

ABSENCE ET ABSENTÉISME LORS DES ENQUÊTES SCOLAIRES : UN EXEMPLE DE CORRECTION DE LA NON- RÉPONSE NON-IGNORABLE DANS ESPAD 2015

Stéphane Legleye¹, Stanislas Spilka², François Beck³

1. *Insee, division Recueil et traitement de l'information, CESP, Inserm*
2. *OFDT*
3. *Insee*

Contexte

Dans les enquêtes scolaires, le taux de réponse est très élevé

- Redressement simple : calage sur les effectifs par sexe et niveau

MAIS

- Environ 10-15% des élèves sont absents le jour donné (Hibell et al., 2012)
- **Lien fort entre absentéisme et comportements déviants** (Vaughn, Maynard, Salas-Wright, Perron, & Abdon, 2013; Maynard et al., 2018)
 - Corrélations avec les absences déclarées
 - Confirmé par étude sur les non-répondants: (Michaud, Delbos-Piot, & Narring, 1998)

➔ Il faudrait tenir compte de ces absences pour redresser!

Les difficultés

- Les enquêtes scolaires mesurent les pratiques éducatives, les modes de vie et les comportements déviants (drogues, violence etc.)
- La non-réponse totale est non-ignorable (Rubin & Little, 2002) pour bon nombre de ces sujets: elle est liée au sujet d'étude
- Non-ignorable? Oui car impossible de modéliser en l'absence d'enquêtes sur les non-répondants:
 - L'absent est absent, pas de mesure pour lui (ni d'historique d'absence)
 - L'élève est soit absent, soit présent

Présentation des données: ESPAD 2015

- ESPAD (European survey project on alcohol and other drugs)
- Piloté par le CAN (conseil suédois d'information sur l'alcool et les drogues).
- Enquête quadriennale visant les élèves de 15-16 ans depuis 1995
- La France participe depuis 1999 (dernière enquête en 2015)
- Le questionnaire est unique dans tous les pays participants.
- En France, l'échantillon a été tiré par la DEPP

Mesures dans l'enquête

- Y =usages d'alcool, de tabac et de cannabis au cours des 30 derniers jours:
 - 0 fois, 1-2 fois, 3-6, 9-10, 10-19, 20-39, 40+ fois, pour alcool et cannabis
 - 0 cigarette, <1/semaine, <1/jour, 1-5/jour, 6-10/jour, 11-20/jour, 20+/jour
 - Usages réguliers:
 - usage quotidien pour le tabac
 - 10+ usages pour l'alcool et le cannabis
- X_i =fréquences d'absences passées au cours des 30 derniers jours suivant trois motifs:
 - Sèche (aucun motif)
 - Maladie
 - Autres motifs (famille etc.)
 - 0, 1 jour, 2 jours, 3-4 jours, 5-6 jours 7+ jours

Rapports de passation

- L'enquêteur dénombre, par sexe :
 - Les présents participants
 - Les présents refusant l'enquête
 - Les absents sans motif (sèche)
 - Les absents avec motif (excusés)
 - Les absents dont la légitimité est incertaine
- Possibilité de distinguer les sécheurs des autres

Description des présences (rapports de passation)

- 284 classes
 - 7169 élèves ciblés (base de sondage)
 - 6185 présents, 984 absents (dont 16 refus parentaux)
 - 6007 ont rendu un questionnaire analysable (178 exclus pour défaut de qualité): 83,7%
 - 359 sont des sécheurs (36,5% des absents, 5,0% des élèves ciblés)
- Au final:
 - 1162 absents ou questionnaires manquants, dont 31% de sécheurs

Description des répondants

nombre de jours d'absence passées par motif des répondants

NR: 5.4% pour les maladie, 8.4% pour la sèche et 7.1% pour d'autres raisons

	Maladie				Sèche (aucune raison)				Autre raison			
	n	n	%	%	n	n	%	%	n	n	%	%
	Brut	Imputé	Brut	Imputé	Brut	Imputé	Brut	Imputé	Brut	Imputé	Brut	Imputé
Manquant	356		5.8		533		8.6		457		7.4	
0	3,198	3403	51,7	55,0	4,145	4536	67,0	73,3	3,592	3913	58,1	63,3
1	949	1003	15,3	16,2	683	752	11,0	12,2	1002	1060	16,2	17,1
2	777	824	12,6	13,3	335	369	5,4	6,0	540	576	8,7	9,3
3.5	560	586	9,1	9,5	236	259	3,8	4,2	324	347	5,2	5,6
5.5	155	163	2,5	2,6	78	80	1,3	1,3	98	106	1,6	1,7
7.5	190	206	3,1	3,3	175	189	2,8	3,1	172	183	2,8	3,0
Total (jours)	6740,5	7107,5			3920,5	4325,5			5045	5422,5		
Ratio (%)	42,9%	42,2%			25,0%	25,7%			32,1%	32,1%		

Corrélations linéaires entre les absences passées et les usages de drogue

	Nombre de jours séchés	Nombre de jours d'absence (toutes causes)	Fréq. Cigarettes 30 jours (0-7)	Fréq. Alcool 30 jours (1-7)	Fréq. Cannabis 30 jours (1-7)	Tabac quotidien (0-1)	Alcool régulier (0-1)	Cannabis régulier (0-1)
Filles								
Nombre de :								
jours séchés	1.00	0.65	0.23	0.04	0.22	0.20	0.03	0.19
jours d'absence	0.65	1.00	0.20	0.04	0.16	0.17	0.06	0.14
Garçons								
Nombre de :								
jours séchés	1.00	0.68	0.27	0.12	0.25	0.24	0.10	0.21
jours d'absence	0.68	1.00	0.19	0.03	0.17	0.17	0.03	0.14

Liens entre absences (toutes causes), sèches et usages

Régression logistique contrôlant le total de nombre d'absences toutes causes, le total de jours de sèches et le sexe

- Pour le tabagisme quotidien
 - OR=**1.28** [1.19; 1.37] pour un jour de sèche supplémentaire
 - OR=1.03 [1.00; 1.07] pour un jour d'absence supplémentaire (toute cause)
- Pour l'alcool régulier
 - OR=**1.14** [1.04; 1.24]
 - OR=0.99 [0.96; 1.03]
- Pour l'usage régulier de cannabis
 - OR=**1.32** [1.21; 1.45]
 - OR=1.02 [0.97; 1.06]

Corrélation nombre de jours de sèche – nombre de jours d'absences toutes causes: $\rho=0,68$

Description des répondants: les questionnaires exclus

- Le nombre moyen de jours d'absences passées est supérieur parmi les questionnaires exclus (6.2 jours vs 2.6 jours) ainsi que la proportion de la sèche dans le total (37% vs 25%).
- Les questionnaires exclus sont donc ceux d'élèves plus souvent absents et sécheurs

Correction proposée (1)

Repondérer les présents ayant déclaré des absences passées pour:

- qu'ils se représentent eux-mêmes ainsi que les absents pour le même motif le jour de l'enquête
 - en fonction des nombres de jours d'absences passées qu'ils ont déclarés
-
- 1. CNRT faite par sexe, car sexe très lié aux usages de drogues
 - CNRT sera suivie du calage classique de l'enquête

Correction proposée (2)

2. CNRT effectuée à trois niveaux:

- Classe : le plus fin, toutes les caractéristiques liées aux usages prises en compte
- Strate (filière et secteur) : la filière (professionnelle ou classique) et le secteur (public/privé) sont fortement associés aux usages et aux absences
- Global : l'ensemble de l'échantillon (confusion des filières et secteurs)

3. CNRT effectuée en distinguant la sèche ou non au sein des absences

Correction proposée (3)

approche par les absences

Pour un niveau donné N , l'approche par les absences A se fonde sur le coefficient de repondération suivant pour l'individu i :

$$pNA_i = W_i + T(\text{absents le jour } J)_{N,W} \times \frac{W_i \times (\text{nb jours absence})_i}{T(\text{nb jours absences})_{N,W}},$$

$(\text{nb jours absence})_i$ = nombre de jours d'absence toutes causes rapporté par le répondant i

$T(\text{nb jours absences})_{N,W}$ = total de nombre de jours d'absence (pondéré) toutes causes rapportés par les répondants du niveau N

$T(\text{absents le jour } J)_{N,W}$ = nombre total d'absents au niveau N le jour J

W_i = poids de sondage

Les types d'absence ne sont pas distingués

Correction proposée (4)

approche par la sèche

Pour un niveau donné N, l'approche par la sèche S impose de distinguer un coefficient pour la correction des absences légitimes et un autre pour la sèche :

$$pNS_i = W_i + C_1(\text{absences légitimes})_i + C_2(\text{sèche})_i$$

$$pNS_i = W_i$$

$$+ T(\text{nb absents non sécheurs le jour } J)_{N,W} \times \frac{W_i \times (\text{nb jours absence légitime})_i}{T(\text{nb jours absence légitimes})_{N,W}}$$

$$+ T(\text{nb sécheurs le jour } J)_{N,W} \times \frac{W_i \times (\text{nb jours sèche})_i}{T(\text{nb jours sèche})_{N,W}}$$

- Si le dénominateur du coefficient pour la sèche vaut 0, alors $pNS_i = pNA_i$ (il n'existe aucune dans laquelle nombre total de jours d'absences déclaré par les répondants est nul)

Tronquage et renormalisation

- Les CNRT de niveau classe (*pCA* et *pCS*) conduisaient parfois à des pondérations parfois extrêmes
 - tronquage au 99^{ème} percentile et réechelonnement sur l'effectif total de la classe
- 7 classes sans questionnaire de garçon retenu, 4 classes sans questionnaire fille
- Les effectifs corrigés par la CNRT étaient donc inférieurs au total théorique (7 117 vs 7 169). Problème corrigé dans l'étape de calage

Statistiques de poids

	Min	Médiane	Moyenne	99th Pctl	Max	Somme	Coeff. de variation
Poids de sondage	0.6	0.6	1.2	2.3	2.3	7126	68.7
Poids de sondage calé	0.4	0.7	1.2	4.5	6.2	7169	76.9
pGA	0.5	0.6	1.2	3.6	6.4	7117	72.8
pGA calé	0.3	0.7	1.2	4.5	11.1	7169	81.0
pGS	0.5	0.7	1.2	3.7	6.5	7117	73.1
pGS calé	0.3	0.7	1.2	4.5	11.5	7169	81.3
pSA	0.5	0.6	1.2	3.8	5.7	7117	73.5
pSA calé	0.3	0.7	1.2	4.5	9.6	7169	81.1
pSS	0.5	0.6	1.2	3.9	6.3	7117	74.1
pSS calé	0.3	0.7	1.2	4.5	10.6	7169	81.8
pCA	0.3	0.6	1.2	4.2	11	7117	79.3
pCA calé	0.3	0.7	1.2	4.8	19.1	7169	90.7
pCS	0.2	0.6	1.2	4.3	13.6	7117	81.7
pCS calé	0.2	0.7	1.2	4.8	22.4	7169	92.6

pGA, pGS, pSA, pSS, pCA, pCS: la première lettre représente le niveau (G=Global, S=Strate, C=Classe), la seconde le niveau de détail (A=toutes absences, S=distinction de la sèche).

Résultats

Niveau global

	Calage direct		pGA calé				pGS calé			
	%	std	%	std	D	Deff	%	std	D	Deff
Tabac										
Vie	60.3	1.2	61.5	1.2	1.2	1.03	61.7	1.2	1.3	1.03
Quotidien	22.6	1.2	24.0	1.3	1.4	1.09	24.1	1.3	1.5	1.10
Alcool										
Vie	86.3	1.0	86.0	1.0	-0.3	1.02	86.1	1.0	-0.2	1.02
Mois	63.4	1.5	63.5	1.5	0.2	1.02	63.6	1.5	0.2	1.02
Régulier	14.4	1.1	14.7	1.1	0.4	1.08	14.8	1.1	0.4	1.09
Cannabis										
Vie	44.0	1.2	45.4	1.3	1.5	1.04	45.6	1.3	1.6	1.04
Mois	22.6	1.0	23.8	1.1	1.2	1.05	24.0	1.1	1.4	1.06
Régulier	7.8	0.7	8.5	0.8	0.8	1.12	8.6	0.8	0.9	1.14

D=différence avec le calage direct ; std=écart-type ; Deff= ratio (std CNRT-calé)/(std calé)

Résultats

Niveau strate

	Calage direct		pSA calé				pSS calé			
	%	std	%	std	D	Deff	%	std	D	Deff
Tabac										
Vie	60.3	1.2	61.5	1.2	1.2	1.03	61.5	1.2	1.2	1.03
Quotidien	22.6	1.2	24.0	1.3	1.4	1.09	24.0	1.3	1.3	1.09
Alcool										
Vie	86.3	1.0	86.0	1.0	-0.3	1.02	86.0	1.0	-0.3	1.02
Mois	63.4	1.5	63.6	1.5	0.2	1.01	63.5	1.5	0.2	1.02
Régulier	14.4	1.1	14.7	1.2	0.3	1.10	14.8	1.2	0.4	1.10
Cannabis										
Vie	44.0	1.2	45.4	1.3	1.5	1.04	45.4	1.3	1.5	1.04
Mois	22.6	1.0	23.8	1.1	1.2	1.06	23.8	1.1	1.2	1.05
Régulier	7.8	0.7	8.5	0.8	0.8	1.13	8.5	0.8	0.8	1.12

D=différence avec le calage direct ; std=écart-type ; Deff= ratio (std CNRT-calé)/(std calé)

Résultats

Niveau classe

	Calage direct		pCA calé				pCS calé			
	%	std	%	std	D	Deff	%	std	D	Deff
Tabac										
Vie	60.3	1.2	61.1	1.2	0.8	1.05	61.1	1.2	0.7	1.05
Quotidien	22.6	1.2	23.5	1.3	0.9	1.11	23.1	1.3	0.5	1.11
Alcool										
Vie	86.3	1.0	86.1	1.0	-0.3	1.02	86.1	1.0	-0.3	1.03
Mois	63.4	1.5	63.6	1.6	0.2	1.04	63.5	1.6	0.1	1.07
Régulier	14.4	1.1	14.8	1.2	0.5	1.13	14.8	1.2	0.4	1.12
Cannabis										
Vie	44.0	1.2	45.0	1.3	1.0	1.07	44.9	1.3	0.9	1.06
Mois	22.6	1.0	23.5	1.1	0.9	1.09	23.3	1.1	0.7	1.07
Régulier	7.8	0.7	8.2	0.8	0.4	1.13	8.1	0.8	0.3	1.12

D=différence avec le calage direct ; std=écart-type ; Deff= ratio (std CNRT-calé)/(std calé)

Bilan rapide

Hypothèses :

Les sécheurs/absents passés sont similaires aux sécheurs/absents le jour J :
raisonnable?

En particulier si le jour d'enquête peut être considéré tiré au hasard: les chances qu'un élève soit absent le jour J sont bien proportionnelles à son total de jours d'absences passées.

➔ Plus raisonnable que l'hypothèse classique de redressement/calage où les non-répondants (absents le jour J) sont identiques en moyenne aux répondants conditionnellement à des variables peu liées aux absences

Bilan rapide

- Très peu d'effet sur la consommation d'alcool
 - Interprétation : très faible corrélation avec les absences et la sèche
- Effets plus importants pour les corrections de niveau strate et global que classe
 - Interprétation : variabilité interne des profils de classe au sein des strates/tronquage au sein des classes. Les non-absents / non-sécheurs remplacent des absents/sécheurs, ce qui est moins bon.
- Peu d'écarts entre les corrections distinguant la sèche des autres types d'absence
 - Interprétation : forte corrélation entre le total des absences et les sèches ($\rho=0,68$)

Est-ce important? Significatif?

- Plan de sondage complexe équilibré (assimilable à une stratification) et en grappe (établissements, classes) \Rightarrow perte de précision
- Aucune sortie des IC95% du calage direct...
- Sans cet effet de grappe, presque toutes les estimations corrigées des niveaux de tabagisme ou d'usage régulier de cannabis sortent des intervalles de confiance des estimations classiques

Limites

Correction très prudente et conservatrice:

- Les 178 questionnaires exclus sont proches des 359 sécheurs
 - On aurait pu les considérer sécheurs, ce qui n'a pas été fait ici: le taux de sèche passerait de 31% des absents à 46%
 - Mais à partir de quel seuil de sèche passée un exclu est-il sécheur le jour J?

Incertitude sur certaines absences le jour J:

- Les sèches sont avérées mais 116 absences sont de statut incertain : une partie doit être de la sèche... mais laquelle?
- Fiabilité des déclarations d'absences passées, notamment la sèche

Recommandations et applications possibles

- Nécessite un bon recueil des motifs d'absence en classe...
- Tester les corrélations (absences, sèches, variables d'intérêt)
- Applicable à toute enquête où :
 - les motifs d'absences peuvent être relevés
 - Les motifs d'absences sont liés aux variables cibles
 - On peut questionner les répondants sur leurs absences passées
- Exemple :
 - Enquête sur les mobilités et les déplacements...

Merci pour votre attention

Références

- Hibell, B., Guttormsson, U., Ahlström, S., Balakireva, O., Bjarnason, T., Kokkevi, A., & Kraus, L. (2012). *The 2011 ESPAD Report: Substance Use Among Students in 36 European Countries*. Retrieved from Stockholm:
[http://espad.org/Uploads/ESPAD_reports/2011/The 2011 ESPAD Report FULL 2012 10 29.pdf](http://espad.org/Uploads/ESPAD_reports/2011/The_2011_ESPAD_Report_FULL_2012_10_29.pdf)
- Vaughn, M. G., Maynard, B. R., Salas-Wright, C. P., Perron, B. E., & Abdon, A. (2013). Prevalence and correlates of truancy in the US: Results from a national sample. *Journal of Adolescence*, 767-776. doi:<http://dx.doi.org/10.1016/j.adolescence.2013.03.015>
- Maynard, B. R., Vaughn, M. G., Nelson, E. J., Salas-Wright, C. P., Heyne, D. A., & Kremer, K. P. (2018). Truancy in the United States: Examining temporal trends and correlates by race, age, and gender. *Children and Youth Services Review*.
- Michaud, P. A., Delbos-Piot, I., & Narring, F. (1998). Silent dropouts in health surveys: are nonrespondent absent teenagers different from those who participate in school-based health surveys? *J Adolesc Health*, 22(4), 326-333.
- Rubin, D. B., & Little, R. J. A. (2002). *Statistical Analysis with Missing Data*: John Willey & Sons.