

Résumé des méthodes de recherche sur le tabagisme chez les adolescents - RDC

Ce document résume l'enquête « **Données sur les Jeunes et le Tabac en Afrique** » (*Data on Youth and Tobacco in Africa, DaYTA*) utilisée dans la page « Consommation de tabac et de nicotine chez les adolescents en RDC » pour la République démocratique du Congo (RDC). Il récapitule la méthode d'étude afin de mieux comprendre comment l'enquête a été menée.

L'étude DaYTA a été menée dans le cadre du programme DaYTA, mis en œuvre par Development Gateway : An IREX Venture financée par la Fondation Gates. Le programme DaYTA est une initiative de recherche globale visant à comprendre la consommation de tabac chez les adolescents âgés de 10 à 17 ans dans trois pays d'Afrique subsaharienne, à savoir la RDC, le Kenya et le Nigeria. En RDC, Development Gateway : An IREX Venture a confié la réalisation de cette étude à *Research Initiatives for Social Development* (RISD) et au *Laboratory for Survey and Research for Development* (SRD Lab).

1 Taille de l'échantillon

La conception de l'étude a permis d'obtenir un échantillon représentatif de la population adolescente à l'échelle nationale, couvrant à la fois les zones urbaines et rurales. La taille de l'échantillon a été calculée à l'aide de la formule des Nations Unies¹ pour estimer la taille des échantillons dans les études de prévalence, avec un niveau de confiance de 95 %. Un effet de conception de 1,5 a été appliqué pour tenir compte de la conception de l'échantillonnage à plusieurs degrés. En outre, un taux de non-réponse de 10 % a été intégré dans le calcul de la taille de l'échantillon, conformément aux études précédentes menées en RDC.^{2,3} La proportion de la population adolescente a été estimée à 23 % et la taille moyenne des ménages a été fixée à 5,25, selon l'Institut National de la Statistique.⁴ Nous avons appliqué une prévalence du

¹ UN. (2008). *Studies in Methods: Designing Household Survey Samples: Practical Guidelines*: Vol. F (98th ed.).

² Mukuku O, Mbuyi J, Tshipuka D. *Pratiques alimentaires des nourrissons : Connaissances, attitudes et pratiques des mères d'une commune urbaine de la ville de Lubumbashi, République Démocratique du Congo*. ResearchGate [internet]. 2018.

³ Ntambue AM, Tshiala RN, Malonga FK, Ilunga TM, Kamonayi JM, Kazadi ST, Matungulu CM, Musau AN, Mulamba D, Dramaix-Wilmet M, Donnen P. Utilisation des méthodes contraceptives modernes en République Démocratique du Congo : prévalence et barrières dans la zone de santé de Dibindi à Mbuji-Mayi. *Pan African Medical Journal*. 2017; 26:199. Available from: <https://doi.org/10.11604/pamj.2017.26.199.10897ijssass.com+2>

⁴ Institut National de la Statistique de la RDC. *Annuaire Statistique RDC*. 2021;Page 142.

tabagisme chez les adolescents de 25 %, sur la base d'une enquête précédente qui faisait état de taux de tabagisme de 22,3 % à Kinshasa et de 24,6 % à Lubumbashi.⁵

En utilisant ces paramètres dans la formule d'échantillonnage, nous avons calculé une taille minimale requise pour un échantillon représentatif au niveau national de 4323 adolescents en RDC. Nous avons appliqué un taux de non-réponse de 10 %, ce qui, combiné aux ajustements d'arrondi effectués à chaque étape de l'échantillonnage (provinces, zone de santé, aire de santé et villages/avenues), a abouti à une taille finale de l'échantillon cible de 4892 personnes. Après avoir identifié les ménages éligibles, en particulier ceux comportant des adolescents âgés de 10 à 17 ans, la taille finale de l'échantillon DaYTA en RDC était de 4675 adolescents.

2 Procédure d'échantillonnage

L'étude était une enquête transversale nationale, basée sur la population et menée auprès des ménages, qui s'est concentrée sur les adolescents âgés de 10 à 17 ans et leurs tuteurs en RDC de mars à mai 2024. L'enquête a utilisé un plan d'échantillonnage à plusieurs degrés afin de garantir une couverture et une représentativité complètes. À l'aide de la formule de Taro Yamane,⁶ la collecte des données a été effectuée dans 16 des 26 provinces du pays. Les 26 provinces ont été regroupées en six strates, correspondant aux six anciennes provinces du pays (1947-1963) : Katanga, Kasai, Léopoldville, Équateur, Orientale et Kivu. Les provinces de chaque strate partagent des caractéristiques historiques, culturelles et socio-économiques communes susceptibles d'avoir influencé les comportements et les attitudes à l'égard du tabagisme.

DaYTA a utilisé un plan d'échantillonnage en grappes stratifié à plusieurs degrés afin d'obtenir un échantillon représentatif à l'échelle nationale. Lors de la **première étape**, 16 provinces ont été sélectionnées de manière aléatoire en fonction des six strates représentant les 26 régions. Lors de la **deuxième étape**, trois zones de santé ont été choisies au hasard dans chacune des 16 provinces sélectionnées, pour un total de 48 zones de santé. Au cours de la **troisième étape**, dans chacune des 48 zones de santé sélectionnées, trois aires de santé ont été échantillonnées de manière aléatoire : deux dans les zones rurales et une dans les zones urbaines pour chaque zone sanitaire. Au total, 144 aires de santé ont ainsi été sélectionnées. Dans chaque aire de santé, une avenue (en zone urbaine) ou un village (en zone rurale) a été choisi pour la collecte des données. Ces avenues et villages constituaient les zones de recensement (**pour la quatrième étape**). Enfin, lors de la **cinquième étape**, dans les avenues et les villages sélectionnés, un recensement des ménages a été effectué afin d'identifier les ménages éligibles, en particulier ceux comprenant des adolescents âgés de 10 à 17 ans.

⁵ Mbuyu MR, Banza LNC. Rapport de l'enquête globale sur le tabagisme chez les jeunes, Kinshasa, RDC. 2008.

⁶ Taro Y. Statistics, An Introductory Analysis. "New York: Harper and Row, 2 Editions. 1967.

3 Pondérations de l'échantillon

Les pondérations de l'échantillon ont été obtenues en appliquant une procédure de pondération en deux étapes. Tout d'abord, les pondérations ont été calculées pour chaque participant à l'échantillon. Ensuite, ces pondérations de base ont été ajustées afin de corriger les différences de non-réponse.

Les pondérations ont été calculées comme l'inverse de la probabilité globale de sélection pour chaque participant. Les probabilités de sélection ont pris en compte plusieurs étapes, notamment la probabilité de sélectionner des unités d'échantillonnage primaires (UEP), qui correspondaient aux zones de recensement représentées par des avenues dans les zones urbaines et des villages dans les zones rurales. Elles ont également tenu compte de la probabilité de sélectionner des ménages au sein de chaque UEP et de la probabilité de sélectionner des adolescents éligibles au sein des ménages échantillonnés.

Les pondérations finales ont ensuite été normalisées afin de garantir que la somme de toutes les pondérations corresponde à la population cible totale. Pour ce faire, chaque pondération finale a été divisée par la somme de toutes les pondérations finales, puis multipliée par la taille de la population souhaitée. Ces pondérations ont été appliquées dans l'analyse des données afin de garantir que les résultats de l'enquête reflètent fidèlement l'ensemble de la population adolescente en RDC.

4 Collecte et analyse des données

Les activités de collecte de données ont été menées dans les provinces échantillonnées entre mars et mai 2024. Les questionnaires ont été administrés par des enquêteurs à l'aide de tablettes et réalisés en face à face.

Les données ont été saisies sur la plateforme SurveyCTO. Afin de réduire les problèmes de qualité des données, des outils intégrés ont été utilisés pour effectuer des contrôles de qualité automatisés, des sauts logiques et des contraintes. Les données de terrain ont été collectées hors ligne et téléchargées quotidiennement sur un serveur afin de minimiser les pertes en cas de perte ou de dysfonctionnement des tablettes. Les données sur les tablettes et en transit ont été cryptées afin d'en garantir la confidentialité. Une fois transférées vers le serveur central, les données ont été protégées par des mots de passe afin d'en restreindre l'accès au personnel autorisé, puis les données sur les tablettes ont été effacées. Les données soumises ont été transférées vers STATA et des contrôles fréquents ont été effectués sur toutes les données.

soumises. Des rapports signalant tout problème de qualité des données ont été partagés chaque semaine. Les problèmes signalés ont été résolus en consultation avec les équipes de terrain et les mesures appropriées ont été documentées.

Des statistiques descriptives ont été calculées afin de résumer les caractéristiques de la population étudiée au niveau des ménages et des individus, ainsi que pour évaluer la prévalence de la consommation actuelle et au cours de la vie pour chaque produit du tabac et de la nicotine. Les fréquences et les pourcentages ont été calculés pour les variables catégorielles telles que le type de résidence (rurale/urbaine), le sexe du chef de ménage, le niveau d'éducation, l'état civil, l'activité principale génératrice de revenus, la tranche d'âge des adolescents et la situation scolaire, entre autres. Des statistiques descriptives ont également été calculées pour d'autres indicateurs d'intérêt tels que la fréquence d'utilisation, l'âge de début de consommation, l'auto-efficacité à résister à la consommation, la susceptibilité à la consommation, l'intensité et la fréquence de consommation, et l'exposition à la fumée secondaire. Afin de garantir que ces estimations de prévalence soient précises et représentatives de l'ensemble de la population adolescente au niveau national, des pondérations ont été appliquées pour tenir compte de la probabilité de sélection, de non-réponse et de post-stratification.

Des modèles de régression logistique binaire ont été utilisés pour estimer les rapports de cotes ajustés des facteurs potentiels associés à la consommation actuelle de tout produit du tabac, au tabagisme actuel et à la consommation actuelle de tabac sans fumée. Des modèles de régression logistique distincts ont été construits pour ajuster les facteurs de confusion potentiels à chaque niveau. Les résultats sont présentés sous forme de rapports de cotes ajustés avec des intervalles de confiance à 95 %. Toutes les analyses ont été effectuées à l'aide du logiciel STATA version 18, les refus et les réponses « ne sais pas » étant traités comme des valeurs manquantes.