

ABSENCE ET ABSENTEISME LORS DES ENQUETES SCOLAIRES : UN EXEMPLE DE CORRECTION DE LA NON-REPONSE NON-IGNORABLE DANS ESPAD 2015

Stéphane Legleye¹, Stanislas Spilka², François Beck³

¹ *Insee, 88 Avenue Verdier, 92120 Montrouge, France ; CESP, Inserm, Hôpital Paul Brousse, 12 Avenue Paul Vaillant Couturier, 94800 Villejuif*

² *OFDT, 3 avenue du Stade de France, Saint-Denis, France*

³ *Insee, 88 Avenue Verdier, 92120 Montrouge, Franc*

Résumé.

La recherche sociologique montre qu'absentéisme scolaire (« sèche ») et usages de drogues sont fortement liés. Or, dans les enquêtes scolaires, les usages des absents (sêcheurs compris) sont inconnus et l'absentéisme impossible à modéliser. C'est donc une non-réponse non-ignorable, conduisant à la sous-estimation des usages. Nous proposons une correction utilisant les nombres d'absents par classe (sêcheurs ou excusés) et les nombre de jours d'absences au cours des 30 derniers jours par motif (sèche, maladie, autre) déclarés par les répondants. Les absents sont représentés par les répondants, proportionnellement à leur nombre d'absences passées déclarées sur le total. Les données sont issues du volet français d'Espad 2015 (European survey project on alcohol and other drugs). Sur les 7169 élèves des 284 classes sélectionnées, 6007 ont répondu : 1162 étaient absents, (16%), dont 359 sècheurs (31% des absents). La correction est opérée par sexe, à la classe, à la strate ou globalement, en distinguant ou non la sèche. Après calage, la distinction des types d'absence apparaît négligeable ; les corrections augmentent surtout les estimations nationales de tabagisme quotidien (de 22.6% à 24.1%) et d'usage régulier de cannabis (de 7.8% à 8.6%). En raison du sondage en grappe, aucune ne sort des intervalles de confiance des estimations classiques, mais la proportion de sècheurs semble être un indicateur de qualité des estimations. La méthode est applicable dans d'autres enquêtes où les comportements étudiés sont liés aux absences, pourvu que les motifs d'absence des personnes sélectionnées et les absences passées des répondants soient recueillis.

Mots-clés. Enquête scolaire, non-réponse non-ignorable, redressement, absentéisme

Abstract.

Sociological research has proved that school absenteeism (truancy) and drug use are strongly linked. In school surveys, drug use of the absentees (including truants) are unknown and absenteeism cannot be modelled: truancy is thus a non-ignorable non-response, leading to the underestimation of drug use. We propose a correction based on the numbers of absents by class (truants and others) and on the report by the respondents of their numbers of past absences in the last 30 days by motive (truancy, illness, other). Absents are represented by respondents in proportion to their number of past absences. Data come from the 2015 French Espad survey (European survey project on alcohol and other drugs). Out of the 7169 pupils in the 284 classes selected, 6007 responded: 1162 were absent (16%), including 359 truants (31% of those absent). The correction is done by gender at three levels: class, stratum and on the entire sample, and by distinguishing or not truancy from legitimate absences. After calibration, the distinction between the types of absences is negligible; all corrections increased the national estimates of daily smoking (from 22.6% to 24.1%) and regular cannabis use (from 7.8% to 8.6%). Because of the clustered sample design, no estimate fell outside the confidence intervals of the conventional estimates, but result suggest that the proportion of truants is an indicator to monitor. The method is applicable in other surveys, as soon as the reasons for absence of the selected persons

and the past absences of the respondents are collected.

Keywords. School surveys, drug use, non-ignorable non-response, weighting, truancy, school absenteeism

1. Introduction

Il est difficile de proposer une définition simple et universelle de la non-réponse dans les enquêtes en raison des difficultés pratiques rencontrées sur le terrain et de la multitude de cas possibles (Skalland 2011). Par ailleurs, le taux de réponse est un indicateur médiocre de la qualité d'une collecte (Groves 2006, Davern, McAlpine et al. 2010), même si bien sûr, un taux proche de 100% (souvent irréaliste) laisse peu de place à des biais. Dans les cas réalistes, les problèmes surgissent lorsque les comportements à estimer sont liés à la non-réponse. Par exemple, si la participation à une enquête sur le revenu est liée au niveau de revenu, alors les estimations de moyennes et de totaux de revenus dans la population seront biaisées même après un redressement classique (sauf si l'on peut estimer le revenu des non-répondants via des informations auxiliaires). Cette situation est celle de la non-réponse non-ignorable (Little and Vartivarian 2005). En pratique, il n'est pas simple de vérifier que la non-réponse est de ce type et la plupart des redressements reposent sur l'hypothèse que ce n'est pas le cas.

Dans les enquêtes scolaires, les problèmes pratiques de définition d'une non-réponse sont apparemment simples et le fort niveau de participation généralement observé peut laisser croire qu'elles sont immunisées contre des biais importants de non-réponse. En effet, le taux de participation est souvent proche de 100%. Mais en fait, cette statistique est trompeuse car si effectivement très peu d'élèves refusent de remplir un questionnaire, une part plus substantielle est absente le jour de l'enquête : entre 10% et 15% voire plus, suivant les pays (Hibell, Guttormsson et al. 2012, Guttormsson, Leifman et al. 2016)(page 18). Or si les absences sont liées aux comportements à estimer, et si en pratique ces comportements sont concentrés de façon disproportionnée parmi les absents, alors le biais de non-réponse risque d'être important.

Dans le rapport de l'enquête ESPAD (European survey project on alcohol and other drugs) de 2003, les chercheurs ont noté une corrélation positive entre le nombre de jours d'absence illégitime (sèche) déclarés par les répondants et leurs niveaux d'usages d'alcool, de tabac et de cannabis au sein des 35 pays participants : (Hibell, Andersson et al. 2004)(page 198). Cette étude méthodologique fait écho à des travaux de recherche en éducation, montrant un lien entre abandon scolaire, sèche et absentéisme et usages de drogues (Wichstrom 1998, Henry and Thornberry 2010, Mounteney and Skutle 2010, Holtes, Bannink et al. 2015), lien qui n'est pas propre aux pays occidentaux (Yoep, Kuang Kuay et al. 2016). La réinterrogation des absents le jour de l'enquête, ou celle de leurs parents, prouve la solidité de ces résultats (Michaud, Delbos-Piot et al. 1998, McAra 2004). Cette non-réponse est très suspecte d'être de type non-ignorable car une partie des absents sont absents en raison de leurs comportements déviants, au nombre desquels leurs usages de drogues.

Corriger ce type de non-réponse implique de pouvoir disposer d'informations auxiliaires permettant de modéliser la sèche.

1.1 Non-réponse et absences dans les enquêtes scolaires

La non-réponse dans les enquêtes scolaires existe en fait à plusieurs niveaux : d'abord au niveau établissement, puis au niveau classe et enfin au niveau élève. Au niveau établissement et classe, elle est due le plus souvent à des refus de chefs d'établissements ou de professeurs principaux, non pas à cause du sujet de l'enquête, mais à cause de la charge de travail. Sans lien avec le sujet, elle peut donc facilement être corrigée par les procédures de redressement classiques.

Au niveau individuel, la situation est plus complexe. On peut distinguer les refus des parents d'élèves et les refus des élèves, qui sont tous deux très rares et a priori sans rapport avec le sujet de l'enquête. Ensuite, il y a les absences des élèves. En France et dans de nombreux pays, les absences sont soit indiquées à l'avance aux professeurs et à l'établissement via un mot établi par les parents, soit constatées immédiatement par l'établissement dès le premier cours, ce qui donne lieu à l'envoi d'un SMS aux parents, qui peuvent alors donner une justification ou non. Autrement dit, lors d'une enquête

scolaire, dès lors qu'elle n'a pas lieu lors du premier cours, le caractère légitime ou non d'une absence constatée est connue avec un grand degré de certitude. Un autre type d'absence légitime est celle d'un groupe d'élèves pour des raisons scolaires (visites médicales, déplacement ou sortie etc.). Enfin, certains élèves rendent des questionnaires qui ne sont pas conservés pour l'analyse en raison de leur mauvaise qualité ; ces questionnaires sont alors assimilables à des questionnaires d'élèves absents, sans qu'on puisse en établir le type d'absence.

1.2 Redressement en une ou deux étapes

Les redressements classiques des enquêtes (scolaires ou non) en une étape se font par calage (Deville and Sarndal 1992). Le principe est de repondérer les questionnaires retenus dans l'analyse afin qu'ils représentent les manquants, pour retrouver le total théorique des élèves au niveau national, suivant quelques critères de base (sexe, âge, région ou académie, secteur public ou privé, filière etc.). Il n'y a pas de distinction entre les types d'absences (légitimes ou non, répétées ou non pour un même individu). L'hypothèse sous-jacente est que les absences sont indépendantes des usages de drogues et qu'absents et présents répondants ont les mêmes usages : tous les répondants représenteront les absents, quels que soient les comportements d'usage des répondants et des absents.

L'intérêt de cette approche est sa simplicité. Son défaut est de ne pouvoir tenir compte de caractéristiques individuelles liées à l'absentéisme et aux usages de drogues car aucune n'est disponible dans les bases de sondages : seuls sont pris en compte le sexe et la filière. Par conséquent, elle ne peut qu'ignorer le lien documenté entre type d'absence (et notamment sèche) et usages. L'absentéisme risque bien de rester une non-réponse non-ignorable dans ce cadre.

Il est parfois possible de procéder à une première étape avant le calage (correction de la non-réponse totale ou CNRT) si l'on peut estimer la probabilité que chaque répondant (et non-répondant) soit répondant. Cette étape peut réduire la variance et diminuer le biais des estimations si les relations entre variables mobilisées et variables cibles de l'enquête (ici usages de drogues) sont fortes (Little and Vartivarian 2005). Dans la plupart des situations, le redressement en deux temps produit de meilleures estimations (Haziza and Lesage 2016). Nous proposons de développer une approche de ce type.

1.3 Objectifs de l'étude

A l'aide des données réelles d'une enquête scolaire récente, nous proposons d'utiliser les informations collectées dans les rapports de passation au niveau de chaque classe ainsi que les historiques d'absences récentes des élèves répondants pour effectuer une correction de la non-réponse totale préalable au calage. Les estimations corrigées obtenues à l'issue de ce redressement en deux étapes seront comparées à celles issues du redressement classique obtenu par calage direct en une étape.

2. Données et méthodes

2.1. Présentation de l'enquête

L'enquête européenne ESPAD (European survey project on alcohol and other drugs) est sous la responsabilité du CAN (conseil suédois d'information sur l'alcool et les drogues). C'est une enquête quadriennale visant les élèves de 15-16 ans qui est en place depuis 1995 et à laquelle la France participe depuis 1999. Le questionnaire est unique dans tous les pays participants. En France, l'échantillon a été tiré par la Direction de l'évaluation, de la prospective et de la performance (DEPP) à partir du recensement scolaire effectué en octobre 2014. Il est équilibré (macro Cube) et vise à représenter les niveaux et filières (général et technique, professionnel) par sexe et secteur académique (public ou privé) ainsi que la zone (urbaine, suburbaine et rurale). Dans chaque établissement scolaire,

deux classes sont tirées au sort. Nous assimilerons le sondage à un sondage stratifié (par filière, niveau et secteur) en grappe pour les besoins de l'étude. La collecte a eu lieu d'avril à début juin 2015.

La collecte se déroule durant 50 mn, sur questionnaire papier, sous la supervision d'un enquêteur formé, chargé de présenter l'enquête, de répondre aux éventuelles questions et de remplir un rapport de passation. Il y dénombre les présents, les absents (suivant le motif : excusé, non excusé) ainsi que les refus, par sexe. Tous les questionnaires, remplis ou non, sont placés dans une enveloppe qui est scellée et envoyée à un opérateur de saisie. Les questionnaires mal ou insuffisamment remplis sont éliminés par l'équipe de recherche.

2.2. Mesures et indicateurs

Les répondants doivent indiquer combien de jours ils ont manqué l'école au cours des 30 derniers jours, en distinguant les cas de maladies, de sèche et enfin les autres raisons. Les modalités de réponse pour chaque motif sont : 0, 1, 2, 3-4, 5-6, 7+. Pour l'étude, nous avons retenu les valeurs médianes des catégories de réponse 3-4 e 5-6, et 7.5 pour la catégorie supérieure. Les usages de drogues retenus pour l'étude sont binaires : usage quotidien de tabac (au moins une cigarette par jour au cours des 30 derniers jours), usage régulier d'alcool et de cannabis (au moins 10 usages durant les 30 derniers jours). Les catégories de réponse originales seront utilisées pour calculer des coefficients de corrélation linéaire : de « jamais », « moins d'une cigarette par semaine » à « plus de 20 cigarettes par jour » (en réponse à la question « combien de cigarettes avez-vous fumées... ») pour le tabac (échelle de 0 à 7) ; 0, 1-2, 3-6, 9-10, 10-19, 20-39, 40+ (en réponse à la question « combien de fois avez-vous... ») pour les usages d'alcool et de cannabis au cours des 30 derniers jours (échelle de 0 à 6).

Les variables auxiliaires de la base de sondage disponibles sont le secteur d'éducation (public ou privé), le niveau et la filière (8 modalités) et le type de commune hébergeant l'établissement scolaire (ville centre ou non). La strate utilisée dans le sondage et pour nos calculs est composée du secteur et de la filière d'éducation.

2.3. Description de l'échantillon

L'échantillon tiré comporte 284 classes et 7169 élèves, parmi lesquels seulement 6185 étaient présents et 984 absents (parmi lesquels 16 refus parentaux ou d'élèves). Parmi les absents, 359 étaient à coup sûr des sécheurs ce jour-là (soit 36.5%). Parmi les présents, seuls 6007 ont rendu un questionnaire analysable (soit 83.8% de l'échantillon théorique) : 1162 questionnaires sont donc manquants (dont 31% de sécheurs) et doivent être représentés par les questionnaires présents.

Parmi les questionnaires saisis, les taux de non-réponse partielle aux déclarations d'absences passées sont modestes : 5.4% pour les absences pour cause de maladie, 8.4% sans motif (sèche) et 7.1% pour d'autres raisons). Parmi les 178 questionnaires éliminés, les taux sont plus élevés (16%, 17% et 18% respectivement). Les résultats des imputations sont donnés tableau 1 : par la suite seules les variables imputées seront utilisées.

Tableau 1 : nombre de jours d'école manqués par motif avant et après imputation, dans l'ensemble des questionnaires analysés et exclus

	Maladie				Sèche (aucune raison)				Autre raison			
	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)	6,740,5	7,107,5			3,920,5	4,325,5			5,045	5,422,5		
Ratio (%)	42,9%	42,2%			25,0%	25,7%			32,1%	32,1%		

Le nombre moyen de jours d'absences passées est supérieur parmi les questionnaires éliminés (6.2 jours vs 2.6 jours) ainsi que la proportion de la sèche dans le total (37% vs 25%). Les questionnaires éliminés sont donc ceux d'élèves plus souvent absents et sécheurs que les élèves dont le questionnaire est retenu pour l'analyse. Parmi ces derniers, les garçons rapportent moins d'absences passées que les filles (63.6% ont été absents contre 69.6%), avec une durée moyenne d'absence parmi les absents identique (3.9 jours), mais autant de sèche : 26.8% disent avoir séché contre 25.4% des filles (en moyenne 2.5 jours sans distinction entre les sexes). La part des jours de sèche parmi les jours d'absence vaut 24.5% (26.8% chez les garçons et 22.9% chez les filles), ce qui est bien inférieur à la proportion de sécheurs parmi les absents le jour de l'enquête (36.5%).

Le tableau 5 (cf. *infra*) illustre la variabilité par strate des effectifs de présents, d'absents et des proportions de sécheurs le jour de l'enquête (partie Description des classes), ainsi que la variabilité des niveaux d'usage régulier de cannabis (colonne Estimation calée classique). Globalement, les taux d'absence et de sèche sont supérieurs dans le secteur public ainsi que dans les filières courtes et professionnalisantes.

2.4. Correction proposée

La correction de la non-réponse totale (CNRT) est classiquement faite à l'aide d'une modélisation, ce qui suppose de disposer des mêmes informations et variables pour les répondants et non-répondants. Dans Espad, les élèves sont soit absents, soit présents le jour de l'enquête mais aucune information individuelle ni d'historique d'absences n'est disponible pour les absents. Les présents indiquent quant à eux combien de jours de classe ils ont manqué au cours des 30 derniers jours, et pour quels motifs (ce qui fait défaut aux manquants). Il est donc impossible de modéliser la probabilité de réponse ou de non-réponse totale des répondants. Le principe de la CNRT dans notre cas sera donc de 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.

La CNRT sera effectuée séparément pour les filles et les garçons, car le sexe est associé aux usages et aux absences (McAra 2004), et à plusieurs niveaux : au niveau global de l'échantillon, au niveau de la classe et au niveau de la strate. La CNRT au niveau global ignore les spécificités des élèves, (filière d'enseignement, le secteur etc.), ce qui implique de remplacer n'importe quel absent par l'ensemble des présents au prorata de leurs absences passées. Les élèves de CAP représenteront aussi bien les élèves de filière générale que l'inverse et cela, en dépit du fait que les niveaux d'usages

varient grandement avec les filières (cf. le tableau 5 pour l'usage régulier de cannabis). La CNRT au niveau de la classe intègre au contraire toutes les spécificités des élèves mais elle pourrait conduire à des repondérations extrêmes dans les cas où les nombres d'élèves répondants sont faibles. Enfin, on peut retenir une CNRT à un niveau intermédiaire qui est la strate de sondage ou une approximation de celle-ci, ce qui assure un nombre d'élèves supérieur à celui des classes tout en préservant une certaine homogénéité de ceux-ci au sein de chaque strate.

Enfin, la CNRT peut ignorer les motifs d'absence passées des répondants et les types d'absence des élèves manquants le jour de l'enquête pour se concentrer sur les totaux (approche par les absences) ou distinguer la sèche et les absences légitimes (approche par la sèche). Dans le premier cas, tous les répondants ayant été absents représenteront tous les absents le jour J (au prorata de leurs absences) ; dans le second, les répondants ayant été sécheurs représenteront les sécheurs du jour J, les répondants ayant été absents pour motif légitime représenteront les sécheurs le jour J, au prorata de leurs déclarations d'absence et de sèche.

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

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

W est le poids de sondage ; $T(\text{absents le jour } J)_{N,W}$ représente le total pondéré par W du nombre d'élèves manquants (refusants l'enquête, absents ou dont le questionnaire a été éliminés de l'analyse) dans le niveau N ; N peut prendre les valeurs G pour global, S pour strate et C pour classe ; $T(\text{nb jours absences})_{N,W}$ est la somme pondérée par le poids de sondage des nombres de jours d'absences déclarées par les répondants analysés dans le niveau N . Si $T(\text{nb jours absences})_{N,W}=0$ alors $pNA=1$.

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 ni aucune strate dans laquelle nombre total de jours d'absences déclaré par les répondants est nul)

Les CNRT de niveau classe (pCA et pCS) conduisant à des pondérations parfois extrêmes, un tronquage au 99^{ème} percentile a été effectué, suivi d'un réechelonnement sur l'effectif total de la classe. Notez enfin que pour certaines classes, il n'y avait soit pas de questionnaire de garçon retenu pour l'analyse (7 classes), soit pas de questionnaire de fille (4 classes). Les effectifs corrigés par la CNRT étaient donc inférieurs au total théorique (7 117 vs 7 169). Ce problème est toutefois corrigé dans l'étape de calage. Après CNRT, l'échantillon fait encore l'objet d'un calage afin d'obtenir les poids finaux CNRT-calés.

2.5. Statistiques

Les effets de la CNRT seront estimés en comparant les estimations calées directement (redressement classique) et les estimations CNRT-calées. Les calculs ont été faits sous SAS V9.4. Les imputations des non-réponses partielles aux déclarations d'absence passées des répondants ont été faites à l'aide de la procédure PROC SURVEYIMPUTE par un hotdeck aléatoire incluant le sexe, la strate et les trois types d'absence. Toutes les estimations ont été faites à l'aide des procédures adaptées à des plans

de sondage complexe : PROC SURVEYFREQ et PROCSURVEYLOGISTIC. Le plan de sondage (équilibré) a été assimilé à un plan stratifié (variable strate) et en grappes (classe). La variance additionnelle causée par la CNRT est estimée par le ratio d'écart-type entre les estimations obtenues par CNRT-calage et les estimations directement calées.

3. Résultats

3.1. Absences passées et usages de drogues

Les coefficients de corrélation de Pearson entre le nombre de jours d'absences passées déclarées et les fréquences d'usages de drogues ainsi que les indicateurs binaires d'usages sont présentés Tableau 2. Pour les deux sexes, ceux de la sèche sont supérieurs à ceux du total des absences. Les corrélations sont faibles pour l'alcool (0.03) mais plus élevées pour le tabac et le cannabis (autour de 0.2).

Tableau 2 : Corrélations entre les nombres de jours d'école manqués et de jours d'école séchés au cours des 30 derniers jours et quelques usages de drogues

	Nombre de jours séchés	Nombre de jours d'absence	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

Données pondérées par le poids de sondage

Une régression logistique contrôlant le sexe, le nombre total de jours d'absences (incluant la sèche) ainsi que le nombre total de jours de sèche déclarés au cours des 30 derniers jours montre que c'est surtout la sèche qui est associée aux usages. En effet, pour le tabagisme quotidien, l'OR associé à un jour de sèche supplémentaire vaut OR=1.28 [1.19; 1.37], tandis que l'OR pour un jour d'absence supplémentaire vaut OR=1.03 [1.00; 1.07]. Les valeurs correspondantes pour l'usage régulier d'alcool sont OR=1.14 [1.04; 1.24] et OR=0.99 [0.96; 1.03], et celles pour l'usage régulier de cannabis OR=1.32 [1.21; 1.45] et OR=1.02 [0.97; 1.06]. Toutefois, comme le souligne le fort coefficient de corrélation linéaire entre le total de jours d'absence et de sèche ($r=0.68$ chez les garçons, 0.65 chez les filles, $r=0.66$ pour l'ensemble), il est légitime de considérer une CNRT ne distinguant pas la sèche des autres absences légitimes et donc mobilisant uniquement le total de jours d'absences toutes causes (dans ce cas, les OR valent : OR=1.12 [1.09 ; 1.14] pour le tabac quotidien, OR=1.04 [1.01 ; 1.06] pour l'alcool régulier et OR=1.12 [1.09 ; 1.16] pour le cannabis régulier).

Toutes les associations décrites perdurent (et sont pratiquement inchangées) si on introduit la filière scolaire et le secteur (public/privé) dans les modèles, qui sont les principales variables auxiliaires de la base de sondage : ces résultats étayaient bien la thèse suivant laquelle l'absentéisme est bien lié aux comportements à estimer, les usages de drogues, de façon non-ignorable pour un redressement classique utilisant les variables auxiliaires disponibles. Ils n'en sont toutefois pas une preuve, puisqu'ils ne sont obtenus que pour les répondants, mais il est peu probable que les relations soient inversées ou même largement modifiées parmi les sécheurs et absents le jour de l'enquête.

3.2. Effet de la CNRT

Le tableau 3 montre la dispersion des poids issus des 6 procédures de CNRT. Les poids finaux sont issus d'une troncature au 99^{ème} percentile des poids bruts (avant calage). Cette troncature a été opérée afin de limiter la dispersion des poids avant calage. Elle a pour effet de faire porter une partie du poids des absents sur les présents quel que soit leur profil d'absences passées : ainsi des répondants n'ayant jamais été absents voient leur poids augmenter un peu dans certains cas.

Les dispersions finales sont assez élevées et diminuent avec la taille des niveaux (classe puis strate puis global). L'élévation de la dispersion des poids est cependant assez mesurée comparativement à la dispersion du poids de sondage ou du poids de sondage calé directement dans l'approche classique (CV max=92.6 pour *pCS* vs 76.9 pour le poids de sondage calé). On note cependant que certains élèves ont un poids très fort : jusqu'à 22.4 pour la méthode *pCS*.

Tableau 3 : Statistiques des poids CNRT et CNRT-calés

	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
<i>pGA</i>	0.5	0.6	1.2	3.6	6.4	7117	72.8
<i>pGA</i> calé	0.3	0.7	1.2	4.5	11.1	7169	81.0
<i>pGS</i>	0.5	0.7	1.2	3.7	6.5	7117	73.1
<i>pGS</i> calé	0.3	0.7	1.2	4.5	11.5	7169	81.3
<i>pSA</i>	0.5	0.6	1.2	3.8	5.7	7117	73.5
<i>pSA</i> calé	0.3	0.7	1.2	4.5	9.6	7169	81.1
<i>pSS</i>	0.5	0.6	1.2	3.9	6.3	7117	74.1
<i>pSS</i> calé	0.3	0.7	1.2	4.5	10.6	7169	81.8
<i>pCA</i>	0.3	0.6	1.2	4.2	11	7117	79.3
<i>pCA</i> calé	0.3	0.7	1.2	4.8	19.1	7169	90.7
<i>pCS</i>	0.2	0.6	1.2	4.3	13.6	7117	81.7
<i>pCS</i> calé	0.2	0.7	1.2	4.8	22.4	7169	92.6

Note : *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).

Le tableau 4 illustre l'effet des 6 CNRT sur les estimations des indicateurs binaires d'usage de drogue. Comme attendu, les estimations corrigées sont supérieures à celles calées directement. Les augmentations sont importantes pour le tabagisme quotidien (passant de 22.6% à 24% pour *pGA*, *pGS*, *pSA*, *pSS*) et l'usage régulier de cannabis (passant de 7.8% à 8.5% ou 8.6% pour *pGA*, *pGS*, *pSA*, *pSS*). Elles varient assez peu avec l'approche (distinguer la sèche ou non) et le niveau, bien que les CNRT opérées à la classe soient généralement inférieures à celles opérées dans les niveaux supérieurs. Les augmentations de variance sont relativement importantes, puisqu'elles peuvent s'élever jusqu'à 14% pour l'usage régulier de cannabis (*pGS*). Aucune estimation corrigée n'est en dehors de l'intervalle de confiance de l'estimation classique.

Le tableau 5 montre l'effet de la CNRT de niveau strate sur l'estimation de la consommation régulière de cannabis. Les variations sont très variables suivant les filières et le secteur et peuvent dépasser un point. Toutefois, aucune estimation corrigée n'est en dehors de l'intervalle de l'estimation calée classique correspondante.

Tableau 4 : Estimations des indicateurs binaires d'usages de drogues au niveau national suivant les différentes options de CNRT

	Calage direct		pGA calé				pGS calé				pSA calé				pSS calé				pCA calé				pCS calé				
	%	std	%	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	61.5	1.2	1.2	1.03	61.5	1.2	1.2	1.03	61.1	1.2	0.8	1.05	61.1	1.2	0.7	1.05	
Quotidien	22.6	1.2	24.0	1.3	1.4	1.09	24.1	1.3	1.5	1.10	24.0	1.3	1.4	1.09	24.0	1.3	1.3	1.09	23.5	1.3	0.9	1.11	23.1	1.3	0.5	1.11	
Alcool																											
Vie	86.3	1.0	86.0	1.0	-0.3	1.02	86.1	1.0	-0.2	1.02	86.0	1.0	-0.3	1.02	86.0	1.0	-0.3	1.02	86.1	1.0	-0.3	1.02	86.1	1.0	-0.3	1.03	
Mois	63.4	1.5	63.5	1.5	0.2	1.02	63.6	1.5	0.2	1.02	63.6	1.5	0.2	1.01	63.5	1.5	0.2	1.02	63.6	1.6	0.2	1.04	63.5	1.6	0.1	1.07	
Régulier	14.4	1.1	14.7	1.1	0.4	1.08	14.8	1.1	0.4	1.09	14.7	1.2	0.3	1.10	14.8	1.2	0.4	1.10	14.8	1.2	0.5	1.13	14.8	1.2	0.4	1.12	
Cannabis																											
Vie	44.0	1.2	45.4	1.3	1.5	1.04	45.6	1.3	1.6	1.04	45.4	1.3	1.5	1.04	45.4	1.3	1.5	1.04	45.0	1.3	1.0	1.07	44.9	1.3	0.9	1.06	
Mois	22.6	1.0	23.8	1.1	1.2	1.05	24.0	1.1	1.4	1.06	23.8	1.1	1.2	1.06	23.8	1.1	1.2	1.05	23.5	1.1	0.9	1.09	23.3	1.1	0.7	1.07	
Régulier	7.8	0.7	8.5	0.8	0.8	1.12	8.6	0.8	0.9	1.14	8.5	0.8	0.8	1.13	8.5	0.8	0.8	1.12	8.2	0.8	0.4	1.13	8.1	0.8	0.3	1.12	

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).

%=estimation; std=écart-type; D= différence avec l'estimation obtenue par calage direct; Deff= design effect calculé comme le ratio des écarts-types.

Tableau 5 : Effets de la correction de la non-réponse totale au niveau de la strate sur l'usage régulier de cannabis

Description de la strate			Description des absences le jour J					Estimations pondérées					
Strate n°	Filière et niveau	Sector	Classes	Elèves	Elèves manquants	Quest. éliminés	Sécheurs parmi les manquants	Calage direct		<i>pSA</i> calé		<i>pSS</i> calé	
			N	N	%	%	%	%	Std	%	Std	%	Std
2211	Seconde	Privé	17	520	6.2%	1.7%	15.6%	3.5	1.4	3.8	1.4	3.8	1.4
		Public	59	1887	10.6%	1.1%	19.5%	4.9	0.9	5.1	0.9	5.1	0.9
2212	Première	Privé	14	368	8.4%	2.7%	19.4%	5.5	1.4	5.7	1.5	5.7	1.5
		Public	36	997	13.3%	1.8%	27.8%	8.5	1.8	9.5	2.0	9.3	2.0
2213	Terminale	Privé	8	182	10.4%	3.8%	26.3%	6.4	3.4	7.2	4.0	7.0	3.8
		Public	29	810	14.9%	1.4%	42.1%	8.7	2.0	10.0	2.4	9.9	2.4
2321	CAP 1	Ensemble	11	160	21.3%	3.8%	17.6%	3.9	1.7	6.5	2.6	6.5	2.6
2322	CAP 2	Ensemble	7	125	22.4%	1.6%	82.1%	9.6	2.8	8.0	2.8	8.3	2.9
2381	BEP 1	Privé	9	159	9.4%	1.9%	26.7%	11.8	4.8	12.4	4.9	12.4	4.9
		Public	41	912	16.4%	4.6%	54.0%	8.3	1.5	9.8	1.7	9.9	1.7
2382	BEP 2	Privé	6	112	7.1%	3.6%	0.0%	9.2	1.4	9.4	1.8	9.3	1.6
		Public	21	405	16.0%	5.2%	56.9%	10.3	2.5	10.5	2.4	10.7	2.3
2383	BEP 3	Ensemble	26	532	21.6%	4.5%	53.0%	13.5	3.9	14.8	4.2	15.1	4.2
-	Ensemble	Privé	61	1476	8.2%	2.5%	20.7%	7.3	1.2	7.8	1.3	7.8	1.3
-	Ensemble	Public	223	5693	14.6%	2.5%	39.8%	7.9	0.8	8.7	0.9	8.7	0.9

Std: écart-type. *pSA* calé (resp. *pSS* calé): CNRT au niveau de la strate distinguant (*pSS*) ou ne distinguant pas (*pSA*) la sèches des autres absences, suivie d'un calage.

4. Discussion

4.1. Résumé des résultats

Malgré une littérature scientifique solide et l'existence d'un absentéisme souvent important et susceptible d'affecter les estimations de niveaux d'usages de drogues, aucune proposition de correction de ce biais de non-réponse non-ignorable n'a jamais été proposé dans les enquêtes scolaires. Nous avons tiré profit de l'existence de rapports de passation détaillant les absences et présences au sein de chaque classe lors de la collecte de l'enquête pour en proposer une, reposant sur le principe de représenter les absents/sècheurs le jour de l'enquête par des présents ayant un historique déclaré d'absence ou de sèche, au prorata des totaux d'absences/sèches qu'ils déclarent. Les résultats montrent une augmentation notable des estimations pour le tabagisme quotidien et l'usage régulier de cannabis, mais sans que les estimations corrigées sortent des intervalles de confiance des estimations classiques.

4.2. Interprétation

Aucune estimation corrigée ne sort pas des marges des estimations traditionnelles. Cela peut être jugé rassurant pour les publications antérieures, mais cette robustesse est en partie due au plan de sondage à probabilité inégales avec un fort effet de grappe du au regroupement des élèves dans des classes entières. Autrement dit, la robustesse des estimations classiques repose en partie sur l'imprécision de ce type de sondage : si l'on ignore le grappage des individus, presque toutes les estimations corrigées des niveaux de tabagisme ou d'usage régulier de cannabis sortent des intervalles de confiance des estimations classiques.

4.3. Limites

Les limites de notre méthode de CNRT sont nombreuses. En premier lieu, elles s'appuient sur les déclarations d'absence passées des élèves répondants. On peut suspecter que les absences illégitimes (sèche) soient quelque peu minorées par les élèves, au profit par exemple des absences pour autre raison ou pour maladie. Ce problème est toutefois sans effet sur les CNRT opérées sans distinguer la sèche des autres absences, pourvu que le nombre total d'absences passées soit correct. La faible différence entre les deux approches montre que notre méthode est robuste de ce point de vue, ce qui est probablement dû à la corrélation entre le total des jours d'absences déclarées et le total des jours de sèche déclarés par les élèves.

Ensuite, on pourrait suspecter un biais de déclaration des niveaux d'usages de drogues qui varierait avec les nombres de jour d'absence et de sèche passés. Par exemple, si les sècheurs minorent davantage que les autres leurs niveaux d'usage, les estimations seraient biaisées à la baisse. Ce biais différentiel est cependant peu vraisemblable. En revanche, on ne peut écarter une sous-déclaration générale des usages par tous les élèves répondants, ce que notre méthode ne peut corriger.

Un troisième biais provient de la précision du recueil des types d'absence le jour de l'enquête au sein des rapports de passation. Si les sèches sont bien avérées, il est probable que les absences au statut indéterminé (n=116) soient pour partie des sèches : les considérer ainsi devrait élever nos corrections.

Enfin, nous avons assimilé les élèves ayant produit un questionnaire éliminé (n=178) à des élèves absents légitimement et non à des sècheurs, alors même que leurs nombre d'absences illégitimes déclarées étaient plus élevés. Les assimiler à des sècheurs ferait passer la proportion de sècheurs à de 31% du total des questionnaires absents à représenter à 46%. Cela resterait toutefois assez hasardeux car ces élèves étaient présents et n'ont pas tous déclaré avoir séché au cours des 30 derniers jours et de plus, le taux de non-réponse partielle à ces dénombrement d'absences/sèches passées est conséquent et choisir un seuil au-delà duquel on devrait considérer qu'un questionnaire non retenu

est assimilable une absence de questionnaire pour motif de sèche serait quelque peu arbitraire. Au vu de ces différents biais, notre correction apparaît globalement prudente et pourrait bien sous-estimer la réalité.

4.4. Recommandations et applications de la méthode

Le bien-fondé de la méthode repose sur l'existence d'une proportion importante de sécheurs, sur la fiabilité de leur repérage ainsi que sur l'existence d'un lien fort entre niveau d'usage de drogues et sèche. Il semble donc important de mesurer précisément ces aspects lors des collectes scolaires, qui peuvent être considérés comme des indicateurs de qualité dans une enquête scolaire. Cela impose notamment l'établissement de rapports de passation fiables.

La méthode pourrait peut-être être étendue à quelques enquêtes auprès des ménages ou des individus lorsqu'on suspecte un lien entre la réponse à l'enquête et les comportements étudiés et plus spécifiquement une non-réponse non-ignorable. Pour cela, il convient de recueillir le motif d'absence ou de refus des personnes sélectionnées non répondantes et les historiques d'absences des répondants ainsi que leurs motifs, lorsque cela est possible. Par exemple, dans l'enquête Mobilité, il pourrait être possible de recueillir le motif d'absence du ménage ou de la personne sélectionnée auprès du voisinage ou des autres membres du ménage, notamment s'il s'agit d'un déplacement de longue durée durant la période de collecte ; tandis qu'il est possible de demander aux répondants d'indiquer la fréquence de leurs déplacements et leur disponibilité durant les 3 derniers mois.

Bien sûr, nous reconnaissons que l'applicabilité d'une telle méthode hors cadre scolaire reste déterminée par l'obtention des données auprès des non-répondants (absents ou refusants), généralement bien plus ardue que dans les enquêtes scolaires. Dans celles-ci en effet, le recueil est global et non individuel (il est impossible de relier une absence lors de l'enquête à des informations de la base de sondage hors le sexe et la classe), ce qui en limite quelque peu la portée ; toutefois il est exhaustif, peu coûteux et fiable car effectué par le superviseur au niveau de la classe, par simple comptage.

Références

- Davern, M. D., T. McAlpine, J. Beebe, J. Y. Ziegenfuss, T. H. Rockwood and K. T. Call (2010). "Are lower response rates hazardous to your health survey? An analysis of three state health surveys." Health Serv Res **45**(5): 1324-1344.
- Deville, J. and C.-E. Sarndal (1992). "Calibration estimators in survey sampling." Journal of the American Statistical Association **87**(418): 376-382.
- Groves, R. M. (2006). "Nonresponse Rates And Nonresponse Bias In Household Surveys." Public Opinion Quarterly **70**(5): 646-675.
- Guttormsson, U., H. Leifman, L. Kraus, S. Arpa, S. Molinaro, K. Monshouwer, M. Trapencieris, J. Vicente and J. Svensson (2016). *Espad 2015 methodology*. Espad. Lisbon, EMCDDA: 40.
- Haziza, D. and E. Lesage (2016). "A discussion of weighting procedures for unit nonresponse." Journal of Official Statistics **32**(1): 129-145.
- Henry, K. H. and T. P. Thornberry (2010). "Truancy and Escalation of Substance Use During Adolescence." Journal of Studies on Alcohol and Drugs **71**: 115-124.
- Hibell, B., B. Andersson, T. Bjarnason, S. Ahlström, O. Balakireva, A. Kokkevi, M. Morgan, Conseil de l'Europe, Groupe Pompidou and CAN (2004). *The ESPAD Report 2003. Alcohol and other drug use among students in 35 European countries*. Stockholm, CAN (Council for Information on Alcohol and other Drugs): 355 p.
- Hibell, B., U. Guttormsson, S. Ahlström, O. Balakireva, T. Bjarnason, A. Kokkevi and L. Kraus (2012). *The 2011 ESPAD Report: Substance Use Among Students in 36 European Countries*. Stockholm, CAN: 394.
- Holtes, M., R. Bannink, E. Joosten-van Zwanenburg, E. Van As, H. Raat and S. Broeren (2015). "Associations of truancy, perceived school performance, and mental health with alcohol consumption among adolescents." Journal of School Health **85**(12).
- Little, R. J. A. and S. Vartivarian (2005). "Does weighting for nonresponse increase the variance of survey means?" Survey methodology **31**(2): 161-168.
- McAra, L. (2004). *Truancy, school exclusion and substance misuse*. Edinburgh, University of Edinburgh, UK: 30.
- Michaud, P. A., I. Delbos-Piot and F. Narring (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.
- Mounteney, J. and A. Skutle (2010). "Truancy, alcohol use and alcohol-related problems in secondary school pupils in Norway." Health Education Research **25**(6): 945-954.
- Skalland, B. (2011). "An alternative to the response rate for measuring a survey's realization rate of the target population." Public Opinion Quarterly **75**(1): 89-98.
- Wichstrom, L. (1998). "Alcohol intoxication and school dropout." Drug Alcohol Rev **17**(4): 413-421.
- Yoep, N., L. Kuang Kuay and N. Safiza (2016). "Prevalence of Truancy and Its Associated Factors among School-Going Malaysian Adolescents: Data from Global School-Based Health Survey 2012." Psychology **7**: 1053-1060.