

Maximiser l'utilisation des données existantes pour renforcer la conception, l'évaluation et l'impact des programmes

HealthBridge Foundation of Canada
Peter R. Berti, Rebecca Brodmann

Université de Montréal
Milena Nardocci, Malek Batal

Center for Creative Initiatives in Health and Population
Tran Hung Minh

University College London
Naomi Saville

Statistics for Sustainable Development
Nicolas Greliche

Novembre, 2020

INTRODUCTION : Les organisations non gouvernementales (ONG) consacrent d'importantes ressources à la collecte de données, à la conception et au suivi de leurs interventions de santé. Les coûts varient en fonction des objectifs, du type d'étude, des méthodes, de la taille de l'échantillon, du contexte local et du travail requis pour planifier et réaliser la collecte de données. Selon le nombre de questions ou d'objectifs, les entretiens peuvent pendre beaucoup de temps, ce qui rend difficile la participation des répondants tout au long du processus. De plus, la précision des estimations des indicateurs dans les enquêtes menées par les ONG est souvent insuffisante pour la conception et le suivi des projets en raison des tailles d'échantillons relativement petites et de la forte variabilité inhérente des indicateurs d'intérêt, ce qui peut résulter en une conception de programme et des mesures de résultats inefficaces.

Cependant, les données pour de nombreux indicateurs de santé et sociaux de plusieurs pays existent déjà dans des bases de données publiques, comme l'Enquête démographique et de santé (DHS, en anglais, soutenu par l'USAID) et l'Enquête à grappes à indicateurs multiples (MICS, en anglais, soutenu par l'UNICEF). La DHS et la MICS fournissent des données standardisées recueillies à l'aide de méthodes rigoureuses, par de recenseurs bien formés, utilisant de grands échantillons. L'un des défis liés à l'utilisation des bases de données publiques comme la DHS ou la MICS pour l'évaluation de base est que ces enquêtes sont conçues pour être représentatives aux niveaux national, régional et provincial, mais les données sont rarement représentatives aux niveaux géographiques inférieurs, tels que les districts ou les villages où travaillent souvent les ONG. De plus, la DHS et la MICS sont menées chaque 3-10 ans, de sorte qu'il peut y avoir un écart de dix ans entre le moment où les données DHS/MICS ont été collectées et les conditions de base que l'ONG souhaite caractériser.

Malgré les défis, l'utilisation de bases de données publiques pour compléter ou remplacer la collecte de données primaires des ONG pour les mesures de base et le suivi des projets permettrait d'économiser des ressources précieuses, réduisant ainsi le fardeau pour le personnel des ONG et pour les répondants. Notez que nous ne considérerions pas les bases de données publiques comme appropriées pour les évaluations finales car le travail des ONG aura tenté de faciliter le changement par rapport à ces conditions de base.

HYPOTHÈSES: Nous émettons l'hypothèse que des bases de données publiques peuvent fournir des estimations de conditions de base similaires à celles dans les rapports de base des ONG lorsque le lieu, l'année et la saison de collecte des données sont les plus similaires possibles. De plus, nous émettons l'hypothèse que l'impact des différences pour l'année, le niveau géographique et la saison varient selon les indicateurs de santé.

Cette étude respecte les normes d'éthique de la recherche en vigueur et a été approuvée par le Comité d'éthique de la recherche en santé de l'Université de Montréal (CERSES-19-030-D).

Nos analyses ont été effectuées en quatre étapes : la *Partie 1* présente les résultats des analyses comparant les données de deux sources : les indicateurs de base des ONG et les estimations calculées à l'aide de la DHS ou de la MICS. La *Partie 2* présente les résultats

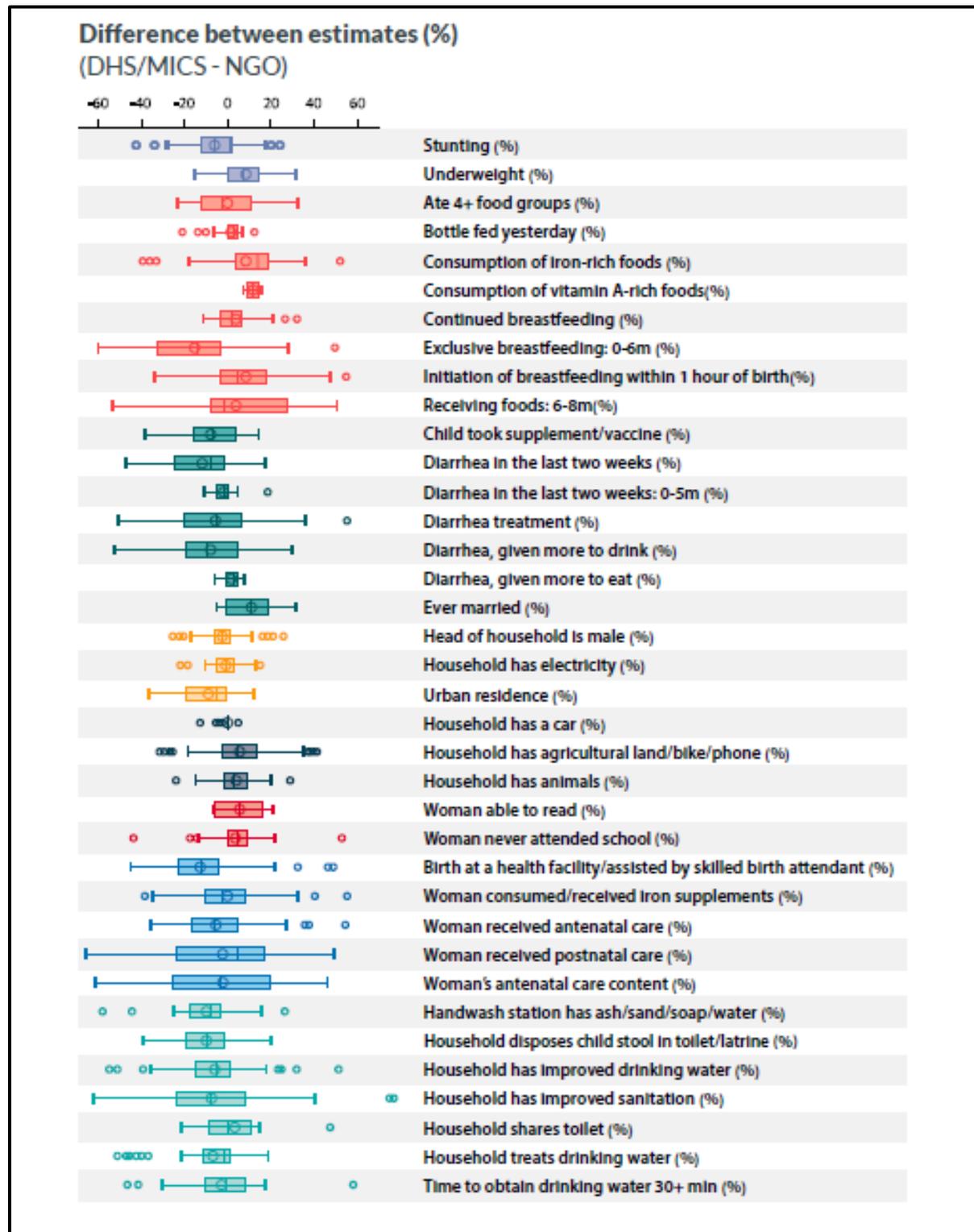
d'analyses comparant les données de la DHS de différentes années et différents niveaux géographiques. Ces analyses nous ont permis de travailler avec une taille d'échantillon beaucoup plus grande et de comparer les données de la DHS provenant d'années ou de niveaux géographiques différents. La *Partie 3* comprend une simulation qui mesure l'impact de l'erreur d'échantillonnage lorsqu'elle est isolée des autres sources d'erreur. La *Partie 4* présente deux études de cas réalisées au Vietnam et au Népal comparant les données d'un rapport de base d'une ONG dans chaque pays avec des estimations calculées à l'aide de la MICS (Vietnam) et de la DHS (Népal).

PARTIE 1 : Analyse ONG vs DHS/MICS

L'objectif de ces analyses était de comparer les estimations des indicateurs des rapports de base des ONG avec les estimations des indicateurs calculées à l'aide de base des données publiques (DHS ou MICS). Un échantillon de 46 rapports de base des ONG sur des projets de santé maternelle, néonatale et infantile réalisés dans 23 pays à revenu faible ou moyen a été compilé. À partir des rapports, nous avons extrait : le nom du pays, le nom de l'ONG, la date de début et de fin de la collecte des données, la population cible et les critères d'inclusion et d'exclusion, le nom de l'indicateur, la taille de l'échantillon (totale et pour chaque indicateur), le lieu de la collecte des données, et l'estimation de l'indicateur. Le lieu où les données ont été recueillies (par exemple, le nom de la région, province, district ou/et village) a été utilisé pour coder le niveau géographique. Le premier niveau représente la plus petite division géographique d'un pays, comme le village, la ville, la localité, l'autorité traditionnelle; le deuxième niveau représente le district (une division plus grande qu'un village mais plus petite que le troisième niveau); le troisième niveau représente la province, l'état, le département, le comté ou le district (s'il se réfère à une division équivalente à une province ou à un état); le quatrième niveau représente la région; et le cinquième niveau représente le pays.

Les estimations des indicateurs provenant de la DHS (pour vingt pays) et de la MICS (pour trois pays : Vietnam, Laos et Soudan du Sud) ont été appariées aux estimations extraites des rapports de base des ONG, en utilisant les données collectées au niveau géographique le plus proche possible. Au total, il y avait 139 indicateurs, regroupés en 41 sous-groupes, qui ont été associés entre les rapports des ONG et dans la DHS/MICS. Six des indicateurs étaient dérivés de variables numériques (p. ex., nombre d'enfants de moins de 5 ans à la maison, moyenne) et 133 étaient dérivés de variables binaires (p. ex., retard de croissance, %). Au total, 2 174 paires d'indicateurs DHS/MICS-ONG ont été retenues pour nos analyses. La différence (et la différence absolue) entre les paires ont été calculées, ainsi que la différence pour le niveau géographique, l'année de collecte des données et si les données ont été collectées au cours de la même saison ou non.

Graphique 1: Différence entre les estimations (DHS/MICS moins NGO) par sous-groupe d'indicateurs.



La différence entre certaines estimations clés est résumée dans le tableau suivant, qui montre que la plupart des indicateurs avaient des paires d'estimations qui avait des différences de 20 % ou plus, et plusieurs indicateurs avaient moins d'un tiers des paires avec une différence de 5 % ou moins. Les indicateurs avec une prévalence très faible ou très élevée étaient les plus proches.

Tableau 1. Différence entre les estimations de certains indicateurs clés.

Sous-groupe	Proportion de paires de indicateurs avec une différence absolue de			
	5 % ou moins	10 % ou moins	15 % ou moins	20 % ou moins
Anthropometries de l'enfant				
Retard de croissance (%)	38	66	83	93
Poids insuffisant (%)	31	55	74	81
Alimentation de l'enfant				
A mangé 4+ groupes d'aliments (%)	24	47	73	90
Allaitement exclusif : 0-6m (%)	15	32	43	52
Initiation de l'allaitement dans l'heure suivant la naissance (%)	36	50	64	75
Santé de l'enfant				
Diarrhée au cours des deux dernières semaines (%)	32	55	68	71
Ménage				
Ménage a de l'électricité (%)	60	70	85	95
Le chef du ménage est un homme (%)	56	79	88	92
Ménage possède une voiture (%)	91	98	100	100
Caractéristiques/Santé de la mère				
Femme peut lire (%)	12	62	75	87
Femme a accouché dans un établissement de santé/assisté par une accoucheuse qualifiée (%)	18	37	53	70
Femme a reçu des soins périnataux (%)	24	44	58	75
WASH				
Ménage a accès à de l'eau améliorée (%)	35	52	67	82
Ménage possède une installation sanitaire améliorée (%)	12	34	45	54

Afin de mieux comprendre comment les différences de niveau géographique, d'année et de saison pourraient expliquer les différences DHS/MICS-ONG, nous avons utilisé l'ANOVA pour partitionner la variance due à l'indicateur, à la différence de niveau géographique, d'année ou de saison. En bref, environ 17 % de la variance était attribuée à l'indicateur, moins de 1 % était attribuée aux différences par rapport au niveau géographique, à l'année ou à la saison, et la partie majeure de la variance était simplement aléatoire et n'était pas expliquée par le modèle.

PARTIE 2 : Analyse de DHS vs DHS

Nous avons comparé les données de la DHS de plusieurs années ou plusieurs régions pour examiner les sources de variation des différences d'estimations dues à l'erreur d'échantillonnage, à l'année de collecte des données ou au niveau géographique, mais pas dues aux erreurs des méthodes (puisque les méthodes de la DHS sont largement cohérentes entre

les enquêtes). Les données DHS de sept pays ont été utilisées : Bangladesh, Éthiopie, Kenya, Malawi, Pakistan, Tanzanie et Zambie, et les mêmes indicateurs qui ont été utilisés dans la *Partie 1* ont été retenus. Les indicateurs DHS de différents cycles et niveaux géographiques ont été comparés en utilisant différentes combinaisons imitant les scénarios réels observés dans DHS/MICS-ONG : indicateurs du même niveau géographique mais d'années différentes (Scénario 1), indicateurs de la même année mais des niveaux différents (Scénario 2), et indicateurs de différentes années et niveaux géographiques (Scénario 3). Les données du cycle le plus récent inclus étaient les données des niveaux géographiques inférieurs (pour imiter les données des ONG), et les données du cycle plus ancien étaient les données du niveau géographique supérieur (pour imiter les données DHS/MICS). La différence absolue entre les deux estimations DHS est résumée dans le tableau suivant.

Tableau 2. Différence absolue entre les deux estimations DHS par différences d'année et de niveaux géographiques.

Variable	Différence absolue entre les estimations (%)			
	Moyenne	ET	Médiane	Max
Différences d'années				
0-2.5 ans	9.4	10.2	6.1	93.3
3.0-5.0 ans	8.7	9.7	5.5	84.1
5.5-6 ans	13.0	14.1	8.6	97.6
Différences de niveaux géographiques				
0	9.6	11.4	5.7	97.4
1	9.8	10.8	6.3	95.6
2+	11.2	12.3	7.2	97.6

PARTIE 3 : Simulation d'erreur d'échantillonnage

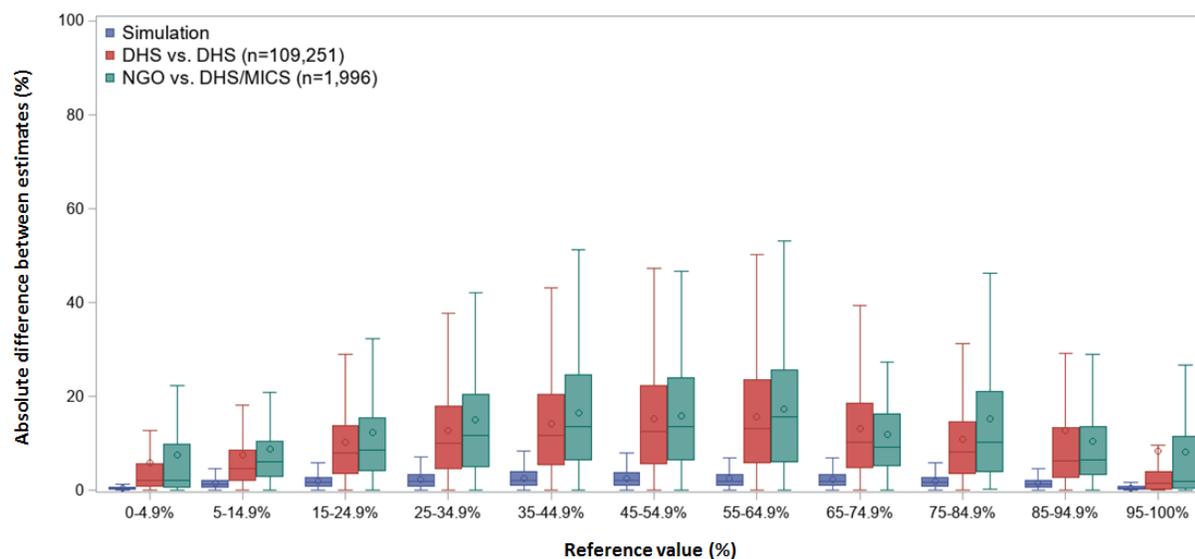
Nous avons simulé une situation où la seule source d'imprécision des mesures de l'indicateur proviendrait de l'erreur d'échantillonnage. Les échantillons de simulation étaient pour une « vraie » prévalence de 1 %, 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 % et 99 %. Nous avons supposé un n de 500, qui était la taille approximative de l'échantillon de la DHS et de l'ONG dans notre base de données. Nous avons ensuite généré une « Estimation de base 1 » en tirant au hasard une distribution binomiale avec une moyenne $n \cdot p$ et une variance $np(1-p)$. Une « Estimation de base 2 » a été générée de la même façon et la différence entre la première et la deuxième estimation a été calculée. Nous avons exécuté 1 000 itérations pour estimer la distribution des différences entre les estimations. Les différences entre les Estimations de base 1 et 2 simulées étaient beaucoup plus faibles que les différences observées entre DHS/MICS-ONG ou DHS-DHS.

PARTIE 4 : COMPARAISON des résultats des Parties 1, 2 et 3

La boîte à moustaches des différences absolues entre les deux estimations (Partie 1 : DHS/MICS – ONG; Partie 2 : DHS – DHS; Partie 3 : Estimation simulée 1 – Estimation simulée 2) est présentée dans le graphique suivante. La différence DHS/MICS – ONG est la plus importante lorsque l'indicateur estimé est compris entre 35 et 65 % et moins aux extrêmes. Les différences DHS – DHS sont très similaires aux différences DHS/MICS – ONG, ce qui suggère que les

différences des méthodes ne sont pas responsables pour les différences importantes observées dans DHS/MICS – ONG (puisque les méthodes DHS sont largement cohérentes entre les années et les niveaux géographiques). Les différences DHS – DHS ne sont probablement que partiellement dues à un échantillonnage aléatoire, car la simulation indique que l'erreur d'échantillonnage ne contribue que faiblement aux différences entre les estimations.

Graphique 2. Différence absolue entre les estimations par la valeur de référence de l'indicateur



Valeur de référence: estimation DHS ou l'estimation qui simule DHS.

DISCUSSION : Notre intérêt pour la réalisation de cette recherche était, qu'en cas de succès, c'est-à-dire si les estimations de la DHS/MICS et des ONG étaient similaires, les ONG pourraient éviter de devoir mener des collectes de données de base et utiliser comme substitut les estimations de la DHS, MICS ou d'autres bases de données publiques. Cela permettrait à l'ONG d'économiser du temps et de l'argent, et de réduire le fardeau pour les répondants.

Bien que nous ne puissions pas donner une recommandation générale selon laquelle DHS et MICS pourraient toujours remplacer les enquêtes de base des ONG, il existe au moins certaines situations dans lesquelles la DHS/MICS pourrait être utilisées à l'avantage des ONG :

- Lorsqu'il est prévu que l'estimation sera inférieure à 15 % ou supérieure à 85 %;
- Lorsque l'indicateur d'intérêt est l'un des rares à présenter une similitude constante entre les estimations de la DHS/MICS et des ONG;
- Lorsque l'ONG a une tolérance pour les estimations avec une précision faible ou inconnue et n'a pas besoin d'estimations pour 100+ indicateurs.

Nous avons émis l'hypothèse que les données publiques peuvent fournir des estimations similaires à celles présentées dans les rapports de base des ONG lorsqu'elles sont les plus proches possibles par rapport au lieu, à l'année et à la saison de collecte de données, et que l'impact des différences d'année, de niveau géographique et de saison varie selon les

indicateurs de santé. Nous avons constaté que la différence moyenne entre les estimations augmentait légèrement en fonction du nombre d'années de différence. Aussi, les estimations dérivées des données de niveaux géographiques inférieurs (comme village ou district pour l'ONG et province pour la DHS/MICS) contribuaient à une différence absolue moyenne plus élevée entre les estimations. En général, des échantillons de plus grande taille ont été obtenus dans les niveaux géographiques plus élevés, et plus la taille de l'échantillon de l'ONG ou de la DHS/MICS est grande, plus la différence absolue moyenne entre les estimations est faible. Le fait que les données ont été recueillies dans la même saison ou non n'impactait pas les différences entre les estimations.

Les analyses de la partition de la variance ont montré que les estimations DHS/MICS et ONG différaient, pour la plupart du temps, de manière imprévisible, et que les différences de niveaux géographiques, d'années et de saison n'expliquaient qu'une petite partie de la variation. Cela indique que la variabilité entre les estimations est plutôt aléatoire ou due aux autres facteurs non capturés dans les modèles.

Plusieurs graphiques et modèles statistiques ont été établis et examinés pour essayer de comprendre comment les différences entre les estimations variaient selon les différences d'année, de niveau géographique et de saison. Aucun modèle ne fournissait d'explications cohérentes du rôle de ces effets dans la différence entre les estimations.

Nous émettons l'hypothèse que des grandes différences entre les estimations des rapports de base des ONG et des bases de données publiques pourraient être dues à trois raisons principales : (i) elles ne mesurent pas la même valeur réelle sous-jacente; (ii) elles ne mesurent pas l'indicateur de la même manière; ou (iii) l'ONG ou la DHS/MICS mesurent l'indicateur avec une erreur technique de mesure élevée.

(i) Il est possible que les estimations des ONG portent sur des populations différentes (la DHS/MICS sont représentatives au niveau national, alors que les ONG essaient souvent de cibler les villages plus marginalisés); ce qui fait que les estimations des ONG semblent moins bonnes que les estimations de la DHS/MICS. Cependant, en comparant les indicateurs liés à la richesse des ménages, nous avons observé que les différences moyennes pour ces indicateurs étaient plus faibles que celles observés pour la plupart des autres indicateurs. Néanmoins, les utilisateurs des données de la DHS/MICS devraient garder à l'esprit la possibilité de différents niveaux de richesse entre leur population cible et les populations de la DHS/MICS.

(ii) Les différentes méthodes utilisées pour l'échantillonnage, la collecte, le traitement et l'analyse des données peuvent également avoir contribué aux différences importantes entre les estimations de la DHS/MICS et des ONG. De nombreux rapports de base des ONG ne fournissaient pas de détails sur les méthodes d'échantillonnage ou d'analyse utilisées, de sorte que nous ne savons pas comment faire correspondre les analyses avec les données de l'enquête DHS/MICS.

(iii) Plusieurs indicateurs liés à la santé maternelle et infantile inclus dans cette étude n'ont pas été validés et certains ont une faible validité, comme l'auto-déclaration des mères sur l'assistance à l'accouchement par un personnel qualifié. La faible précision des auto-déclarations peut entraîner des biais dont la direction et la dimension sont imprévisibles, entraînant de grandes différences entre les estimations.

Quelle que soit la cause des grandes différences entre les estimations, il n'a pas été possible de savoir laquelle des sources de données (DHS/MICS ou ONG) a fourni l'estimation la plus précise de la prévalence réelle dans la population cible des ONG. De plus, bien que nous comparions les estimations ponctuelles de la DHS/MICS et d'ONG, les indicateurs sont mesurés avec une erreur et un intervalle de confiance typique de 95 % pour les estimations de la DHS/MICS et d'ONG serait de +/- 5 % à 10 %.

Une grande partie des analyses de ce rapport documentent les différences entre les estimations d'ONG et de la DHS/MICS. Néanmoins, dépendamment des objectifs, les données publiques peuvent être utilisées dans certains cas pour compléter ou remplacer les études de base et pour obtenir des estimations d'indicateurs importants. Les études de cas du Népal et du Vietnam ont montré des exemples d'indicateurs pour lesquels les estimations des DHS/MICS et des ONG étaient similaires.

Les grandes enquêtes sur la santé dans les populations peuvent être une source précieuse de données pour comprendre les relations complexes, telles qu'entre pauvreté et santé, et lors de la planification des interventions, car elles fournissent des données standardisées pour différents pays et régions. L'utilisation de données secondaires peut être intéressante dans des situations de restriction budgétaire ou de mobilité réduite, comme lors de la pandémie de la COVID-19 durant laquelle les opportunités de collecte de données sont limitées. Cependant, lors de l'utilisation des données de la DHS/MICS, l'utilisateur doit tenir compte des limites mises en évidence dans ce rapport.

CONCLUSION : Notre première hypothèse était que les données publiques peuvent fournir des estimations de conditions de base similaires à celles dans les rapports des ONG lorsqu'elles sont les plus proche possible par rapport au lieu, à l'année et à la saison de collecte des données. Notre réponse, en bref, est que les données publiques peuvent être utilisées, si l'ONG tolère des estimations imprécises.

Notre deuxième hypothèse était que l'impact des différences d'année, de niveau géographique et de saison varie selon les indicateurs de santé. Notre réponse est oui, mais la variation due à l'année, au niveau géographique et à l'indicateur de santé n'explique qu'une petite partie de la variation totale, et même après avoir tenu compte de ces sources de variation, il y aura toujours des différences entre les estimations à partir des deux sources de données.

Bien qu'une ONG puisse utiliser les résultats présentés ici pour justifier de ne pas faire une enquête de base, elle doit être consciente que la DHS et la MICS ne fournissent des estimations que pour certains indicateurs. En moyenne, nous avons estimé 18 des estimations rapportées

par l'ONG à l'aide de la DHS/MICS, mais les ONG rapportaient souvent 100+ estimations, ce qui pourrait justifier la réalisation d'une enquête de base complète. De plus, être sur le terrain peut fournir des informations précieuses pour la conception et la mise en œuvre de projets.



CanWaCH
Canadian Partnership for
Women and Children's Health

CanSFE
Partenariat canadien pour la
santé des femmes et des enfants

