des données (3 - 5%)
- Enregistrement des données (0.1 - 1.2%)
- Consolidation des données

Patient ID	Age	Gender	City	Diagnosis	Occupation
1	32	M	Geneva	COPD	Employee
2	34	F	Geneva	Infarct	Employee
3	35	F	Geneva	Polytrauma	Social worker
4	0	M	Geneva	Diabetes	Student
5	46	M	Geneva	HTA	Student
6	48	F	Geneva	Fracture	Independant
7	67	M	Geneva	Polytrauma	Independant
8	54	M	Geneva	Diabetes	Independant
9	23	M	Geneva	Cancer	Student
9	23	M	Geneva	Cancer	Student
9	23	F	Geneva	Cancer	Student
10	33	F	Geneva	COPD	Employee
11	567	F	Geneva	Infarct	Employee
12	34	F	Geneva	Polytrauma	Social worker
13	67	M	Geneva	Diabetes	Student
14	87	F	Geneva	HTA	Student
15	76	M	Geneva	Fracture	Independant
16	65	M	Geneva	Diabetes	Independant
17	54	M	Geneva	Cancer	Independant
18	34	F	Geneva	Cancer	Student
19	12	F	Geneva	Cancer	Student
20	23	M	Geneva	HTA	Student

Méthodes pour limiter les erreurs

- Utilisation d'outils de mesure ayant une bonne fiabilité et précision
- Préparation soigneuse des outils d'acquisition des données
- Développement d'un manuel de procédures
- Certification et formation du personnel de recherche
- Choix et design adapté de la base de données
- Pré-test
- Contrôles réguliers et nettoyage final

Outil d'acquisition des données

Outil d'acquisition des données

- Clair et concis
- Bien organisé et logique
- Limité dans les champs “libres”
- Sans questions pilotes!

Exemple questions

Postmenopausal Estrogen/Progestin Intervention (

“ Some women find that menopause and estrogen replacement therapy may affect their sexual relationships. Although the following questions are sensitive and personal, they are important in determining how treatment affects this part of your life. Please be assured that your responses to these questions will remain confidential. ”

A votre avis....

Version 1:

Are you currently having sexual intercourse with a partner?

Yes No

Version 2:

Are you currently sexually active?

Yes No

Version 3:

Do you ever engage in any sexual activity with a partner?

Yes No



Manuel de Procédures



Manuel de Procédures

- Procédures standardisées
- Instructions clairement détaillées
- Updates et clarifications régulières
- Accessibilité
- Contacts de référence

Base de données

Base de données

- Adaptée à l'utilisateur + checks (1%→0.1%)
- Sécurisée (confidentialité-données sensibles)
- Protégée (virus; accès non-autorisé)
- Régulièrement sauvegardée

Exemple web-based



Site Map

Login Name: PHOGAN (Staff ID: 30011) Login date/time: 04/27/2006 3:48:22PM
Automatic logout: 04/27/2006 4:08:23PM

Staff	Study Administration	Diabetes Support and Education	Publications and Presentations
Directory	Accelerometry Data Management System Data Management Reports	Session Documentation Posters/Flyers DSE Out Years/Phase III Information	Study Slides Abstracts/Publications Manuscript Registry Publications Poster Template
Committees	Certification Query Engine Potential Participant Transfers	Adjudication	Bibliography Citations Manuscript Proposal Form Logos Acknowledgments for Baseline Papers NIH Public Access Policy
Committee Committee Roster page 1 Committee Roster page 2 Working Group Roster	Cross-study Documents	Adjudication Forms Adjudication Data Entry Adjudication Manual	
News	Manual of Procedures Duality of Interest Form Industry Policy Duality of Interest Policy Model Consents Privacy in Communications DSMB Report - April 2004 Central Lab Manual of Procedures Phase III IRB Letter	GXT Certification	
Look AHEAD Newsletter Forum/Bulletin Board FAQ		Certification Update Memo Instructions for Certification Maximal GXT Certification SubMax GXT Certification GXT Clinic Certifications	

Sécurité

Web-based

- Données encryptées
 - Protocoles de transmission sécurisée(HTTPS)
 - Données transmises encryptées 128 bit SSL (Secure Socket Layer)
- Accès protégé
 - Sessions automatiquement limitées en temps
 - Accès par mot de passe régulièrement renouvelé
 - Accès spécifiques et limités en étendue
 - Traçabilité de tous les accès

Contrôle continu des activités

- Rapport régulier et feed-back aux centres participants
- Audit des données
- Visites de site
- Sessions de formation continue du personnel de recherche

Nettoyage des données

- Vérifier plusieurs enregistrements visuellement (fichier informatique versus papier)
- Tableaux de fréquences de toutes les variables:
 - valeurs aberrantes (ex: sexe autre que M, F)
 - valeurs manquantes
- Comparer les sources (ex: âge déclaré versus date de naissance sur dossier médical)
- Examiner des relations connues entre variables (ex: les hommes devraient peser plus que les femmes)



Bonne Chance

