 Société Française de Pharmacologie et de Thérapeutique			Programme de la journée du 4 septembre 2024
10h	Accueil café		
Prévention, traitement et interprétation des données manquantes - Modération : Bruno Laviolle, Anne Ego			
10h15	Structures des données manquantes et impact sur l'estimation de l'effet du traitement	Silvy Laporte	
10h45	Méthodes de gestion des données manquantes	Antoine Vilotitch	
11h15	Interprétation : comment décider en cas de données manquantes	Clara Locher	
11h45	Discussion		
12h15	Déjeuner sur place		
Actualités et nouveautés des méta-analyses - Modération : Michel Cucherat, Matthieu Roustit			
13h30	Les différents types de revues	Céline Chapelle	
14h00	Les différents types de méta-analyses	Charles Khouri	
14h45	Interprétation des méta-analyses ; méthode GRADE	Rémy Bousageon	
15h30	Discussion		
16h00	Fin de la journée		

1

1

La rentrée de la méthodo
5^{ème} édition
Lyon, 4 septembre 2024

 Société Française de
Pharmacologie et de Thérapeutique

Structures des données manquantes et impact sur l'estimation de l'effet du traitement

Silvy Laporte

Service de Pharmacologie Clinique
Unité de Recherche Clinique
INSERM U1059









2

Biometrika (1976), **63**, 3, pp. 581-92
 Printed in Great Britain

581

Inference and missing data

By DONALD B. RUBIN

Educational Testing Service, Princeton, New Jersey

SUMMARY

When making sampling distribution inferences about the parameter of the data, θ , it is appropriate to ignore the process that causes missing data if the missing data are 'missing at random' and the observed data are 'observed at random', but these inferences are generally conditional on the observed pattern of missing data. When making direct-likelihood or Bayesian inferences about θ , it is appropriate to ignore the process that causes missing data if the missing data are missing at random and the parameter of the missing data process is 'distinct' from θ . These conditions are the weakest general conditions under which ignoring the process that causes missing data always leads to correct inferences.

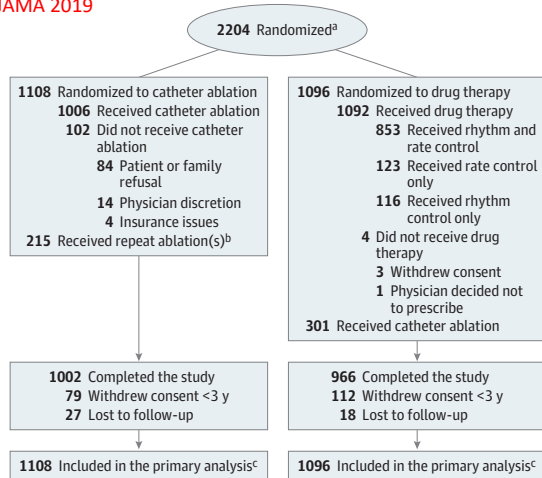
Some key words: Bayesian inference; Incomplete data; Likelihood inference; Missing at random; Missing data; Missing values; Observed at random; Sampling distribution inference.

3

3

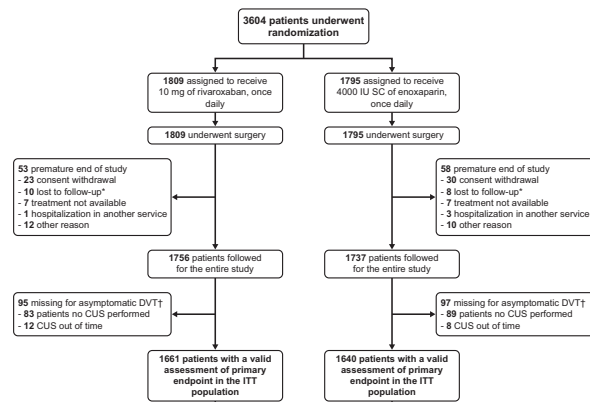
Effect of Catheter Ablation vs Antiarrhythmic Drug Therapy on Mortality, Stroke, Bleeding, and Cardiac Arrest Among Patients With Atrial Fibrillation The CABANA Randomized Clinical Trial

JAMA 2019



Rivaroxaban or Enoxaparin in Nonmajor Orthopedic Surgery The PRONOMOS study

N Engl J Med 2020



4

4

Pourquoi des données manquantes ?

Différentes raisons

- Le patient déménage à l'étranger
- Le patient ne veut plus venir en consultation
- Le patient abandonne l'étude
- Le médecin annule la visite car urgence au bloc et oublie de redonner un rdv
- Le médecin de l'étude a quitté le centre
- La radiographie permettant de statuer sur l'outcome est interprétable
- Le patient est décédé avant la mesure de l'outcome
- Le dossier est perdu après mutation du patient

5

5

Données manquantes, est-ce vraiment un problème ?

Essai clinique dimensionné pour

- montrer une quantité d'effet
- avec un risque de base supposé

Impact données manquantes

- moins d'information que prévu
- perte de puissance statistique

Essai randomisé pour

- contrôler le biais de sélection
- équi-répartir les patients à risque

Impact données manquantes :

- sélection indirecte des patients analysés
- possible biais d'attrition / de sélection

Perte d'information non informative

Impact nul sur l'estimation de l'effet traitement

Impact sur la précision

Perte d'information informative

Biais possible sur l'estimation de l'effet traitement

et impact sur la précision

Importance de l'impact fonction : - du taux de données manquantes

- de la nature des données manquantes

6

6

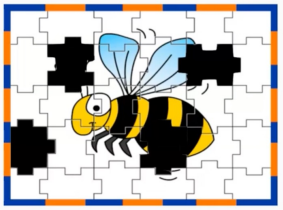
The image shows four puzzle pieces, each containing a cartoon bee. The missing pieces are represented by black squares.

- MCAR (missing completely at random):** The missing pieces are scattered randomly across the puzzle.
- MAR (missing at random):** The missing pieces form a regular, repeating pattern.
- MNAR (missing not at random):** The missing pieces form a large, irregular shape, suggesting the missingness depends on the unobserved data.

Stat Mia
<https://www.youtube.com/watch?v=YpqUbirqFxQ>

7

MCAR (missing completely at random) Manquant de manière complètement aléatoire





Y : outcome = perte de poids
 X1 : sexe X2 : groupe de rando Exp/C

- Probabilité d'une donnée manquante sur le poids : ne dépend ni de sa valeur, ni du sexe ni du groupe de rando
- Probabilité de ne pas avoir l'information : idem pour tous
- Echantillon de patients sans donnée manquante = sous-groupe aléatoire des patients randomisés

Les causes des données manquantes ne sont pas liées aux données elles-mêmes

- Hypothèse forte
- Situation idéale

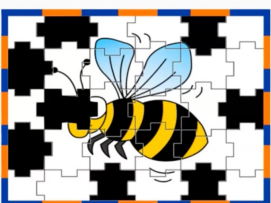
8

	groupe Exp N=150	groupe C N=150	
 Données complètes	DM	0	0
	évaluables	150	150
	evt	15	30
	pas d'evt	135	120
	10%	20%	
			RR 0.50 [0.28 ; 0.89] P value 0.019
 Scénario 1 MCAR 33%	DM	50	50
	évaluables	100	100
	evt	10	20
	pas d'evt	90	120
	10%	20%	
			RR 0.50 [0.25 ; 1.01] P value 0.055

9

9

MAR (missing at random) Manquant de manière aléatoire



Y : outcome = perte de poids
X1 : sexe X2 : groupe de rando Exp/C



- Probabilité d'une donnée manquante sur le poids : plus importante chez les femmes, dépend du sexe (ou dans le groupe B, dépend du trt)
- Et observation manquante chez les femmes de façon aléatoire (ne dépend pas de sa valeur Y)

L'absence de données sur l'outcome ne dépend que de valeurs observées

- Hypothèse moins contraignante, plus générale
- Situation sûrement plus réaliste que MCAR

10

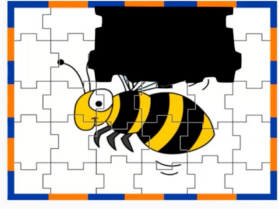
10

	groupe Exp N=150	groupe C N=150	
 Données complètes	DM	0	0
	évaluables	150	150
	evt	15	30
	pas d'evt	135	120
		10%	20%
			RR 0.50 [0.28 ; 0.89] P value 0.019
 Scénario 2 MAR 33% groupe C et MCAR groupe C	DM	0	50
	évaluables	150	100
	evt	15	20
	pas d'evt	135	120
		10%	20%
			RR 0.50 [0.27 ; 0.93] P value 0.028

11

11

MNAR (missing not at random) Manquant de manière non aléatoire





Y : outcome = perte de poids
X1 : sexe X2 : groupe de rando Exp/C

- Probabilité d'une donnée manquante sur le poids :
 - différente entre les patients
 - varie pour des raisons inconnues
 - peut être plus élevée pour les patients qui n'ont pas réussi à perdre du poids
- Echantillon de patients sans donnée manquante ≠ sous-groupe aléatoire des patients randomisés

L'absence de données sur l'outcome dépend de données non observées, et peut donc dépendre de la valeur elle-même



12

12

	groupe Exp N=150	groupe C N=150	
 Données complètes	DM	0	0
	évaluables	150	150
	evt	15	30
	pas d'evt	135	120
		10%	20%
			RR 0.50 [0.28 ; 0.89] P value 0.019
 Scénario 3 MNAR 33% groupe C sur evt seulement	DM	0	10
	évaluables	150	140
	evt	15	20
	pas d'evt	135	120
		10%	14%
			RR 0.70 [0.37 ; 1.31] P value 0.26

13

13

	groupe Exp N=150	groupe C N=150		
 Données complètes	DM	0	0	
	évaluables	150	150	
	evt	15	30	
	pas d'evt	135	120	
		10%	20%	
				RR 0.50 [0.28 ; 0.89] P value 0.019
 Scénario 1 MCAR 33%	DM	50	50	
	évaluables	100	100	
	evt	10	20	
	pas d'evt	90	120	
		10%	20%	
				RR 0.50 [0.25 ; 1.01] P value 0.055
 Scénario 2 MAR 33% groupe C et MCAR groupe C	DM	0	50	
	évaluables	150	100	
	evt	15	20	
	pas d'evt	135	120	
		10%	20%	
				RR 0.50 [0.27 ; 0.93] P value 0.028
 Scénario 3 MNAR 33% groupe C sur evt seulement	DM	0	10	
	évaluables	150	140	
	evt	15	20	
	pas d'evt	135	120	
		10%	14%	
				RR 0.70 [0.37 ; 1.31] P value 0.26

pattern des données manquantes ignorable

pattern des données manquantes non ignorable

14

Peut-on savoir si on est plutôt scénario 2 ou scénario 3 ?



Scénario 2
MAR 33% groupe C
et MCAR groupe C

	groupe Exp N=150	groupe C N=150
DM	0	50
évaluables	150	100
evt	15	20
pas d'evt	135	120
	10%	20%

RR 0.50 [0.27 ; 0.93]
P value 0.028



Scénario 3
MNAR 33% groupe C
sur evt seulement

	groupe Exp N=150	groupe C N=150
DM	0	10
évaluables	150	140
evt	15	20
pas d'evt	135	120
	10%	14%

RR 0.70 [0.37 ; 1.31]
P value 0.26

15

15

Peut-on savoir si on est plutôt scénario 1 ou scénario 2-3 ?



Scénario 1
MCAR 33%

	groupe Exp N=150	groupe C N=150
DM	50	50
évaluables	100	100
evt	10	20
pas d'evt	90	120
	10%	20%

RR 0.50 [0.25 ; 1.01]
P value 0.055



Scénario 2
MAR 33% groupe C
et MCAR groupe C

	groupe Exp N=150	groupe C N=150
DM	0	50
évaluables	150	100
evt	15	20
pas d'evt	135	120
	10%	20%

RR 0.50 [0.27 ; 0.93]
P value 0.028



Scénario 3
MNAR 33% groupe C
sur evt seulement

	groupe Exp N=150	groupe C N=150
DM	0	10
évaluables	150	140
evt	15	20
pas d'evt	135	120
	10%	14%

RR 0.70 [0.37 ; 1.31]
P value 0.26

16

16

Test de l'hypothèse MCAR Est-ce que les données manquent de façon aléatoire ?

Surtout approprié pour données quantitatives

H0 : les données sont de type MCAR

Little RJA. J Am Stat Assoc 1988

H1 : les données ne sont pas MCAR

Principe du test

Grouper les individus en K sous-groupes de profils de valeurs manquantes distincts

Distribution asymptotique des moyennes et variances conditionnelles aux K groupes de profils selon une loi du χ^2

Règle de décision

$p < 0.05$: on rejette H0, les données ne sont pas MCAR

$p > 0.05$: aucune preuve pour suggérer que le mécanisme est MAR ou MNAR

Test partiel

N'indique pas quelles variables ne sont pas MCAR, et on ne sait pas si elles sont MAR ou MNAR...

17

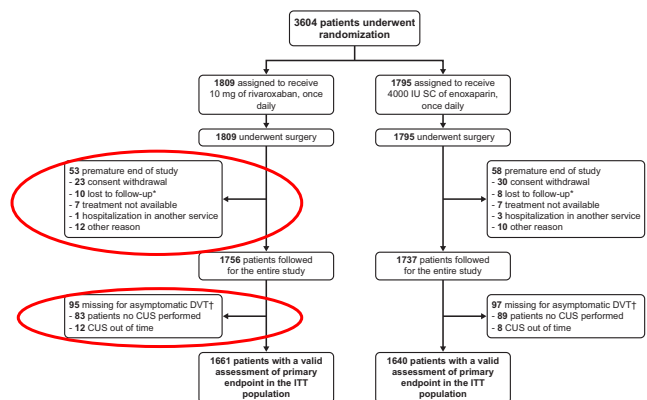
17

Rivaroxaban or Enoxaparin in Nonmajor Orthopedic Surgery The PRONOMOS study

N Engl J Med 2020

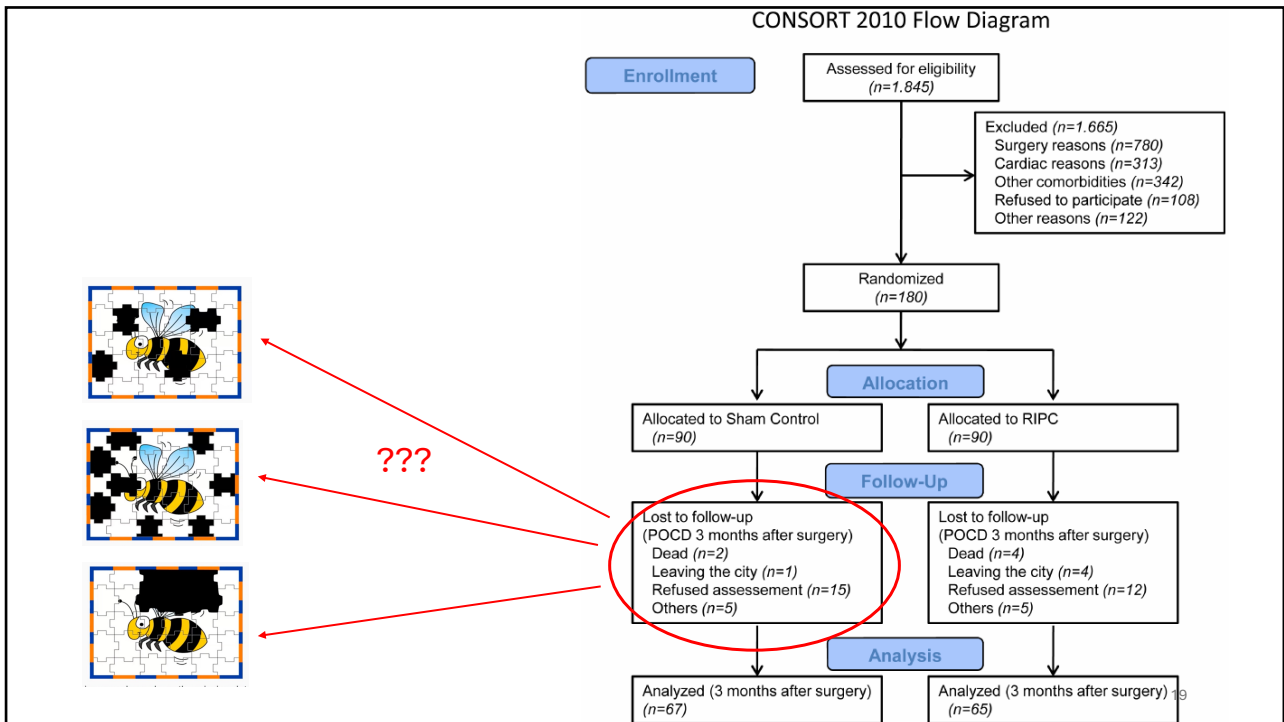


???



18

18



19

Discussion

Plusieurs mécanismes pour les données manquantes :

- Connus
- Inconnus

Connaître les raisons peut s'avérer très utile pour prévenir les manques

STOP à la génération de données manquantes

20

20

L'abus de langage « sortie d'étude »

Arrêt d'étude	Patient qui arrête prématurément le protocole - plus de prise de sang, d'EKG, de scanner... - plus de questionnaire
Arrêt de traitement	Patient qui arrête prématurément le traitement de l'étude - survenue d'événements indésirables - évolution de la maladie - résultat insuffisant ou satisfaisant...
Écart au protocole	Patient qui n'a pas été traité / suivi en conformité avec le protocole
Inclus à tort	Patient inclus alors qu'il ne présentait pas tous les critères d'inclusion ou présentait au moins un critère d'exclusion
Perdu de vue	Patient dont le devenir est inconnu et pour lequel le critère de jugement est indisponible
Non évaluable	Patients pour lequel le critère de jugement est indisponible Retrait du consentement au traitement des données

**Sortie
d'étude**

21

21

L'abus de langage « sortie d'étude »

Arrêt d'étude	Patient qui arrête prématurément le protocole - plus de prise de sang, d'EKG, de scanner... - plus de questionnaire
Arrêt de traitement	Patient qui arrête prématurément le traitement de l'étude - survenue d'événements indésirables - évolution de la maladie - résultat insuffisant ou satisfaisant...
Écart au protocole	Patient qui n'a pas été traité / suivi en conformité avec le protocole
Inclus à tort	Patient inclus alors qu'il ne présentait pas tous les critères d'inclusion ou présentait au moins un critère d'exclusion
Perdu de vue	Patient dont le devenir est inconnu et pour lequel le critère de jugement est indisponible
Non évaluable	Patients pour lequel le critère de jugement est indisponible Retrait du consentement au traitement des données

**Critère
d'évaluation
mesurable si
suivi poursuivi**

**Donnée
manquante**

22

22

Et pour les arguments concernant le maintien des données d'un patient dans l'étude

https://health.ec.europa.eu/document/download/c3042973-b36d-4094-a1fb-a6fc980f065e_en

Q6.

23