

Planification d'un système d'enregistrement par sondage

Plans d'échantillonnage et utilisation des données

Agbessi Amouzou, PhD

Université Johns Hopkins

8 septembre 2025

Aperçu

- Pourquoi se soucier de la conception d'un échantillonnage dans un SRS ?
- Éléments d'un plan d'échantillonnage
- Taille de l'échantillon
- Procédures d'échantillonnage et pondération
- Autres considérations
- Utilisation des données
- Exemple de la Zambie – (Stephen)

PLANS D'ÉCHANTILLONNAGE

Pourquoi se soucier de la conception de l'échantillonnage dans un SRS ?

1. Garantir la représentativité de l'échantillon SRS aux niveaux national et infranational (provinces, districts)
 - La probabilité que tout le monde soit inclus dans l'échantillon est connue
2. Estimer rigoureusement la précision ou les intervalles de confiance autour des indicateurs dérivés des données SRS
3. Garantir que les indicateurs dérivés sont crédibles, fiables et dignes de confiance
4. Garantir la transparence dans la sélection de la population incluse
5. Promouvoir la reproductibilité et la reproductibilité
6. Comparer les estimations issues d'un échantillonnage aléatoire simple à d'autres ensembles de données

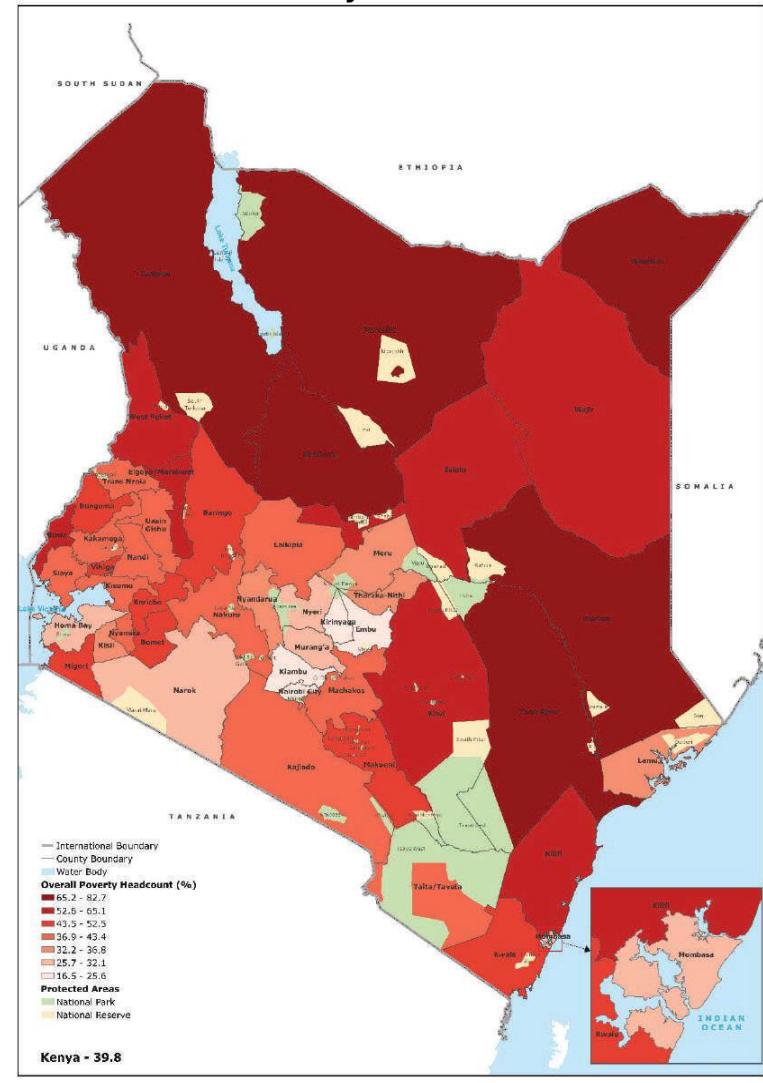
Éléments d'un plan d'échantillonnage

1. Compréhension des priorités gouvernementales en matière de santé et des utilisateurs potentiels des données du SRS
2. Niveau de représentativité de l'échantillon (domaines statistiques)
 - National, régions/provinces, districts, urbain/rural ?
3. Strates de l'échantillon, c'est-à-dire groupes à partir desquels un échantillon indépendant est tiré
 - par exemple, régions urbaines/rurales
4. Plus petites unités géographiques d'échantillonnage (ou unités d'échantillonnage primaires ou grappes)
 - Par exemple, zones de recensement, villages, districts, sous-districts
5. Cadre d'échantillonnage complet des grappes, organisé par domaines statistiques (souvent disponible auprès de l'Office national de la statistique)
 - La base de recensement est recommandée car elle est souvent complète et mise à jour lors des recensements de population
6. Taille de l'échantillon
7. Procédures de sélection des échantillons et pondérations des échantillons

1. Priorités sanitaires du gouvernement et utilisateurs potentiels des données SRS

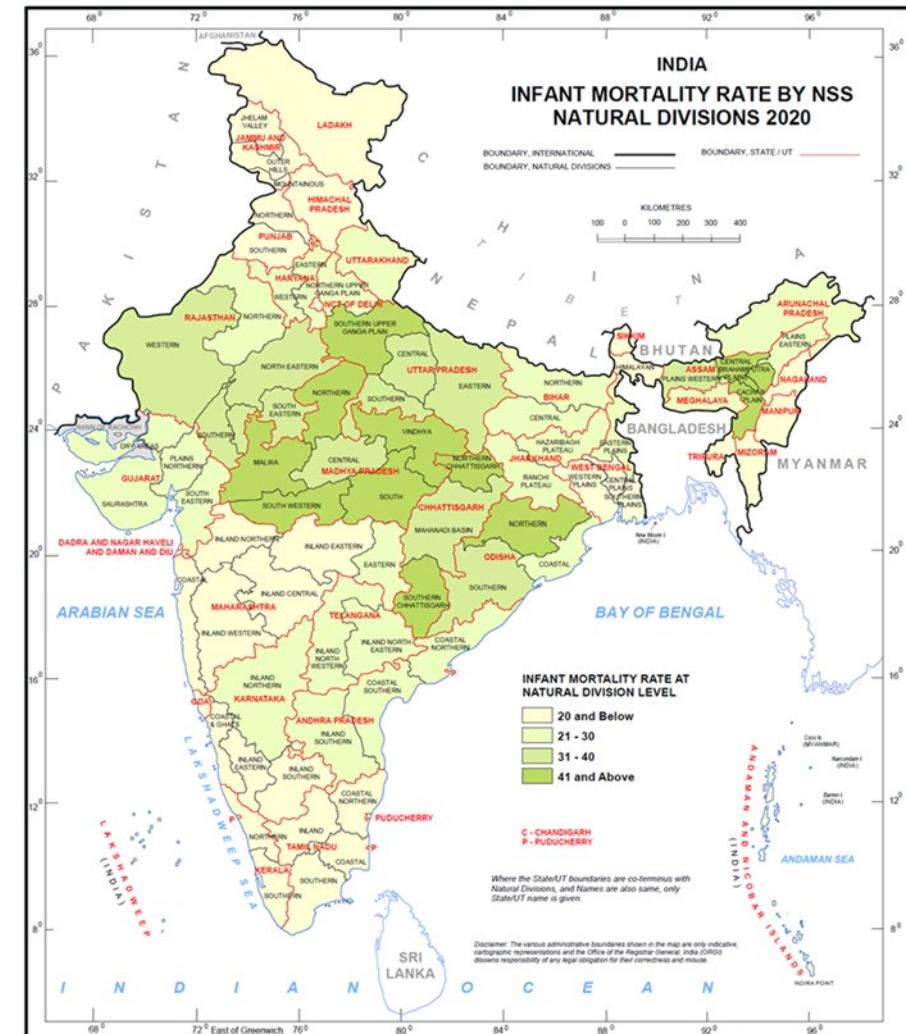
- Comprendre les domaines prioritaires du gouvernement au niveau infranational qui nécessitent des données fréquentes et plus précises pour des interventions sanitaires ciblées ou l'évaluation des performances
 - Par exemple, à partir des plans stratégiques nationaux en matière de santé
 - Impliquer le ministère de la Santé et d'autres secteurs
- Qui sont les autres utilisateurs directs des données et quel niveau de granularité souhaitent-ils voir ?
 - Partenaires mettant en œuvre de grands programmes de santé pluriannuels
- Lien potentiel avec d'autres données existantes

Niveau global de pauvreté au niveau des comtés,
Kenya 2022



2. Niveau de représentativité de l'échantillon (domaines statistiques)

- Les domaines statistiques sont les zones géographiques les plus petites pour lesquelles des indicateurs précis seront générés.
 - Par exemple, l'Inde utilise les divisions naturelles
 - Ceci est déterminé en consultation avec le gouvernement et les parties prenantes
 - Utilise généralement des zones administratives infranationales
 - Par exemple, régions/provinces, zones urbaines/rurales, districts
 - Certaines zones infranationales peuvent faire l'objet d'un suréchantillonnage
 - Par exemple, suréchantillonner une région pauvre pour produire des estimations au niveau du district
 - Plus vous créez de domaines statistiques, plus votre échantillon est important et plus votre coût est élevé



3. Strates d'échantillonnage

- Groupes d'unités d'échantillonnage à partir desquels un échantillon indépendant est prélevé
 - Par exemple, zones rurales/urbaines dans chaque région, populations/établissements particuliers
- La stratification peut être utilisée pour garantir la représentation de certains groupes dans votre échantillon, en particulier les groupes minoritaires qui pourraient ne pas être sélectionnés dans un échantillon non stratifié
- Les domaines statistiques peuvent être utilisés comme strates
- La stratification crée des échantillons à probabilité inégale qui nécessitent l'utilisation de pondérations d'échantillonnage (voir les diapositives suivantes)

4. Plus petites unités géographiques d'échantillonnage (grappes)

- Plus petites unités d'échantillonnage dans lesquelles une surveillance complète est mise en place
- Doivent être déterminées en fonction de la disponibilité de la liste complète
- Les zones de recensement de la population (EA) offrent des unités d'échantillonnage plus complètes et plus rigoureuses
 - Issues du dernier recensement de la population
 - Souvent 100 à 150 ménages
 - Des cartes numériques et des croquis sont souvent disponibles auprès de l'Office national de la statistique
 - Couvrent l'ensemble du pays et sont stables
 - Cependant, elles peuvent traverser des villages ou des communautés
- Les zones infranationales, telles que les villages/blocs, les districts/communes, les sous-districts, peuvent être utilisées, mais peuvent être instables dans le temps
- Les limites doivent être clairement identifiables



5. Cadre d'échantillonnage complet des grappes

- Liste complète des grappes (par exemple, les zones économiques) classées par unités administratives, domaines statistiques, strates et population
- Disponible auprès de l'Office national de la statistique, avec des cartes et, dans certains cas, les coordonnées géographiques des limites et des points de repère clés
 - Il faudra travailler avec l'ONS, car la liste peut ne pas être accessible à tout le monde
- Il n'est pas nécessaire que la liste soit entièrement à jour en termes de taille de la population, mais elle doit couvrir l'ensemble de la zone d'échantillonnage
- Éviter les bases de sondage qui changent trop rapidement (par exemple, liste des villages, districts)

Région	District	Village/commune	Id Zone de dénombrement	Nombre de ménages	Population
1	11	111	1111111	150	750
1	11	112	1111112	125	700
1	11	113	1111131	90	500
1	11	113	1111132	100	500
1	11	114	1111141	120	600
...

6. Taille de l'échantillon

Facteurs déterminants de la taille de l'échantillon :

1. Indicateurs à mesurer
2. Précision statistique et domaines
3. Période de mesure des indicateurs de mortalité
4. Priorités sanitaires du gouvernement
5. Budget

6. Taille de l'échantillon : choix des indicateurs

- Indicateurs de mortalité
 - ✓ **Taux brut de mortalité** = décès annuels / population en milieu d'année
 - ✓ **Taux de mortalité infantile** = décès infantiles (T) / naissances vivantes (T)
 - ✓ **Taux de mortalité des moins de cinq ans** = décès des moins de cinq ans (T) / naissances vivantes (T)
 - ✓ **Taux de mortalité des adultes** = décès annuels (15-64 ans) / population adulte en milieu d'année
 - ✓ **Taux de mortalité maternelle** = décès maternels (T) / naissances vivantes (T)
 - ✓ **Taux de mortalité par cause** = décès par cause (T) / population cible (T)
 - ✓ Autres...
- Choisissez l'indicateur susceptible de générer des échantillons suffisamment grands pour mesurer la plupart des autres indicateurs, mais pas trop grands au point d'être irréalistes.
 - Par exemple, le TMI
- Une fois la taille de l'échantillon calculée, calculez la précision des autres indicateurs.
- La taille de l'échantillon doit être calculée pour chaque domaine statistique

6. Calcul de la taille de l'échantillon

- m = taux de mortalité (en proportion)
- d = marge d'erreur absolue, c'est-à-dire que l'intervalle de confiance est $m \pm d$
- R = taux de réponse des ménages
- $Deff$ = effet de conception (emprunté à des enquêtes existantes, par exemple DHS)
- N est la population cible correspondant au dénominateur de l'indicateur (par exemple, N est le nombre de naissances si l'on utilise le TMI)
- N doit être converti en population totale et en nombre de ménages correspondants

$$N = Z_{\alpha/2}^2 * \frac{m * (1 - m)}{d^2 * R} * Deff \quad \longrightarrow \quad d = \text{SQRT}(Z_{\alpha/2}^2 * \frac{m * (1 - m)}{N * R} * Deff)$$

6. Taille de l'échantillon : Exemple

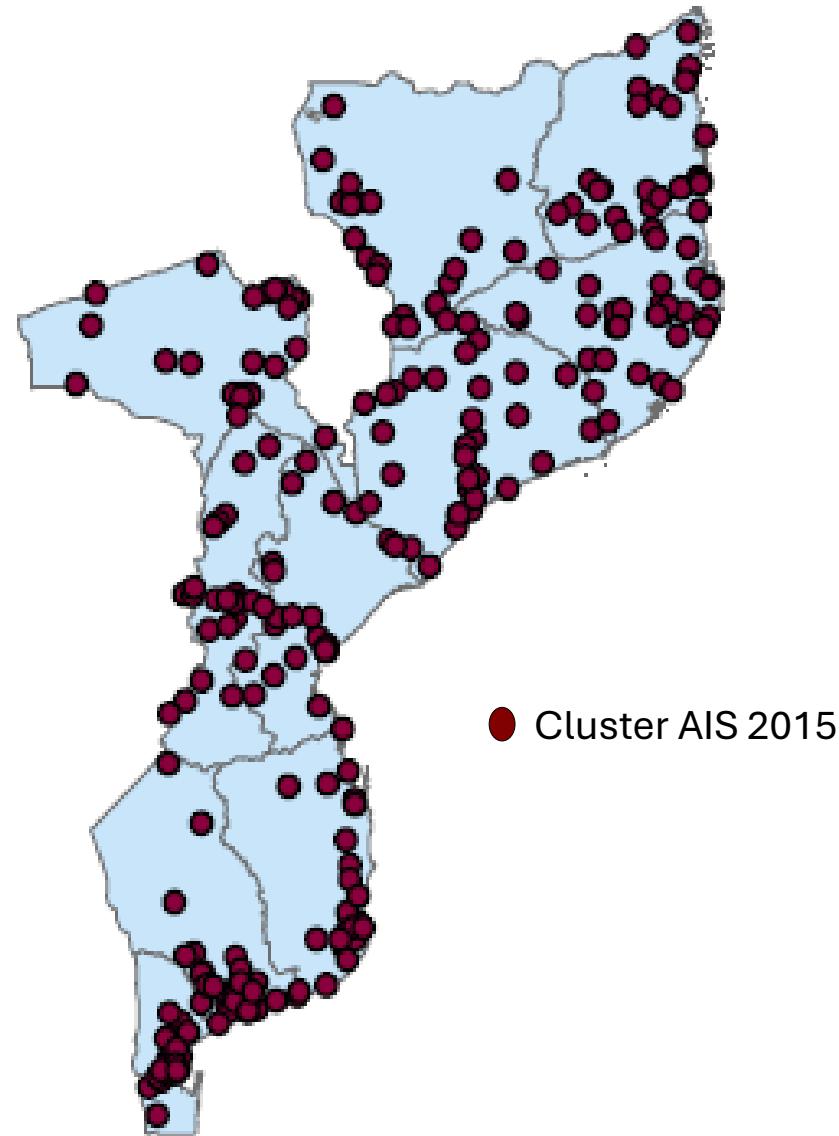
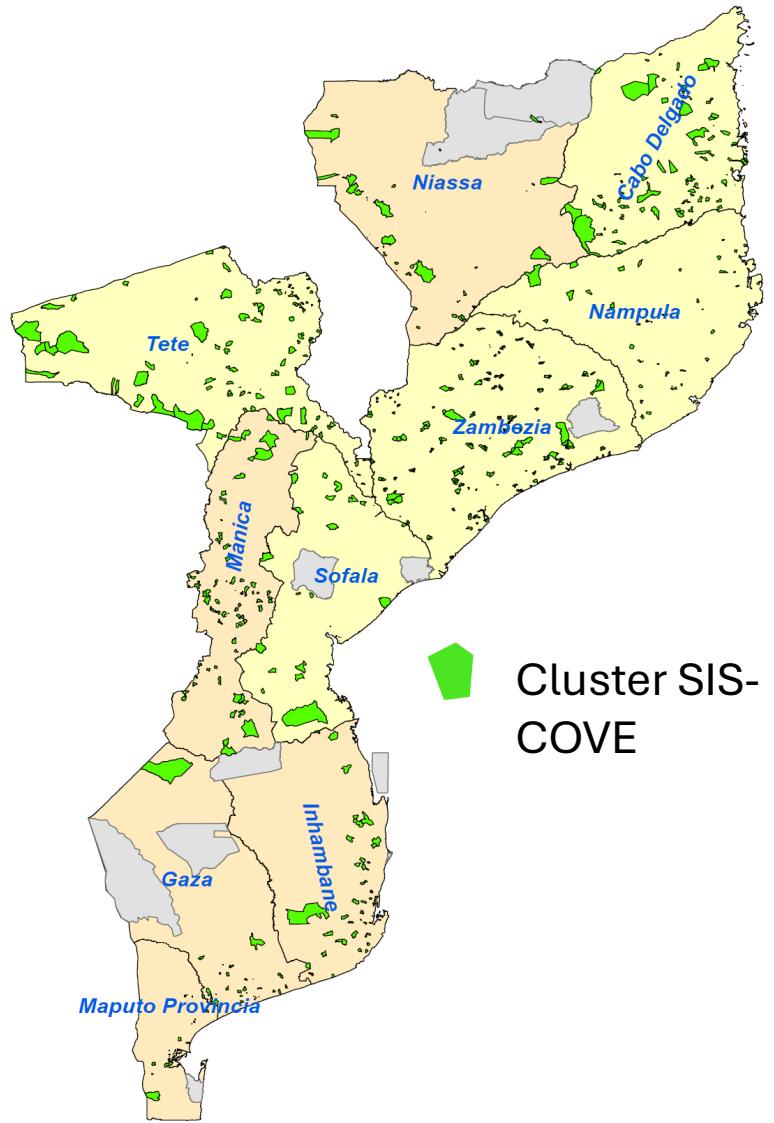
Domaines	IMR (pour 1 000 LB)	Marge d'erreur relative (2SE/M)	Marge d'erreur absolue	Nombre estimé de naissances (annuel)	Taux brut de natalité (pour 1000)	Taille moyenne des ménages	Nombre correspondant de ménages	Population correspondante	Nombre d'unités d'échantillonnage (en supposant 125 ménages par unité)	Nombre estimé de décès infantiles
Domaine 1	100	12	12,0	3 961	40	5	19 805	99 023	158	396
Domaine 2	90	12	10,8	4 450	35	5	25 428	127 141	203	400
Domaine 3	50	20	10,0	3 010	30	5	20 069	100 344	161	151
Domaine 4	80	12	9,6	5 061	40	5	25 306	126 530	202	405
Domaine 5	150	10	15,0	3 591	45	5	15 961	79 806	128	539
TOTAL				20 074		5	106 569	532 844	853	1 891

Hypothèse : effet de conception = 1,5 ; taux de non-réponse = 10 %

7. Procédures de sélection des échantillons et pondérations de conception

- Sélectionner l'échantillon au sein de chaque strate identifiée
- Organiser la base de sondage par unités administratives (stratification implicite)
- Sélectionner les unités d'échantillonnage primaires (grappes) de manière aléatoire avec une probabilité proportionnelle à leur taille
 - Échantillonnage aléatoire systématique
- Représenter les grappes échantillonnées sur une carte et vérifier que l'échantillonnage est acceptable
- Comparer avec d'autres données existantes si nécessaire

Échantillon SIS-COVE du Mozambique comparé à l'enquête nationale AIS 2015



7. Pondérations de conception : inverse de la probabilité de sélection au sein de la strate de l'échantillon

- Lorsque toutes les unités de l'échantillon n'ont pas la même probabilité de sélection, un poids de conception est nécessaire pour ajuster l'échantillon pendant l'analyse afin d'obtenir des estimations représentatives

	Strates			National
	1	2	3	
Population totale	P1	P2	P3	$P = (P1+P2+P3)$
Population de l'échantillon	S1	S2	S3	$S=(3*S)$
Probabilité de sélection	$S1/P1$	$S2/P2$	$S3/P3$	



Pondérations de l'échantillon	$P1/S1$	$P2/S2$	$P3/S3$	
-------------------------------	---------	---------	---------	--

→ S n'est pas toujours représentatif de P

7. Pondération de la conception

- Procédure d'échantillonnage : échantillonnage en grappes stratifié à un seul degré avec probabilité proportionnelle à la taille de la population
 - C_s = nombre total de grappes dans la strate S
 - N_s = nombre de grappes sélectionnées dans la strate S
 - Grappes numérotées de i à C_s dans la strate S
 - C_{is} = taille de la population du groupe I de la strate S
 - M_s = population totale dans la strate S

$$\text{Probability of selection of cluster } i \text{ in stratum } S = P_{is} = N_s * \frac{C_{is}}{M_s}$$

$$\text{Design Weight for cluster } i \text{ in stratum } S = W_{is} = \frac{1}{P_{is}}$$

- Les pondérations de conception peuvent être ajustées pour tenir compte des non-réponses au niveau des grappes et des ménages
- Il est essentiel de documenter les procédures d'échantillonnage afin de permettre le calcul des pondérations de conception

Qui impliquer dans la conception de l'échantillon ?

- Les experts en échantillonnage de l'Office national de la statistique
 - Calculer la taille de l'échantillon
 - Tirer l'échantillon
 - Calculer les pondérations de l'échantillon
- Ministère de la Santé et autres parties prenantes
 - Pour obtenir des informations sur la répartition de l'échantillon
- Démographe, statisticien ou épidémiologiste

Ressources sur Viva

- Rendez-vous sur viva.jhuhost.org pour accéder à la bibliothèque d'outils qui facilitent la conception d'échantillonnages.

Devoirs pour les pays qui envisagent de mettre en place un SRS

Remplissez ce tableau pour résumer votre plan d'échantillonnage

Activités	Décrire	Qui a été impliqué
1 Priorités sanitaires du gouvernement et utilisateurs potentiels des données SRS		
2 Niveau de représentativité de l'échantillon (domaines statistiques)		
3 Strates de l'échantillon		
4 Plus petites unités géographiques d'échantillonnage (grappes)		
5 Disponibilité d'une base d'échantillonnage complète des grappes		
6 Taille de l'échantillon		
7 Procédures de sélection de l'échantillon et pondérations de l'échantillon		

UTILISATION DES DONNÉES

Utilité des données SRS

1. Comprendre et surveiller les tendances et les disparités en matière de santé
 - ✓ Tendances en matière de mortalité toutes causes confondues
 - ✓ Principale cause de décès
 - ✓ Charge de morbidité
 - ✓ Amélioration de la recherche en matière de santé spécifique à chaque pays
 - ✓ Ventilation par âge, sexe, zone géographique et statut socio-économique
2. Production de données probantes pour éclairer les interventions prioritaires
 1. Population la plus exposée (par exemple, nouveau-nés, mères, personnes âgées)
 2. Disparités
 3. Allocation des ressources
3. Évaluation de l'efficacité des politiques et des programmes
 1. Modèles d'évaluation rigoureux (avant/après avec comparaison)
 2. Évaluations d'impact
4. Alerte précoce en cas de crise
 1. Pandémies
 2. Intervention d'urgence en cas de catastrophe
 3. Tendances inhabituelles en matière de mortalité
5. Relier et évaluer d'autres systèmes de données
 1. CRVS, HMIS, autres

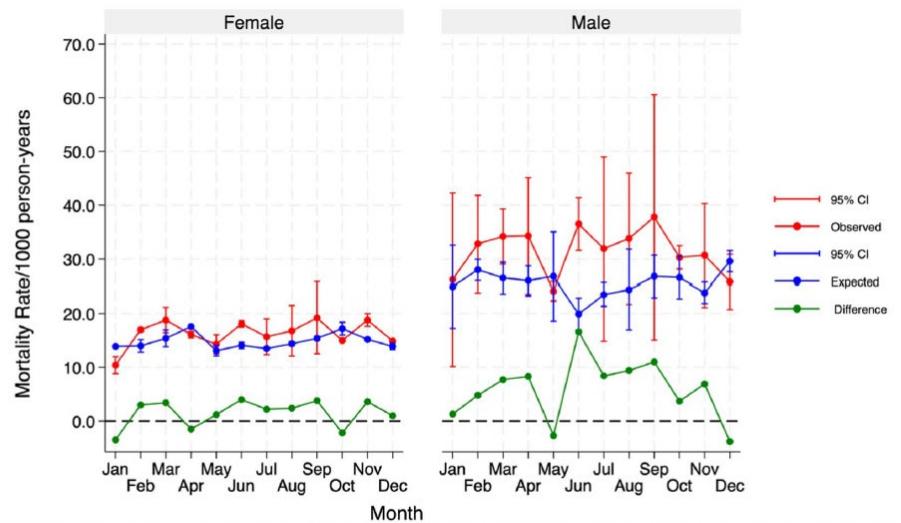
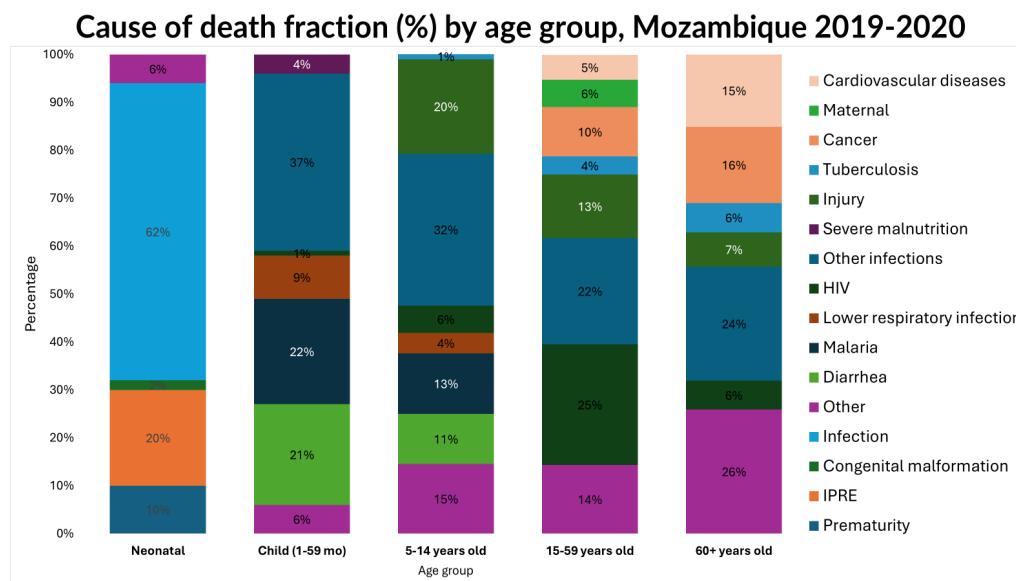


Fig. 4 Monthly trends in excess mortality by gender during the COVID-19 pandemic (2020–2021) (using outputs from the predictions of the Poisson generalised additive model) in the Navrongo HDSS study area

Azongo et al. *Population Health Metrics* (2025) 23:31
<https://doi.org/10.1186/s12963-025-00389-7>

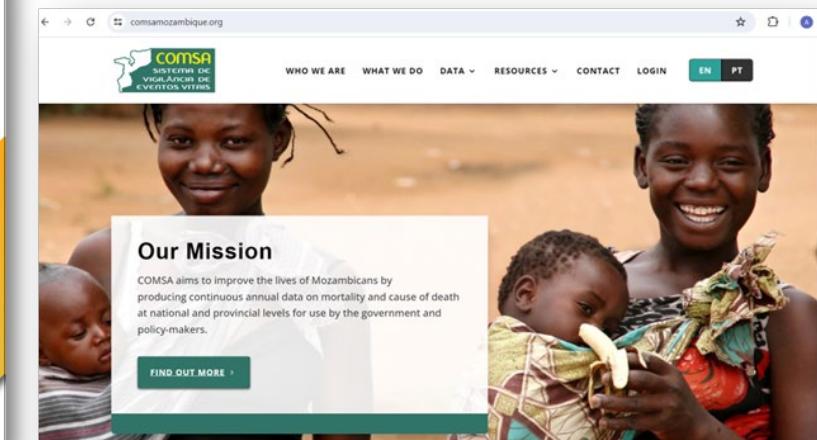
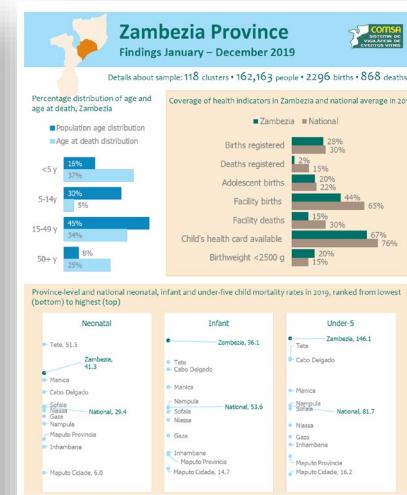


Qui sont les utilisateurs des données ?

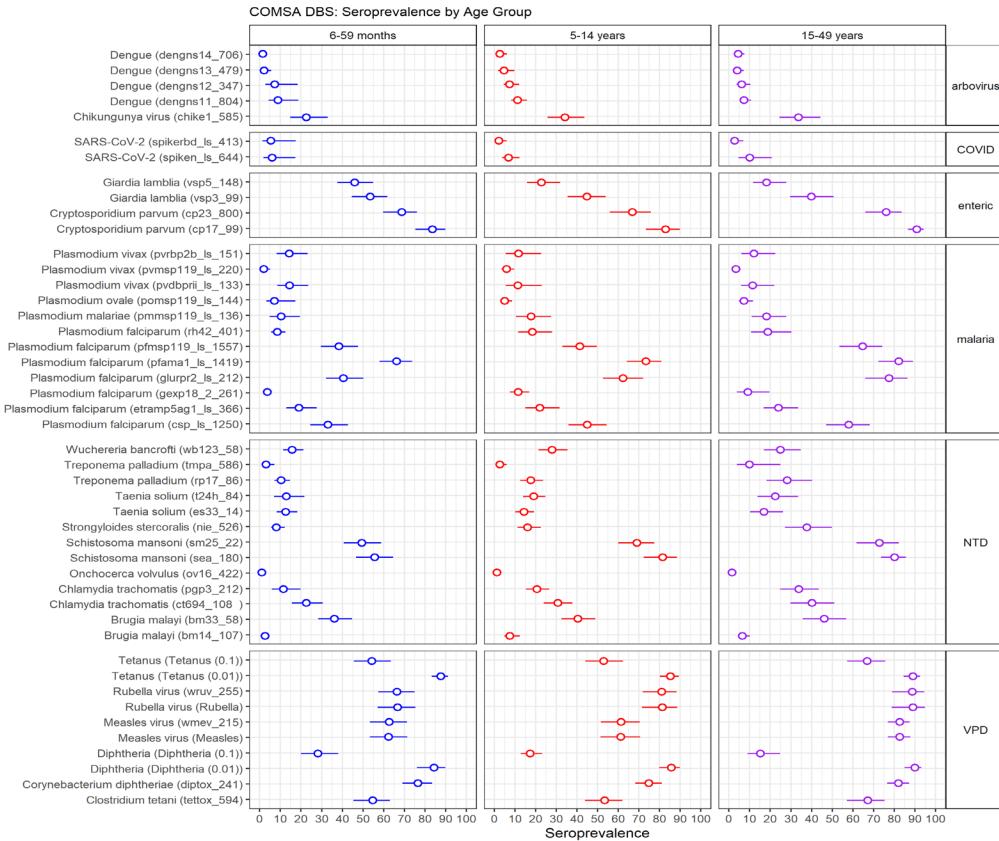
UTILITÉ	UTILISATEURS
1. Compréhension et suivi des tendances et des disparités en matière de santé	Ministère de la Santé, autres secteurs (par exemple, nutrition, éducation, genre et protection sociale) ; responsables sanitaires infranationaux ; partenaires ; institutions de recherche et de statistiques en santé publique ; associations de professionnels de santé ; communauté mondiale de la santé
2. Produire des données probantes pour éclairer les interventions prioritaires	Décideurs du ministère de la Santé ; responsables sanitaires infranationaux ; bailleurs de fonds ;
3. Évaluer l'efficacité des politiques et des programmes	Responsables de programmes de santé ; bailleurs de fonds ; partenaires techniques ; responsables sanitaires infranationaux
4. Servir d'alerte précoce en cas de crise	Instituts de santé publique ; département de la santé publique ; bureau de la santé humanitaire
5. Relier et évaluer d'autres systèmes de données	CRVS, HMIS,

Quelles plateformes pour la diffusion et le partage des données ?

1. Rapports annuels contenant des données détaillées
2. Rapport de synthèse et fiches d'information aux niveaux national et infranational
3. Groupes nationaux spécifiques (par exemple, groupe consultatif national, groupes de direction du ministère de la Santé)
4. Plateformes web avec modalités d'accès aux données publiques (site web avec visualisation des données actualisées et interactives)
5. Lien avec la plateforme d'autres systèmes (par exemple, DHIS-2)
6. Publications évaluées par des pairs

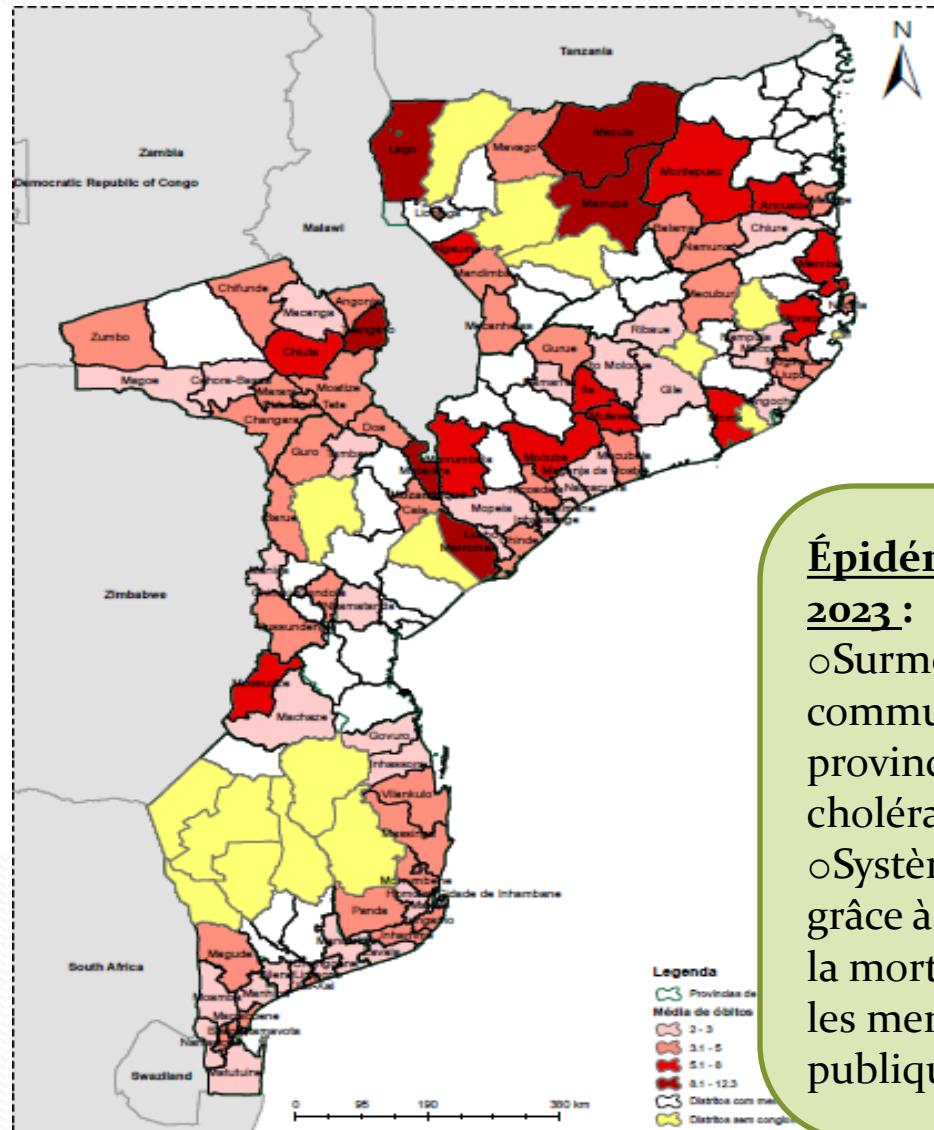


Utilisation du SIS-COVE pour améliorer la surveillance communautaire des maladies et les enquêtes sur les épidémies



Séroprévalence des maladies évitables par la vaccination en Zambézie :

- Diphthéries : 87,2 % (IC à 95 % : 84,3-89,6)
- Tétanos : 86,9 (IC à 95 % : 83,6-89,6)
- Rubéole : 83,0 % (IC à 95 % : 78,3-86,8)
- Rougeole : 69,1 % (IC à 95 % : 63,3-74,4)



Épidémie de choléra en 2023 :

- Surmortalité dans une communauté isolée de la province de Tete due au choléra
- Système d'alerte précoce grâce à la surveillance de la mortalité pour détecter les menaces pour la santé publique

Mozambique : les données sur la mortalité du SIS-COVE sont utilisées pour améliorer le niveau d'alphabétisation général et pour la prise de décisions politiques dans différents secteurs

SIS-COVE et DHS sont les principales sources de données utilisées pour élaborer le Plan stratégique du secteur de la santé



Réunions



Radio



Journaux

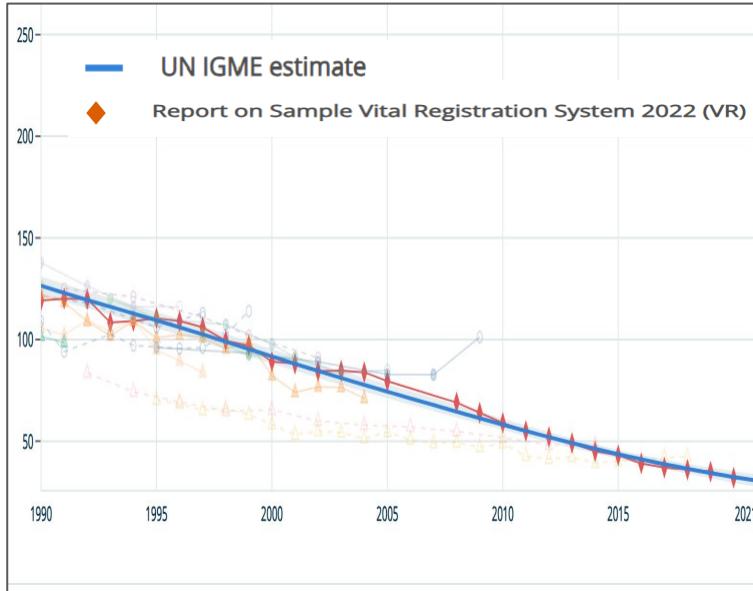


Interviews en direct

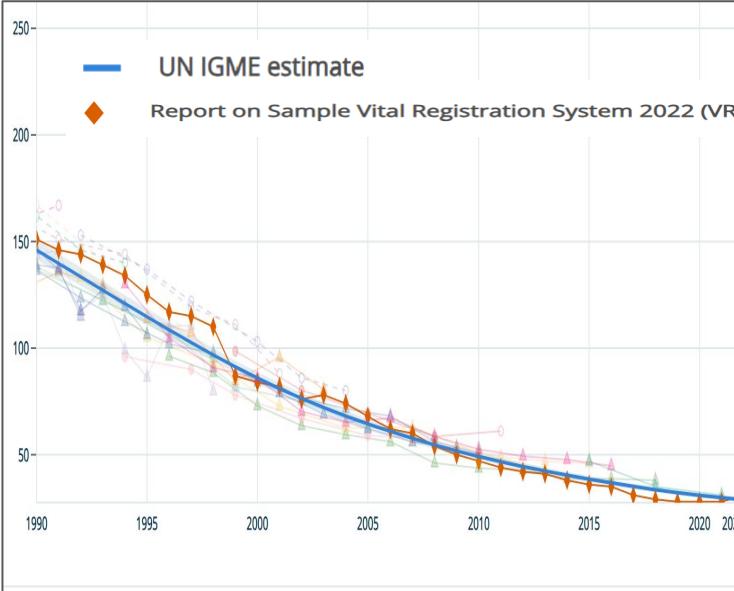


Les données SRS alimentent les estimations de la mortalité infantile par l'UN-IGME

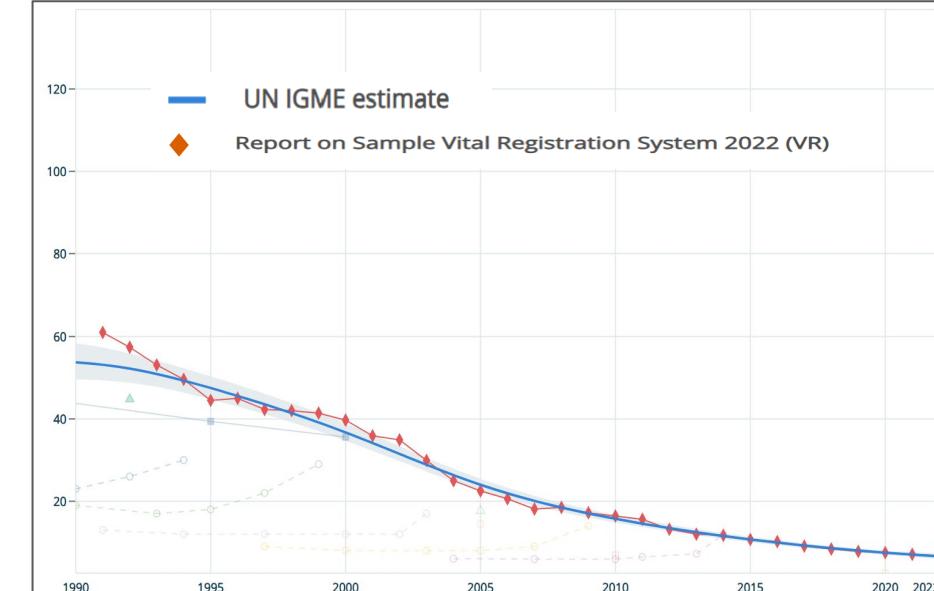
Inde



Bangladesh



Chine



- Lorsqu'il est disponible et fonctionne correctement, le SRS peut être l'une des meilleures sources de données sur la mortalité.

Soyez conscient des obstacles à l'utilisation des données

1. Caractère inopportun des résultats
2. Qualité des données et manque de transparence sur la manière dont les données et les résultats sont obtenus
3. Lien ou interopérabilité limités avec d'autres systèmes
4. Faible capacité d'analyse et d'interprétation des données
5. Préoccupations éthiques et juridiques concernant la confidentialité des données

Merci !

Données minimales dans un SRS

Numérateurs :

- Nombre de décès et leur répartition par âge, sexe, zone géographique et cause de décès

Dénominateurs :

- Naissances (mort-nés, mortalité périnatale, taux de mortalité infantile)
- Population par âge et par sexe (années-personnes ou population à risque)

> Un SRS doit au minimum collecter des données sur la population par âge et par sexe, les issues de grossesse, les décès et les causes de décès

Taux de mortalité infantile

$$IMR = \frac{\text{Infant deaths (T)}}{\text{Births (T)}}$$

Taux de mortalité par âge

$${}_nM_x = \frac{{}_nD_x(T)}{{}_nPYL_x(T)}$$

PYL : années-personnes vécues

${}_nM_x$ = taux de mortalité entre les âges x et x+n

Probabilité de décès par âge

$${}_nq_x = \frac{{}_nM_x}{[1 + (n-s){}_nM_x]}$$

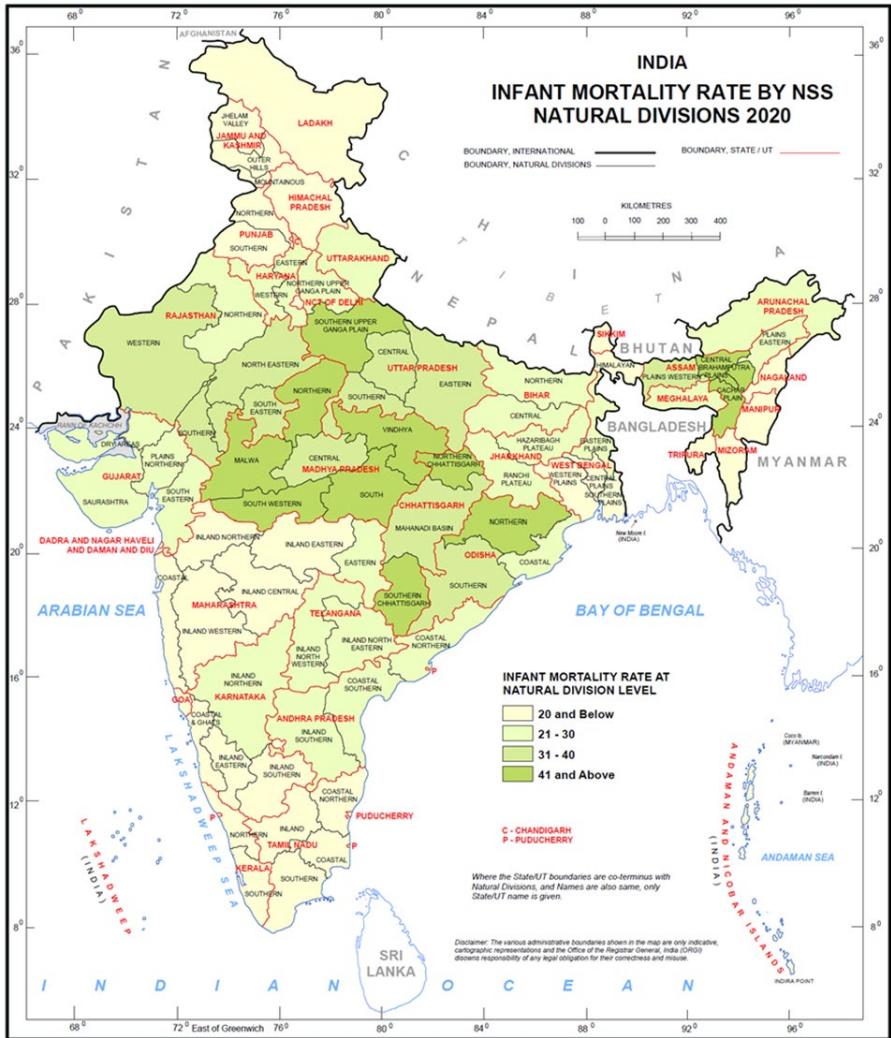
Expériences actuelles en matière d'échantillonnage

Bangladesh



- Début en 1980
- Base de sondage : recensement de 2022 EA
- Taille de l'échantillon en 2024 :
 - 2 766 zones d'échantillonnage
 - 313 140 ménages
 - > 1,3 million d'habitants
- Méthode d'échantillonnage
 - Domaines statistiques et strates : 7 divisions X zones urbaines/rurales/municipalités
 - Échantillonnage aléatoire stratifié à un seul degré des zones d'enquête avec PPS
- L'échantillon représente 0,8 % de la population

Inde



- Début : 1968-70
 - Base d'échantillonnage : recensement de 2011 EA
 - Taille de l'échantillon basée sur le TMI au niveau des divisions naturelles (marge relative de 10 % pour les grands Etats et de 15 % pour les petits États)
 - Échantillon : 8 853 zones d'échantillonnage (2014)
 - Domaines d'échantillonnage : divisions naturelles ou petits États
 - Sélection de l'échantillon : mélange d'échantillonnage aléatoire stratifié à un ou deux degrés (en fonction de la taille des strates)
 - L'échantillon représente 0,7 % de la population totale

Chine

Fig. 2. Surveillance points, national mortality surveillance system, China, 2013



Note: Dark green lines represent 447 counties with national mortality surveillance points; light green lines represent 158 counties which also have disease surveillance points. The map was produced using QGIS version 2.8.3 (QGIS Development Team).

Source: Map data for China from the Natural Resources and Geospatial Base Information Database (National Development and Reform Commission, China). Map data for neighbouring countries from WHO.

- Intégration des points de surveillance VS et des maladies en 2013
- Base d'échantillonnage : recensement de la population de 2010
- Grappe : comtés/districts
- Échantillon : 605 grappes déterminées sur la base d'une consultation d'experts
- Procédure d'échantillonnage : sélection itérative contrôlant certains indicateurs démographiques et socio-économiques afin de garantir la représentativité
- L'échantillon représente 24 % de la population totale

Utilité d'un SRS pour d'autres systèmes de surveillance

Systèmes existants	Rôles du SRS
CRVS	<ul style="list-style-type: none">■ Évaluer l'exhaustivité■ Promouvoir l'enregistrement des événements d'état civil■ Faciliter la mise en œuvre d'études de recherche ciblées afin de soutenir une expansion rapide
HMIS	<ul style="list-style-type: none">■ Compléter avec les données communautaires■ Comprendre la cause des décès au niveau communautaire■ Comprendre le profil des utilisateurs hors établissement■ Servir de plateforme pour évaluer et tester les approches communautaires
Autres systèmes de surveillance spéciaux (par exemple, MPDSR)	<ul style="list-style-type: none">■ Compléter avec les données communautaires■ Évaluer l'exhaustivité et la qualité des données