

Les techniques de dépistage de la fraude

Marianne Paesmans – Lieveke Ameye

Erreurs versus fraude

- Contrôle qualité des données et validation d'une base de données avant analyse sont des étapes standard lors de la réalisation d'un essai clinique
- Monitoring manuel du CRF
- Revue centrale des évaluations
- Monitoring de la base de données
- Monitoring sur site : comparaison du CRF aux données source (qui ne sont pas nécessairement correctes)
- Monitoring statistique
- Objectif : contrôler l'intégrité des données et la compliance au protocole
- Tâche plus complexe en cas de fraude, caractérisée par une intention délibérée de déguiser la réalité
- Monitoring statistique : aussi efficace et moins coûteux que le monitoring sur site ?

La fraude peut se situer à plusieurs niveaux

- Au niveau des données (falsification, invention)
- A niveau de l'analyse : exclusion de patients, de données
- Au niveau de ce que l'on choisit de rapporter
- Au niveau de l'interprétation et de la rédaction de la conclusion d'un article

- Avec un continuum entre erreur et fraude

- Motivations : carriériste, financière, ...

Données : continuum entre erreur et fraude

Type	Exemples	Caractère intentionnel ?
Erreur	Matériel de mesure mal calibré Précision de la mesure non respectée Erreur d'unité (mg ou mg/m ²)	Non intentionnel

Données : continuum entre erreur et fraude

Type	Exemples	Caractère intentionnel ?
Erreur	Matériel de mesure mal calibré Précision de la mesure non respectée Erreur d'unité (mg ou mg/m ²)	Non intentionnel
Paresse (incompétence)	Oubli de données, données mal retranscrites, report de données sur des évaluations ultérieures	Pas de réelle intention de tricher

Données : continuum entre erreur et fraude

Type	Exemples	Caractère intentionnel ?
Erreur	Matériel de mesure mal calibré Précision de la mesure non respectée Erreur d'unité (mg ou mg/m ²)	Non intentionnel
Paresse (incompétence)	Oubli de données, données mal retranscrites, report de données sur des évaluations ultérieures	Pas de réelle intention de tricher
Fraude	Données falsifiées (pour rendre un patient éligible par exemple), données isolées inventées, patients inventés	Intention délibérée de manipuler la réalité

Données : continuum entre erreur et fraude

Type	Exemples	Caractère intentionnel ?
Erreur	Matériel de mesure mal calibré Précision de la mesure non respectée Erreur d'unité (mg ou mg/m ²)	Non intentionnel
Paresse (incompétence)	Oubli de données, données mal retranscrites, report de données sur des évaluations ultérieures	Pas de réelle intention de tricher
Fraude	Données falsifiées (pour rendre un patient éligible par exemple), données isolées inventées, patients inventés	Intention délibérée de manipuler la réalité
Fraude liée au traitement	Données fabriquées ou falsifiées dans le but de favoriser un des traitements	Intention délibérée de manipuler la réalité pour arriver à un résultat précis

Source: IDDI webinar – Central statistical monitoring of clinical trials

Données : continuum entre erreur et fraude

Type	Exemples	Caractère intentionnel ?	Impact
Erreur	Matériel de mesure mal calibré Précision de la mesure non respectée Erreur d'unité (mg ou mg/m ²)	Non intentionnel	<ul style="list-style-type: none"> • Possible perte de puissance • Pas de biais
Paresse (incompétence)	Oubli de données, données mal retranscrites, report de données sur des évaluations ultérieures	Pas de réelle intention de tricher	<ul style="list-style-type: none"> • Possible perte de puissance • Pas de biais
Fraude	Données falsifiées (pour rendre un patient éligible par exemple), données isolées inventées, patients inventés	Intention délibérée de manipuler la réalité	<ul style="list-style-type: none"> • Effet sur la puissance ? • Pas de biais
Fraude liée au traitement	Données fabriquées ou falsifiées dans le but de favoriser un des traitements	Intention délibérée de manipuler la réalité pour arriver à un résultat précis	BIAIS

Caractéristiques et analyse de données authentiques

- Les méthodes statistiques devraient permettre de distinguer effets véritables et effets aléatoires
- Trois éléments importants :
 - Variabilité entre sujets
 - Amplitude de l'effet
 - Taille de l'échantillon
- -> Pour frauder :
 - Diminuer la variabilité entre sujets
 - Augmenter l'amplitude de l'effet
 - Augmenter la taille de l'échantillon
 - Sauf pour une étude d'équivalence !

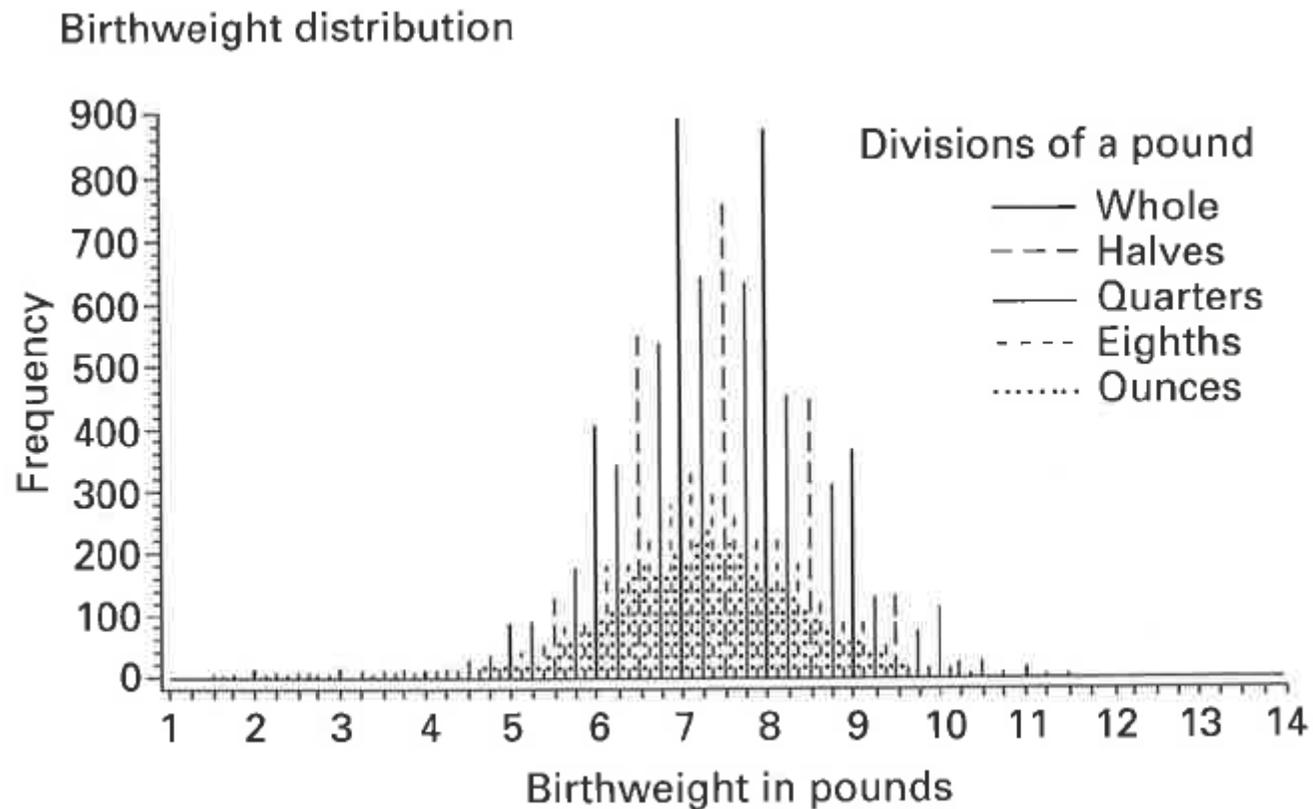
Caractéristiques des données erronées/falsifiées/inventées

- Dépendent de la sophistication des méthodes utilisées pour inventer les données
- Caractéristiques des distributions univariées risquent d'être modifiées par rapport aux distributions de données authentiques :
 - Préférence de la décimale ou des résultats arrondis
 - Distributions plus « plates » qu'attendu
 - Tendance à essayer de répliquer une moyenne -> trop d'observations proches de la moyenne et peu de variabilité autour de la moyenne
 - Peu ou pas de valeurs aberrantes / trop de valeurs aberrantes
- Structure des distributions multivariées et/ou tendances chronologiques sont sensibles aux erreurs (intentionnelles ou non) et sont très difficiles à imiter (fraude)

Méthodes de détection

- Demandent une connaissance du contexte
- Dépendent du type de données erronées/falsifiées
- Univariées ou multivariées
- Méthode très simple en cas de CRF manuel : contrôler l'écriture du CRF : écriture régulière, sans ratures, utilisation du même stylo à travers tout le CRF sont des éléments suspects !
- Exemples

Arrondis et décimale de préférence

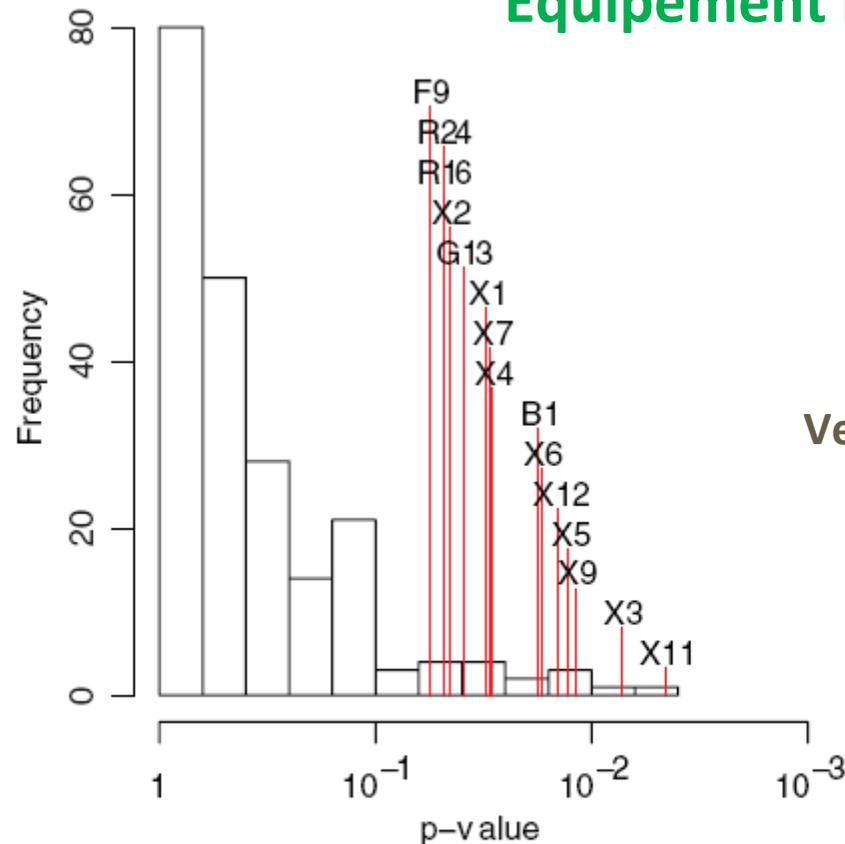


Source – National child development survey

Suspect de retrouver ce genre de distribution pour une variable continue dont la mesure est automatisée

Mauvaise calibration : comparaison des moyennes de chaque centre à celles de tous les autres centres

Equipement mal calibré dans le pays X



Venet et al, Clinical Trials, 2012

Figure 3. Frequency distribution of p -values obtained by comparing a feature mean across centers. Centers are identified by a letter (denoting the country) and a number. Vertical lines indicate the p -values of specific centers (the height of these lines is unimportant). Country X clearly shows a large number of outlying centers.

Falsification des critères d'éligibilité

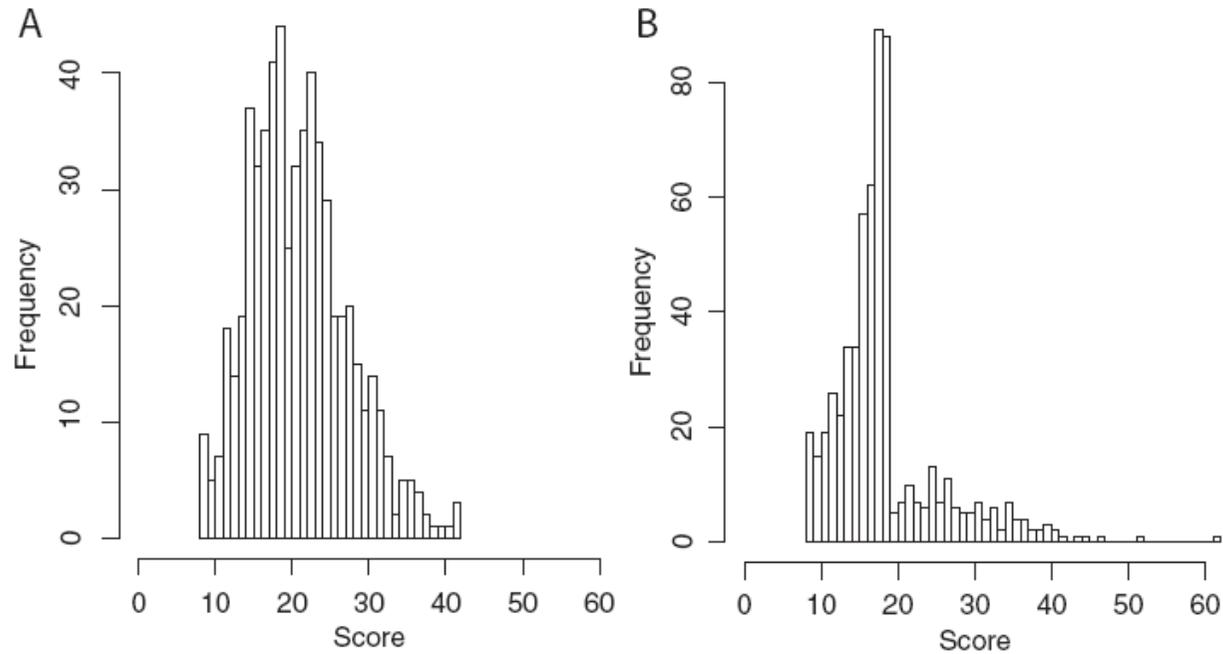


Figure 1. Distribution of psychiatric scores at entry in open-label phase (panel A) and prior to randomization (panel B). Patients with a psychiatric score below 20 were eligible.

Réduction de variabilité dans les mesures répétées

Patient X dans un centre suspect Y

Visit	Systolic blood pressure (mm Hg)	Diastolic blood pressure (mm Hg)	Respiratory rate (breath/min)
Baseline	130	80	16
Visit 1 – pretreatment	130	80	16
Visit 2 – pretreatment	130	80	16
Visit 3 – pretreatment	130	80	16
Visit 4 – pretreatment	130	80	16
Visit 5 – pretreatment	130	80	16
Visit 6 – pretreatment	120	80	16
Visit 7 – pretreatment	120	80	16
Visit 8 – pretreatment	120	80	16

Venet et al, Clinical Trials, 2012

Ceci suggère de comparer les variances intra-patient dans le centre Y et les autres centres

Analyse de tendances chronologiques peuvent identifier un schéma déterministique dans les valeurs -> suspicion de données générées à l'aide d'un algorithme automatisé (Scherrer B, La Fraude dans les Essais Cliniques, Paris 1991)

Evaluer les coefficients d'aplatissement (moments d'ordre 4)

- Permettent de repérer (à dispersion égale) :
 - L'absence d'outliers
 - Les distributions trop uniformes (platicurtiques)

Vérification des dates

- Fréquence des visites ou examens de routine le dimanche ou un jour férié
- Traitements alloués par jour, semaine ou mois de randomisation
- Intervalles entre examens :

Table IV. Frequency distribution of the number of self-evaluation days in suspect centre and in other centres.⁷¹ The patients were supposed to perform a self-evaluation and report it in a diary every day for three weeks (21 days)

Number of self-evaluation days	Suspect centre	Other centres
18 or less	—	0·13
19	—	0·04
20	—	0·10
21	0·98	0·10
22	0·01	0·13
23	0·01	0·17
24	—	0·10
25	—	0·23

Buyse et al,
Stat Med 1999

Méthodes multivariées

- Se basent sur le fait que les associations entre variables sur données inventées sont plus faibles (différentes) que sur des données réelles
- Méthodes bivariées : examiner matrice de corrélations globalement et éventuellement par centre, se référer à des données historiques
 - Exemple réel (BMJ Books, 2001) : modèle animal sur l'infarctus du myocarde : données inventées : mauvaise corrélation entre le poids des chiens et le poids de leur ventricule gauche

Détection d'inliers : distance de Mahalanobis

- Cherche à identifier des données inventées en utilisant la technique de répliquer la moyenne
- Un individu peut être proche de la moyenne sur une variable mais pas sur toutes les variables
- Distance entre un point et la moyenne de la distribution multivariée
- Distribution de cette distance suit une distribution chi carré

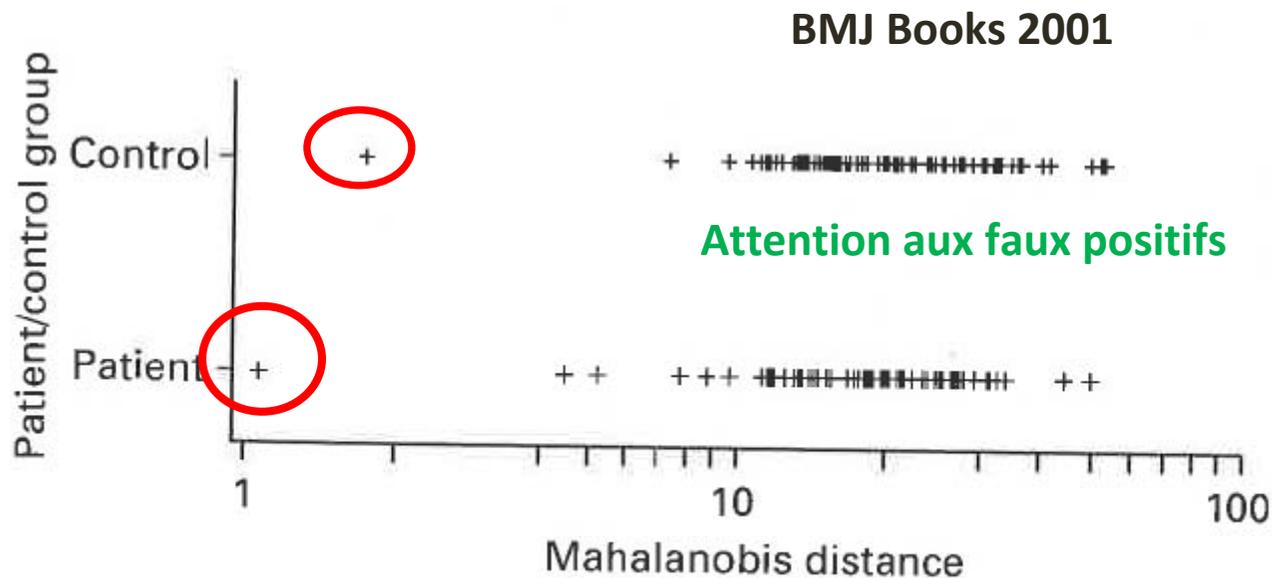
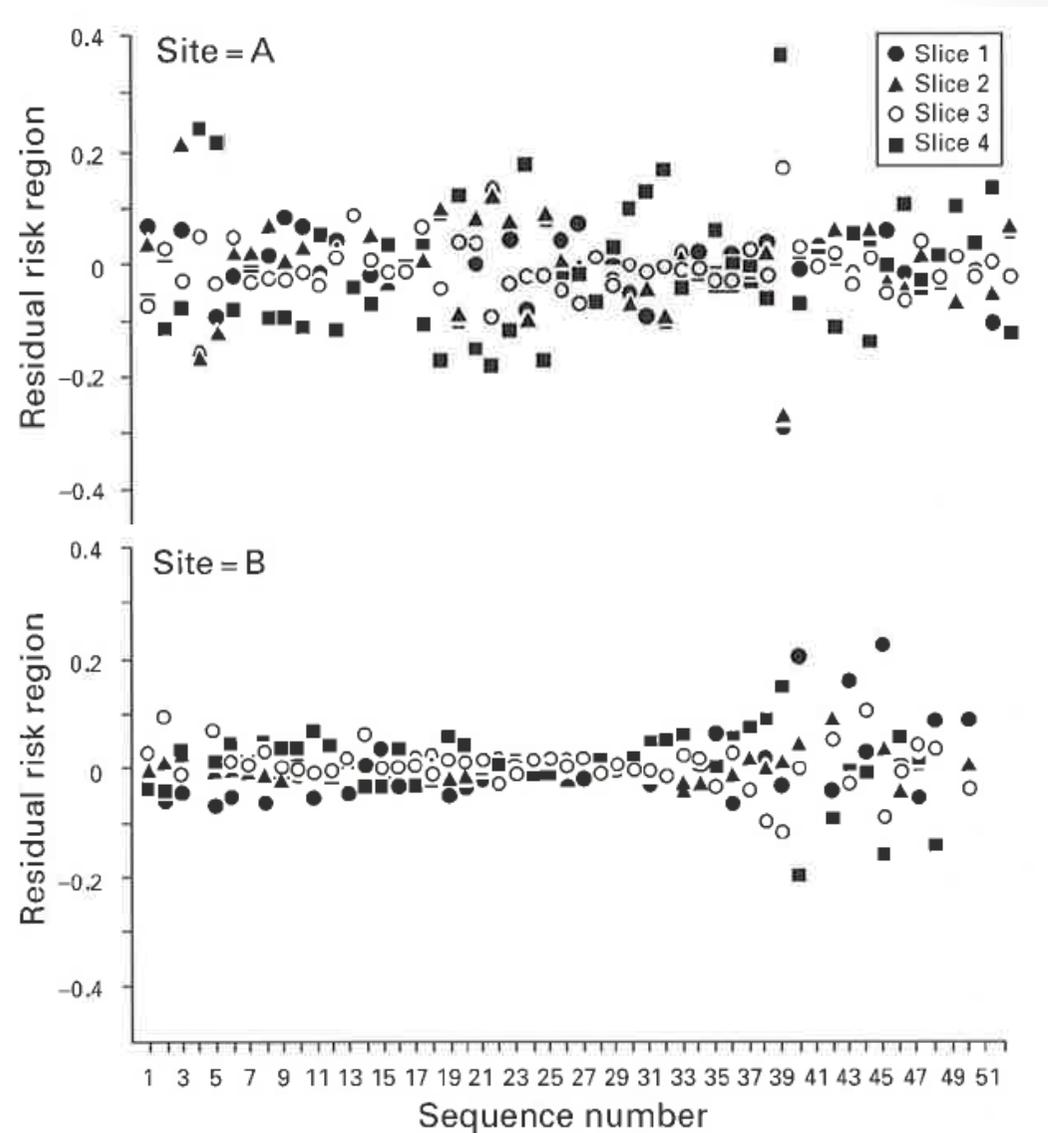


Figure 14.4 Distribution of Mahalanobis distance for a set of data to which two “inliers” have been added.

Examen des résidus d'un modèle

- Variabilité des résidus pourrait être plus faible sur données inventées

BMJ Books 2001



Cluster analysis

- Permet de détecter des proximités entre patients
- Si certains patients sont trop proches :
 - Suspicion de données inventées obtenues en construisant des variantes proches d'un patient réel
 - Suspicion de données construites en envoyant au labo des échantillons du même patient comme s'il s'agissait de patients différents

Analyse en composantes principales

- Pour chaque centre, construction d'un vecteur contenant les probabilités de signification comparant les moyennes de variables sélectionnées
- Centre X suspect au départ (audit précédent avait mis en évidence que le personnel du centre remplissait les questionnaires qui auraient dû être remis aux patients)

Venet et al, Clinical Trials, 2012

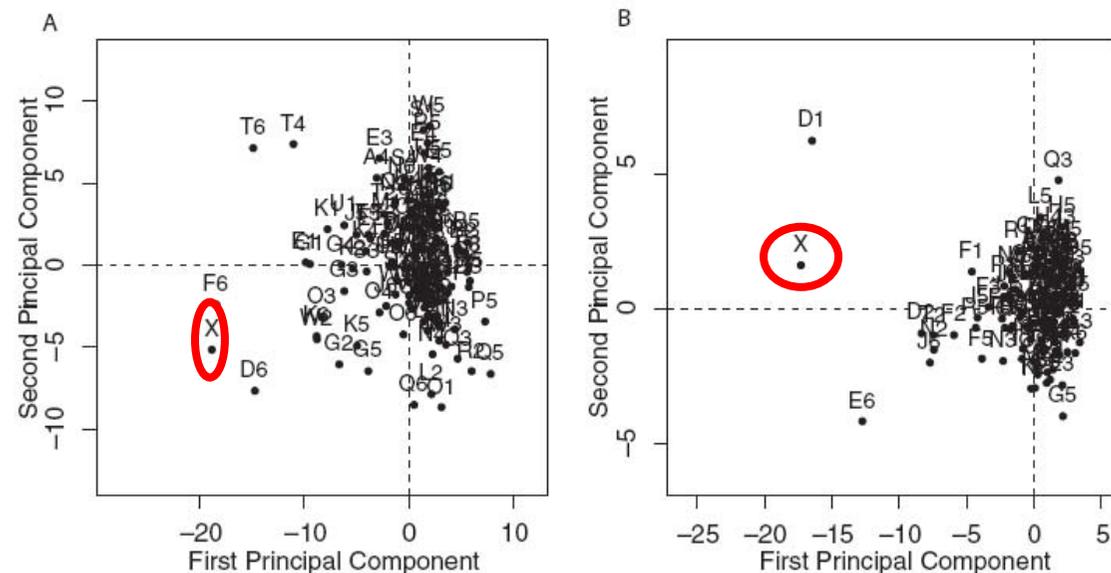


Figure 2. Principal component analysis of p -values. Centers are identified by a letter (denoting the country) and a number, except for center X where fraud had been identified. Centers far from the origin behave differently from the bulk of centers around the origin and are therefore more likely to be problematic. Panel A: analysis based on all data. Panel B: analysis based on patient-questionnaire data. PCA: Principal component analysis.

Quelles méthodes pour les reviewers et les éditeurs de journaux scientifiques ?

- Prêter attention à la dispersion des variables étudiées
 - Refuser les papiers qui ne la rapportent pas, préférer un intervalle interquartiles plutôt que l'étendue
 - La déduire éventuellement de probabilités de signification et vérifier son ordre de grandeur
- Encourager les présentations graphiques sans perte d'information (scatter plots)
- Vérifier la plausibilité de recrutement d'un auteur qui publie beaucoup sur le même sujet
- Comparer les résultats d'études portant sur le même sujet

Exemple : prévention de maladies cardiovasculaires

Al-Marzouki et al (BMJ 2005)

- Deux essais cliniques avec 2 interventions différentes
 - Régime enrichi en fruits et légumes (A)
 - Médicament (B)
- Réserves émises par les reviewers pour l'étude évaluant le régime (incohérences entre les probabilités de signification d'une part et les moyennes, écarts-type et tailles d'échantillon d'autre part)
- Etude A : étude randomisée en simple aveugle, chez des patients à risque de maladie coronarienne, 831 patients, évaluation à 2 ans
- Requête du BMJ d'avoir accès aux données individuelles (fournies sur papier)
- Etude B : étude randomisée chez des patients avec hypertension modérée, 21750 patients, non suspecte, 5 centres sélectionnés au hasard (n=838), évaluation à 2 ans
- Requête du BMJ d'avoir accès aux données individuelles (fournies sur support électronique)

Comparaison des caractéristiques baseline dans les 2 études

Table 1 Baseline variables in the two trials under comparison

	Diet		Drug	
	Intervention	Control	Intervention	Control
Weight (kg):				
Mean	65.74	65.59	70.27	70.08
Median	66	66	70	69
Mode	65	65	70	61
SD	7.89	7.64	11.6	12.4
Min	40	39	40	36
Max	87	85	111	120
Height (cm):				
Mean	165.1	165.28	162.1	162.6
Median	165	165	160	163
Mode	165	165	160	157
SD	6.91	3.93	9.22	9.14
Min	143	143	138	140
Max	179	178	190	188
Systolic blood pressure (mm Hg):				
Mean	134.2	131.9	184.4	184.6
Median	130	130	185	184
Mode	130	130	186	181
SD	18.5	16.9	12.2	12.9
Min	100	100	160	160
Max	200	195	209	210
Diastolic blood pressure (mm Hg):				
Mean	86.5	86.7	91.8	91.2
Median	86	85	92	91
Mode	80	85	101	90
SD	9.98	9.2	10.8	11.4
Min	60	60	46	50
Max	112	120	114	115
Cholesterol (mmol/l):				
Mean	5.46	5.43	6.68	6.57
Median	5.48	5.48	6.6	6.5
Mode	5.42	5.42	6.4	6.1
SD	0.352	0.296	1.26	1.21
Min	4.52	2.95	3.6	3.7
Max	6.52	6.00	12	10.8

Table 2 Baseline comparison of the two intervention groups, diet trial and drug trial

	Diet trial				Drug trial			
	Levene's F test for equality of variances		t test for equality of means		Levene's F test for equality of variances		t test for equality of means	
	F	Significance	t	Significance (two tailed)	F	Significance	t	Significance (two tailed)
Height	71.15	1.4x10 ⁻¹⁶	-0.508	0.612	0.054	0.82	-0.12	0.411
Weight	0.204	0.652	0.284	0.776	2.46	0.12	0.27	0.82
Systolic blood pressure	4.81	0.029	1.89	0.06	0.001	0.97	0.206	0.84
Diastolic blood pressure	4.366	0.037	-0.27	0.788	0.001	0.97	-0.679	0.497
Cholesterol	28.77	1x10 ⁻⁷	1.19	0.235	0.27	0.61	-1.22	0.22
Fasting blood glucose	8.21	0.004	-0.57	0.566	—	—	—	—
Total cholesterol	0.043	0.835	-0.35	0.729	—	—	—	—
Triglycerides	21.98	3x10 ⁻⁸	0.484	0.628	—	—	—	—
Energy	0.98	0.322	-1.57	0.118	—	—	—	—
Total carbohydrate	1.97	0.161	0.236	0.814	—	—	—	—
Complex carbohydrate	12.86	0.0004	14.8	6x10 ⁻⁴⁴	—	—	—	—
Protein	15.18	0.0008	5.02	6x10 ⁻⁷	—	—	—	—
Fat	20.5	7x10 ⁻⁸	-2.88	0.004	—	—	—	—
Saturated	15.2	0.0001	3.9	0.0002	—	—	—	—
Fibre	94.23	4x10 ⁻²⁵	-8.47	2x10 ⁻¹⁶	—	—	—	—
Soluble fibre	10.13	0.002	-6.95	7x10 ⁻¹²	—	—	—	—
Caffeine	2.41	0.121	0.957	0.339	—	—	—	—
Salt	39.72	5x10 ⁻¹⁰	-0.377	0.706	—	—	—	—
Vitamin C	0.007	0.991	-5.6	2x10 ⁻⁸	—	—	—	—
Carotene	51.06	2x10 ⁻¹²	29.8	2x10 ⁻¹³³	—	—	—	—
Vitamin E	25.7	5x10 ⁻⁷	5.9	5x10 ⁻⁹	—	—	—	—
Vitamin A	51.42	2x10 ⁻¹²	4.49	8x10 ⁻⁶	—	—	—	—

Aucune comparaison significative, 10 comparaisons

16 différences significatives pour la comparaison des variances, 10 pour la comparaison des moyennes, probabilités de signification extrêmes; 22 variables testées

Décimale de préférence

Table 3 χ^2 value (with P value) for the final digit at baseline, diet trial and drug trial

	Diet trial*		Drug trial	
	Intervention	Control	Intervention	Control
Height	—	—	239 (1.8×10^{-46})	251 (7.2×10^{-49})
Weight	128 (4×10^{-23})	23 (0.00655)	7.3 (0.60)	6.5 (0.69)
Systolic blood pressure	1796 (U)	1470 (U)	7.6 (0.58)	9.1 (0.43)
Diastolic blood pressure	763 (2×10^{-158})	820 (1×10^{-170})	8.1 (0.52)	13.8 (0.13)
Cholesterol	554 (2×10^{-113})	430 (6×10^{-87})	16.23 (0.062)	5.76 (0.76)
Fasting blood glucose	478 (4×10^{-87})	538 (5×10^{-110})	—	—
Total cholesterol	1053 (6×10^{-221})	1522 (U)	—	—
Triglycerides	642 (2×10^{-132})	963 (2×10^{-201})	—	—
Energy	2151 (U)	2630 (U)	—	—
Total carbohydrates	207 (1×10^{-39})	927 (7×10^{-194})	—	—
Complex carbohydrates	231 (1×10^{-44})	939 (3×10^{-196})	—	—
Protein	54 (2×10^{-5})	251 (5×10^{-49})	—	—
Fat	229 (2×10^{-44})	437 (2×10^{-88})	—	—
Saturated	123 (4×10^{-22})	98 (4×10^{-17})	—	—
Fibre	263 (2×10^{-51})	1127 (9×10^{-237})	—	—
Soluble fibre	273 (1×10^{-53})	1086 (6×10^{-228})	—	—
Caffeine	613 (3×10^{-126})	694 (1×10^{-143})	—	—
Salt	288 (9×10^{-57})	301 (2×10^{-59})	—	—
Vitamin C	304 (5×10^{-60})	411 (6×10^{-83})	—	—
Carotene	1470 (U)	1156 (5×10^{-243})	—	—
Vitamin E	118 (3×10^{-21})	101 (8×10^{-18})	—	—
Vitamin A	705 (6×10^{-146})	799 (3×10^{-166})	—	—

The χ^2 value has 9 degrees of freedom.

* U means that the P value is too small for calculation.

Décimale de préférence : différence entre les bras de traitement ?

Table 4 χ^2 value (with P value) for the final digit at the baseline in the diet and drug trials between the two randomised groups

	Diet trial		Drug trial	
	χ^2 test (P value)	df	χ^2 test (P value)	df
Height	—	—	5 (0.83)	9
Weight	36 (3×10^{-5})	9	10 (0.31)	9
Systolic blood pressure	26 (0.00019)	6	7 (0.69)	9
Diastolic blood pressure	16 (0.046)	8	10 (0.38)	9
Cholesterol	13 (0.182)	9	7 (0.60)	9
Fasting blood glucose	12 (0.2)	9	—	
Total cholesterol	46 (5×10^{-7})	9	—	
Triglycerides	48 (3×10^{-7})	9	—	
Energy	16 (0.064)	9	—	
Total carbohydrate	154 (2×10^{-28})	9	—	
Complex carbohydrate	135 (1.4×10^{-24})	9	—	
Protein	43 (2×10^{-6})	9	—	
Fat	40 (6.4×10^{-6})	9	—	
Saturated	15 (0.08)	9	—	
Fibre	157 (8×10^{30})	8	—	
Soluble fibre	175 (6.5×10^{33})	9	—	
Caffeine	15 (0.059)	8	—	
Salt	28.5 (0.001)	9	—	
Vitamin C	18 (0.03)	9	—	
Carotene	10 (0.266)	8	—	
Vitamin E	20 (0.017)	9	—	
Vitamin A	9.5 (0.4)	9	—	

The degrees of freedom are less than 9 when one or more digits do not appear.

Interprétation :

- Même sans regarder les critères d'évaluation :
 - Données de l'étude A ne proviennent pas d'une étude randomisée
- Fraude portant sur le processus de randomisation ?
- Données totalement falsifiées ?

Conclusions :

- Il existe des méthodes autres que le monitoring sur site traditionnellement utilisé
- A utiliser probablement plus souvent
- Ne sont vraiment utiles qu'avec des nombres suffisants de patients et de patients / centre -> peu pertinentes en début d'étude, peu utiles pour de petits centres
- Il existe des différences authentiques entre centres qui ne doivent pas être niées
- Ces méthodes peuvent induire une suspicion mais ne sont pas des preuves