# Addendum au Guide d'Echantillonnage FANTA par Robert Magnani (1997) : Correction de la Section 3.3.1 Déterminer le nombre de ménages qui ont besoin d'être contactés

Diana Stukel
Megan Deitchler

Mars 2012

**Food and Nutrition Technical Assistance III Project (FANTA)**
FHI 360 1825 Connecticut Avenue, NW Washington, DC 20009-5721
**T** 202-884-8000 **F** 202-884-8432 fantamail@fhi360.org www.fantaproject.org

Cet addendum a été rendu possible grâce au généreux soutien du peuple américain avec le soutien du Bureau de la Nourriture pour la Paix ; le Bureau pour la Démocratie, le Conflit, et l'Aide Humanitaire ; et le Bureau de la Santé, les Maladies Infectieuses, et de l'Alimentation, le Bureau pour une Santé Globale ; l'Agence des États-Unis pour le Développement International (USAID), aux termes de l'Accord de Coopération Numéro AID-OAA-A-12-0005, par le biais du FANTA, géré par la FHI 360.



Publié en mars 2012

**Citation recommandée :**

Stukel, Diana y Deitchler, Megan. 2012. *Addendum au Guide d'Echantillonnage FANTA par Robert Magnani (1997) : Correction de la Section 3.3.1 Déterminer le nombre de ménages qui ont besoin d'être contactés.* Washington, DC: FHI 360/FANTA.

**Contact information :**

Le Pont du Projet III d'Assistance Technique sur l'Alimentation et la Nutrition (FANTA)
FHI 360
1825 Connecticut Avenue, NW
Washington, D.C.  20009-5721
Tél: 202-884-8000
Fax: 202-884-8432
Email: fantamail@fhi360.org
Site web: www.fantaproject.org

# Table des Matières

## Remerciements

Les auteurs tiennent à exprimer leur reconnaissance pour les contributions des membres suivants du personnel de la FHI 360 : Sandy Remancus, Gilles Bergeron, Vicky Michener et Laura Glaeser.

# 1. Introduction

Le Guide d'Echantillonnage FANTA fournit des indications sur la façon de calculer la taille de l'échantillon pour les enquêtes de base et d'évaluation finale effectuées par les programmes Nourriture pour la Paix Titre II (FFP / TII). En règle générale, le calcul de la taille d'échantillon est conduit par un indicateur anthropométrique, comme celui relatif au retard de croissance ou à l'insuffisance pondérale, ce qui nécessite la collecte de données sur les enfants âgés de moins de 5 ans. Dans des cas pareils, la taille initiale calculée de l'échantillon reflète le nombre d'enfants dans cette tranche d'âge pour lesquels les données sont requises. Cependant, la plupart des enquêtes utilisent les ménages (ou les habitations) plutôt que les enfants en tant qu'une base pour l'échantillonnage dans les groupes. Il est de lors plus typique que les ménages soient échantillonnés d'abord, et que les données sur les enfants admissibles qui résident dans les ménages échantillonnés soient recueillies ensuite. Cela remet en question le calcul de taille de l'échantillon parce que le nombre des ménages ne correspond pas forcement au nombre des enfants.

Afin de résoudre ce problème, le Guide d'Echantillonnage FANTA propose une approche (voir Section 3.3.1) pour transformer la taille calculée de l'échantillon pour les enfants en nombre de ménages qui ont besoin d'être visités afin de s'assurer que le nombre d'enfants requis pour l'échantillon est obtenu. En revanche, l'expérience passée sur le terrain sur certaines enquêtes de base et l'évaluation finale menées par les programmes FFP / TII, ont montré que l'approche proposée par le Guide d'Echantillonnage FANTA peut parfois aboutir à une sous-estimation du nombre de ménages qui devraient être visités. Dans cet addendum nous proposons une approche alternative qui donnera lieu à un échantillon des ménages de taille plus importante que celle suggérée par le Guide d'Echantillonnage FANTA actuel, mais qui est plus certain de fournir la taille requise de l'échantillon des enfants.

# 2. Calculer la taille initiale requise pour l'échantillon

La formule pour la taille requise de l'échantillon, n (pour les enfants, si le calcul est basé sur un indicateur anthropométrique), donnée dans la section 3.3.1 du Guide d'Echantillonnage FANTA, est également donnée dans l'annexe 1 du présent addendum. La taille de l'échantillon résultant de cette formule montre le nombre d'unités sur lesquelles des données pour cet indicateur devraient être recueillies dans l'enquête de base et dans l'évaluation finale à la fois. Par conséquent, si l'indicateur utilisé pour le calcul de la taille de l'échantillon est « la prévalence du retard de croissance chez les enfants de 0 à 59 mois », la taille de l'échantillon calculée avec cette formule refléterait le nombre requis d'enfants de 0 à 59 mois qui devraient être échantillonnés dans l'enquête de base et dans l'évaluation finale à la fois.

En revanche, lors de la réalisation d'une enquête de base ou d'une évaluation finale, il peu probable d'avoir une liste complète des enfants à échantillonner. Il est beaucoup plus courant pour les enquêtes d'utiliser les ménages (ou logements) en tant qu'une base de l'échantillonnage dans les groupes et d'échantillonner plutôt les ménages que les enfants. Les données sont ensuite recueillies sur les enfants admissibles résidant dans ces ménages échantillonnés. Cela signifie que le nombre des ménages et le nombre des enfants admissibles ne correspond pas toujours. Bien que certains ménages auront exactement un enfant admissible, d'autres auront plus d'un enfant admissible et certains n'auront aucun enfant admissible. Pour des raisons d'échantillonnage, il est donc essentiel d'avoir non seulement une estimation du nombre d'enfants admissibles qui doivent être échantillonnés, mais aussi une estimation du nombre de ménages à visiter afin d'obtenir l'échantillon requis d'enfants admissibles. (Il faut cependant noter que si le calcul de la taille d'échantillon est basé sur un indicateur de niveau des ménages, comme le Score de Diversité Alimentaire des Ménages ou l'Indice Domestique de la Faim, il n'y a aucun lien de correspondance, puisque pour ces indicateurs les ménages sont les seules unités d'échantillonnage. Dans ce cas, les facteurs d'inflation et de déflation décrits dans l'addendum ne sont pas requis.)

## 3. Faire monter la taille initiale requise pour l'échantillon afin de tenir compte des ménages sans enfants admissibles

Bien que l'exemple de calcul de la taille requise pour l'échantillon décrit ci-dessus est donné en fonction d'unités finales d'échantillonnage (par exemple : les enfants de moins de 59 mois, lorsqu'elles sont fondées sur des indicateurs anthropométriques liés au retard de croissance ou à l'insuffisance pondérale), on ne peut pas connaître la composition actuelle de l'âge des enfants dans un ménage échantillonné jusqu'à ce que le ménage soit contacté et cette information est obtenue grâce à un examen préalable. Ainsi, la taille réelle de l'échantillon des enfants qui sera atteinte en visitant un nombre fixe de ménages ne peut jamais être prédite avant le début du travail sur le terrain. Pour être sûr que la taille requise pour l'échantillon des enfants soit atteinte, avant de commencer le travail sur le terrain, on peut augmenter le nombre d'enfants à échantillonner d'une somme qui tient compte des ménages sans enfants admissibles.

La directive actuelle donnée dans le Guide d'Echantillonnage FANTA (Section 3.3.1) suggère d'augmenter la taille requise pour l'échantillon d'une somme égale à l'inverse de la moyenne estimée des enfants admissibles par ménage. Un exemple est donné dans la Section 3.3.1 du guide, où la taille requise pour l'échantillon des enfants est n = 300. Pour le pays concerné, la taille moyenne des ménages est de 6 et la proportion des enfants qui sont dans la tranche d'âge ciblée pour l'indicateur clé (moins de 24 mois dans l'exemple) n'est pas supérieure à 0,08 (équivalent à 8 pour cent)[1]. Ainsi, le nombre moyen estimé d'enfants de moins de 24 mois par ménage est de 60,08 ou 0,48. Nous référerons ce facteur, 0,48, le nombre moyen d'enfants dans le groupe d'âge souhaité par ménage, comme λ (« lambda »). Pour d'obtenir le nombre correct de ménages qui ont besoin d'être échantillonnés afin de s'assurer que la taille requise pour l'échantillon de 300 enfants soit atteinte, le guide propose de diviser n = 300 par λ = 0,48, pour obtenir une taille d'échantillon de n ÷ λ ménages. Dans cet exemple, nous calculons ainsi 300 ÷ 0,48 = 625 ménages.

Toutefois, l'expérience de terrain passée sur certaines enquêtes de base et évaluation finale menées par les programmes FFP / TII a montré que cette approche peut sous-estimer le nombre de ménages qui devraient être visités pour obtenir la taille requise pour l'échantillon des enfants. Cela a donné en retour des enquêtes qui n'atteignent pas la taille requise pour l'échantillon des enfants lors de la réalisation du travail de terrain. Dans ces circonstances, certains programmes FFP / TII ont opté pour l'augmentation du nombre de ménages échantillonnés et visités en utilisant des techniques d'échantillonnage « à la volée » non-basées sur les probabilités. De telles stratégies devraient être évitées et il est préférable de faire une estimation appropriée de la taille de l'échantillon de ménages dont on a vraiment besoin avant de mener le travail sur le terrain[2]. De telles techniques « à la volée » ont inclus la visite de ménages adjacents supplémentaires jusqu'à ce que la taille requise pour l'échantillon soit atteinte. Pour cette technique particulière, les ménages ne sont pas sélectionnés à l'aide d'un mécanisme aléatoire et par conséquent la technique n'est pas fondée sur la probabilité. Par conséquent, nous recommandons une approche alternative qui estimera plus précisément le nombre de ménages qui doivent être échantillonnés et visités afin de s'assurer que la taille requise pour l'échantillon d'enfants sera atteinte.

L'approche implique d'augmenter la taille requise pour l'échantillon par l'inverse de la proportion de ménages qui ont au moins un enfant admissible. Dans cet addendum nous proposons une approche alternative qui donnera lieu à un échantillon des ménages de taille plus importante que celle suggérée

---

[1] Les chiffres pour à la fois la taille moyenne des ménages et la proportion d'enfants dans la tranche d'âge ciblée sont généralement obtenues à partir du recensement national le plus récent ou de certaines autres enquêtes nationales ou internationalement sponsorisées.

[2] De telles techniques « à la volée » ont inclus la visite de ménages adjacents supplémentaires jusqu'à ce que la taille requise pour l'échantillon soit atteinte. Pour cette technique particulière, les ménages ne sont pas sélectionnés à l'aide d'un mécanisme aléatoire et par conséquent la technique n'est pas fondée sur la probabilité.

par le Guide d'Echantillonnage FANTA actuel, mais qui est plus certain de fournir la taille requise de l'échantillon des enfants.

L'approche alternative inclut un facteur d'inflation de la taille de l'échantillon qui est estimé avec la distribution de Poisson[3]. En utilisant cette distribution, on peut estimer que la proportion de ménages ayant au moins un enfant admissible (1, 2, 3 ou ...) est donnée par $1 - e^{-\lambda}$, alors qu'en moyenne il y a λ enfants admissibles par ménage. Ici, « e » désigne à la fonction exponentielle trouvée sur n'importe quelle calculatrice scientifique sous le symbole « exp » ou « e »x. Dans l'exemple ci-dessus avec λ = 0,48, nous avons $1 - e^{-\lambda} = 1 - e^{-0.48} = 0.38$. Ainsi, si n est la taille d'origine de l'échantillon comme calculée en utilisant la formule donnée dans le Guide d'Échantillonnage FANTA (et dans **l'Annexe 1**), alors une taille ajustée de l'échantillon qui tient compte de ce facteur d'inflation est donnée par :

$$\mathrm{n(1\_ajust\acute{e})} = \frac{\mathrm{n}}{(1 - e^{-\lambda})}.$$

Les détails techniques de la dérivation de n (1_ajusté) sont donnés dans **l'Annexe 2**. Dans l'exemple ci-dessus, rappelons que la taille d'origine requise pour l'échantillon des enfants, n = 300, a été ajustée à n ÷ λ = 300 ÷ 0,48 = 625 ménages, en utilisant le facteur d'inflation donné dans le Guide d'Echantillonnage FANTA. Toutefois, sous la nouvelle approche recommandée ici, la taille d'origine requise pour l'échantillon de n = 300 enfants est ajustée à plutôt

$$\frac{\mathrm{n}}{(1 - e^{-\lambda})} = \frac{300}{(1 - e^{-0.48})} = 786$$

ménages. En utilisant l'approche suggérée dans le Guide d'Echantillonnage FANTA, il est supposé que 325 (ou 625 − 300) des ménages échantillonnés n'auront pas d'enfants admissibles. D'autre part, sous la nouvelle approche préconisée ici, il est supposé que 486 (ou 786 − 300) des ménages échantillonnés n'auront pas d'enfants admissibles.

Du point de vue des opérations sur le terrain, l'équipe d'enquête peut examiner chaque ménage échantillonné (basé sur la taille gonflée de l'échantillon) pour les enfants admissibles à l'aide d'un fichier des ménages[4] obtenu grâce à une visite de contact initiale. Si aucun enfant admissible n'est trouvé, ces ménages peuvent être éliminés à des fins de collecte de données sur l'indicateur relatif aux enfants de la tranche d'âge ciblée. Toutefois, dans la plupart des cas, les données liées aux autres indicateurs relatifs aux autres membres du ménage et le ménage en général seront toujours recueillies auprès de ces ménages (par exemple, pour des indicateurs tels que l'Indice Domestique de la Faim et l'Indice de Diversité Alimentaire des Ménages).

---

[3] La distribution de Poisson est une distribution statistique discrète définie pour les entiers 0, 1, 2, 3 … qui donne une probabilité (ou une proportion) ou le chiffre de fois (0, 1, 2, …), une variable aléatoire apparaît, quand on sait qu'il apparait λ de fois.

[4] Un fichier des ménages est une liste de tous les membres du ménage, qui comprend habituellement les détails tels que le nom, l'âge, le sexe, la relation avec le chef du ménage et les autres informations démographiques pertinentes.

## 4. Décroître la taille ajustée de l'échantillon n (1_ajusté) pour tenir compte des ménages avec deux admissibles ou plus

Bien que l'approche ci-dessus estime le nombre de ménages sans enfants admissibles plus correctement, en comparaison avec l'estimation originale donnée dans le Guide d'Echantillonnage FANTA, la taille ajustée de l'échantillon, n (1_ajusté) ne tient pas compte du fait que certains ménages peuvent avoir deux enfants admissibles ou plus. Par ailleurs, les enquêtes peuvent opter pour la collecte d'informations sur la totalité ou un sous-échantillon d'un ou plusieurs enfants admissibles au sein d'un ménage échantillonné. Toutefois, pour l'enquête de base et l'évaluation finale effectuées par les programmes FFP / TII, il est fortement recommandé que la stratégie de sélection de tous les enfants admissibles au sein d'un ménage soit adoptée, plutôt que de sous-échantillonner un ou plusieurs enfants[5]. Compte tenu de cela, n (1_ajusté) doit être dégonflé légèrement pour tenir compte des ménages qui contribuent avec deux enfants ou plus dans la taille globale requise pour l'échantillon des enfants[6]. Une fois de plus, la distribution de Poisson est utilisée pour calculer le facteur de déflation requis et l'augmentation de la taille de l'échantillon dans la section précédente est utilisée comme un point de départ.

La formule pour l'ajustement de déflation est indiquée ci-dessous, où n (1_ajusté) est le résultat de l'augmentation antérieure de l'échantillon et n (2_ajusté) est le résultat de l'ajustement de déflation :

$$\mathrm{n(ajusté\_2)} = [\mathrm{A} * \mathrm{n(ajusté\_1)}] + \left[\frac{(1 - \mathrm{A}) * \mathrm{n(ajusté\_1)}}{2}\right]$$

où :

$$\mathrm{A} = (1 + \lambda) * e^{-\lambda}.$$

Les détails de la dérivation du n(2_ajusté) peuvent être retrouvés dans **l'Annexe 3**. Si on continue avec l'exemple ci-dessus où n(1_ajusté) = 786 et λ = 0.48, on obtient :

$$e^{-\lambda} = 0.62 \quad \text{and} \quad \mathrm{A} = (1 + 0.48) * 0.62 = 0.92$$

et finalement :

$$\mathrm{n(2\_ajusté)} = [0.92 * 786] + \left[\frac{(1 - 0.92) * 786}{2}\right] = 754 \text{ ménages.}$$

---

[5] Le principal avantage de la sélection de tous les enfants est d'éviter le sous-échantillonnage au sein des ménages. En faisant ainsi, il n'est pas nécessaire de calculer et d'appliquer un poids supplémentaire d'échantillonnage aux données lors de l'analyse afin de montrer cette étape supplémentaire de l'échantillonnage. (Voir la section 5.2 du Guide d'Echantillonnage FANTA pour plus de détails sur la pondération de l'échantillon.) Cet avantage est particulièrement pertinent pour les enquêtes polyvalentes, telles que celles menées par les programmes FFP / TII, où il y a souvent une tentative de collecter des données en faveur d'un certain nombre d'indicateurs, chacun ayant différents groupes d'âge cible (par exemple, les enfants de moins de 6 mois pour l'allaitement maternel exclusif, les enfants de 0 à 59 mois pour les retards de croissance et l'insuffisance pondérale, les enfants de 6 à 23 mois pour le régime minimal acceptable, etc.) Si une enquête devait sélectionner au hasard un enfant admissible par ménage dans chacun des groupes d'âge cible ci-dessus, il devrait y avoir un « poids de l'enfant » pris séparément pour chacun des indicateurs associés. Une telle stratégie d'échantillonnage serait trop complexe à gérer. La stratégie de sélectionner tous les enfants admissibles au sein d'un ménage échantillonné permet d'éviter cette situation.

[6] Il est important de noter que le facteur de déflation décrit dans cette section repose sur le respect rigoureux de la stratégie d'échantillonnage de tous les enfants admissibles dans un ménage échantillonné. Si au contraire, une stratégie de sous-échantillonnage des enfants admissibles dans un ménage échantillonné est adoptée, le calcul de la taille finale de l'échantillon des ménages ne sera pas exact.

L'ajustement de la déflation se traduit par une diminution de la taille de l'échantillon de 786 ménages à 754 ménages. Cela signifie qu'environ 786 − 754 = 32 ménages sont censés contribuer avec deux enfants ou plus dans l'échantillon d'enfants.

## 5. Augmentation de n(2_ajusté) pour tenir compte des ménages sans réponse

La dernière étape dans le calcul de la taille appropriée pour l'échantillon des ménages afin de garantir les données sur le nombre requis d'enfants admissibles est d'appliquer un facteur d'inflation final à la taille de l'échantillon des ménages pour tenir compte des ménages avec une non-réponse anticipée[7]. Comme indiqué dans le Guide d'Echantillonnage FANTA (Section 3.3.6), à moins qu'une information préalable relative au niveau des ménages sans réponse dans le pays ou la région du pays en question soit connue à partir d'enquêtes passées, un minimum de ménage sans réponse de 10 pour cent devrait être considéré et pris en compte dans le calcul de la taille de l'échantillon final. La formule pour l'ajustement de non-réponse est donnée ci-dessous[8]:

$$\mathrm{n(final)} = \{\mathrm{n(2\_ajusté)} \ + \ [\mathrm{n(2\_ajusté)} * (0.1)]\} = \mathrm{n(2\_ajusté)} * (1.1).$$

En continuant avec l'exemple ci-dessus, n (2_ajusté) = 754 est encore augmenté à n (final) = 754 * 1,1 = 829,4 ou 830 ménages. C'est le nombre de ménages sur lequel le plan d'échantillonnage pour la collecte des données devrait être fondé. En utilisant le même exemple, il est intéressant de comparer ces résultats avec ceux qui seraient obtenus sous la stratégie actuelle suggérée par le Guide d'Echantillonnage FANTA, où la taille ajustée de l'échantillon de 625 serait encore augmentée pour tenir compte des ménages non-réponse ainsi : 625 * 1.1 = 687,5 ou 688 ménages. Conformément à la stratégie actuelle, il est prévu que 688 − 625 = 63 ménages ne répondront pas à l'enquête. D'autre part, conformément à la nouvelle stratégie recommandée qui englobe l'ajustement des facteurs n (1_ajusté) et n (2_ajusté), il est prévu que 830 − 754 = 76 ménages ne répondront pas à l'enquête.

---

[7] Il est supposé qu'une certaine non-réponse résiduelle au niveau des ménages restera malgré tout effort concerté pour les contacter et mener des interviews dans tous les ménages échantillonnés. La non-réponse peut être due aux refus, les absences, les barrières linguistiques ou aux autres problèmes.

[8] Notez que le taux de 10 pour cent est considéré comme un minimum. Si l'on sait que le taux de non-réponse dans un pays particulier ou une région du pays est supérieur à 10 pour cent, plutôt le taux le plus élevé doit être utilisé dans la formule pour n (final).

**Sommaire des différents facteurs d'ajustement pour s'assurer du nombre correct de ménages à visiter afin que la taille requise pour l'échantillon d'enfants soit atteinte**

1. Calculez la taille initiale requise pour l'échantillon d'enfants (n) tel que prescrit dans le Guide d'Echantillonnage FANTA (1997) dans la Section 3.3.1 (et aussi dans **l'Annexe 1** de cet addendum).

2. Pour traduire ce nombre en nombre de ménages qui doivent être échantillonnés, augmenter n à n (1_ajusté) pour tenir compte de ménages qui n'ont pas d'enfants admissibles. La traduction du nombre requis d'enfants, n, en nombre requis de ménages, n (1_ajusté), est donnée par :

$$\mathrm{n(1\_ajust\acute{e})} = \frac{\mathrm{n}}{(1 - e^{-\lambda})}.$$

   Ici λ représente la moyenne d'enfants admissibles par ménage, calculée en multipliant la taille moyenne des ménages par la proportion d'enfants admissibles dans la population. Les chiffres relatifs à la fois à la taille moyenne des ménages et à la proportion d'enfants admissibles dans la tranche d'âge ciblée sont généralement obtenus à partir du plus récent recensement national ou de certaines autres enquêtes nationales ou internationalement sponsorisées.

3. Réduisez n (1_ajusté) à n (2_ajusté) pour tenir compte des ménages qui contribuent avec deux enfants admissibles ou plus dans l'échantillon.

$$\mathrm{n(2\_ajust\acute{e})} = [\mathrm{A} * \mathrm{n(1\_ajust\acute{e})}] + \left[\frac{(1 - \mathrm{A}) * \mathrm{n(1\_ajust\acute{e})}}{2}\right]$$

où :

$$\mathrm{A} = (1 + \lambda) * e^{-\lambda}.$$

   Notez que l'Etape 3 considère que la stratégie d'échantillonnage de *tous* les enfants dans un ménage échantillonné est utilisée.

   Enfin, augmentez n (2_ajusté) à n (final) pour tenir compte de ménages avec une non-réponse anticipée. En supposant un taux global de ménages avec une non-réponse de 10 pour cent, cela peut être calculé en utilisant :

$$\mathrm{n(final)} = \mathrm{n(2\_ajust\acute{e})} * (1.1).$$

Si l'on sait que le taux de non-réponse dans un pays particulier ou une région du pays est supérieur à 10 pour cent, plutôt le taux le plus élevé doit être utilisé dans la formule pour n (final).

En conclusion, bien que n (final) des ménages sont échantillonnés au départ, certains ne répondront pas, certains n'auront pas d'enfants admissibles et certains auront deux enfants admissibles ou plus.

Après avoir pris ces facteurs en compte à travers les ajustements approximatifs de la taille de l'échantillon des ménages, le nombre final d'entretiens réalisées avec des enfants admissibles doit être très proche de n.

**N.B :** Si λ «5 » ≥ «7 »1.5, il peut être démontré que les Etapes 2 et 3 combinées se traduiront par une «8 » déflation «14 » globale de la taille originale de l'échantillon n. Par conséquent, dans le cas où λ 1.5, les Etapes 2 et 3 devraient être supprimées et seuls les Etapes 1 et 4 devraient être appliquées.

# Annexe 1. Formule pour la Taille Initiale Requise pour l'Echantillon (des Enfants)

La formule pour la taille requise pour l'échantillon, n (des enfants, si le calcul est basé sur un indicateur anthropométrique) donnée dans le Guide d'Echantillonnage FANTA (Section 3.3.1) est la suivante :

$$n = \frac{D * (Z_{1-a} + Z_{1-b})^2 * [p_1 (1 - p_1) + p_2 (1 - p_2)]}{(p_2 - p_1)^2}$$

où :

| | | |
|---|---|---|
| n | = | taille requise de l'échantillon des enfants ; |
| D | = | effet du plan (nous considérons que D = pour la plupart des programmes FFP/TII) ; |
| p1 | = | a valeur de l'indicateur clé à la base (ou une valeur proxy), exprimée comme une proportion comprise entre 0 et 1 ; |
| p2 | = | la valeur cible prévue pour l'indicateur clé lors de l'évaluation finale, exprimée comme une proportion comprise entre 0 et 1 ; |
| Z1-a | = | le z-score correspondant au niveau de confiance désiré (généralement, nous fixons a = .05, ainsi Z0.95 = 1.645) ; et |
| Z1-b | = | le z-score correspondant au niveau de puissance désiré (généralement, nous fixons b = 0.2, ainsi Z0.8 = 0.840). |

Cette formule de la taille de l'échantillon est basée sur un test statistique de la différence de proportions (ou de la prévalence) pour un indicateur (par exemple, de l'enquête de base à l'évaluation finale), en contrôlant les erreurs d'inférence. Le test statistique est appliqué au moment de l'évaluation finale pour voir si les objectifs fixés par les programmes FFP / TII (en collaboration avec l'Agence des États-Unis pour le Développement International [USAID]) ont été atteints (bien que la réalisation peut ou ne peut pas être attribuable au programme). Par exemple, si le test est basé sur l'indicateur du retard de croissance, il est intéressant de voir s'il y a eu une baisse significative du point de vue de statistique des retards de croissance au cours de la durée du programme proportionnelle à l'objectif fixé au départ.

## Annexe 2. Dérivation de n(1_ajusté)

- Pour dériver le premier facteur d'inflation, nous utilisons la distribution de Poisson, qui est une distribution discrète définie pour les nombres entiers 0, 1, 2, 3, ... et qui donne la probabilité (ou la proportion), notée par le Pr, du nombre d'occurrences (x) d'un événement particulier (X), étant donné que l'on sait que le nombre moyen de fois où l'événement se produit est λ. La répartition se présente comme suit :

$$\Pr(\mathrm{X} = \mathrm{x}) = \frac{(e^{-\lambda}\lambda^{x})}{\mathrm{x}\,!} \qquad \mathrm{x} = 0, 1, 2, 3 \ldots \tag{1}$$

  où « x ! » est appelé « x factoriel » et est défini comme x ! = x ∗ (x − 1) ∗ (x − 2) ∗ … ∗ 1. Notez que 0 ! = 1.

- Si par exemple, nous définissons l'événement, X, comme étant « les enfants âgés de moins de 5 ans dans un ménage », et nous définissons λ comme étant « le nombre moyen d'enfants âgés de moins de 5 ans par ménage », alors la distribution de Poisson donne la probabilité (ou la proportion) du nombre d'enfants âgés de moins de 5 ans dans un ménage donné.

- Par exemple, si nous voulons la probabilité qu'il y ait 0 (ou pas) enfants de moins de 5 ans dans un ménage, en utilisant l'équation (1) avec x = 0 (et en notant que λ0 = 1), nous calculons :

$$\Pr(\mathrm{X} = 0) = \frac{(e^{-\lambda}\lambda^{0})}{0\,!} = e^{-\lambda}. \tag{2}$$

- En supposant que nous souhaitons réaliser des entretiens sur n enfants de moins de 5 ans, nous avons besoin de savoir combien de ménages il faut visiter, y compris ceux où il n'y a pas d'enfant d'âge admissible.

$$\Pr(\mathrm{X} > 0) = 1 - \Pr(\mathrm{X} = 0) = 1 - e^{-\lambda}. \tag{3}$$

- Pour obtenir le nombre de ménages à visiter, nous devrions augmenter n (la taille de l'échantillon calculée pour les enfants âgés de moins de 5 ans) par l'inverse de la proportion donnée dans l'équation (3). Par conséquent, nous avons :

$$\mathrm{n(1\_ajust\acute{e})} = \frac{\mathrm{n}}{(1 - \mathrm{e}^{-\lambda})}. \tag{4}$$

**N.B :** La distribution de Poisson répand les probabilités à travers les points de masse qui vont de la valeur de 0 à l'infini. Cependant, il n'y a pas de nombre infini d'enfants âgés de moins de 5 ans dans un ménage. Par conséquent, pour être le plus correct du point de vue technique, cette dérivation doit être fondée sur une distribution de « Poisson tronquée » qui ne permet pas d'avoir les valeurs supérieures à un certain nombre raisonnable d'enfants d'âge admissible par ménage (par exemple, 5) et cela définit la distribution des valeurs discrètes 0, 1, 2, 3, 4 et 5 seulement. Toutefois, il est possible de démontrer que, pour les petites valeurs de λ (prenons λ <1,5), Pr (X> 5) est proche de 0 et est donc est négligeable. Ainsi, il a été jugé que la précision ajoutée grâce à l'utilisation de la distribution de Poisson tronquée ne justifiait pas la complexité supplémentaire dans la formule. Par conséquent, la répartition de Poisson habituelle a été utilisée à la place de la distribution de Poisson tronquée dans la dérivation ci-dessus.

## Annexe 3. Dérivation de n(2_ajusté)

- La taille ajustée de l'échantillon, n (1_ajusté), de l'Annexe 2 donne le nombre de ménages à échantillonner pour atteindre la taille requise pour l'échantillon des enfants, n, en tenant compte des ménages sans enfants admissibles. Par conséquent, n (1_ajusté) comprend les ménages qui ont exactement un enfant d'âge admissible. Mais n (1_ajusté) comprend aussi les ménages avec deux ou plus enfants d'âge admissible. Dans les cas où un seul enfant d'âge admissible est échantillonné par ménage, ce dernier groupe ne serait pas un souci. Toutefois, Les programmes FFP / TII sont invités à échantillonner tous les enfants d'âge admissible dans un ménage sélectionné et sont donc susceptibles d'atteindre la taille globale désirée pour l'échantillon d'enfants, n, en visitant moins de n (1_ajusté) de ménages. C'est parce que certains ménages inclus dans n (1_ajusté) contiendront deux ou plus enfants d'âge admissible, qui seront tous échantillonnés. Afin de tenir compte des ménages avec deux enfants d'âge admissible ou plus, il faut réduire n (1_ajusté) en conséquence.

- Afin de calculer le déflateur de cela, nous utilisons l'équation (1) avec x = 1 et il faut noter que :

$$\Pr(\mathrm{X}=1)=\frac{(e^{-\lambda}\lambda^{1})}{1\,!}=\lambda e^{-\lambda}. \tag{5}$$

  Par ailleurs :

$$\Pr(\mathrm{X}\ \geq 2)=1\ -\Pr(\mathrm{X}=0)-\Pr(\mathrm{X}=1)=1\ -\ \mathrm{e}^{-\lambda}\ -(\ \lambda\mathrm{e}^{-\lambda}) \tag{6}$$

  en utilisant les équations (2) et (5). En combinant les conditions, nous avons :

$$\Pr(\mathrm{X}\ \geq 2)=1\ -\left[(1\ +\ \lambda)*\ \mathrm{e}^{-\lambda}\right]\ =1-\mathrm{A} \tag{7}$$

  où :

$$\mathrm{A}\ =(1\ +\ \lambda)*\ e^{-\lambda}. \tag{8}$$

  En utilisant les équations (6) et (7), il est utile de noter que :

$$\mathrm{A} = \Pr(\mathrm{X} = 0) + \Pr(\mathrm{X} = 1). \tag{9}$$

- Ensuite, nous utilisons la tautologie :

$$1 = \Pr(\mathrm{X} = 0) + \Pr(\mathrm{X} = 1) + \Pr(\mathrm{X} \geq 2). \tag{10}$$

  L'équation (10) est vraie car la somme de la distribution de Poisson (ou de toute autre distribution discrète) à travers toutes les valeurs possibles est égal à 1. En utilisant les équations (2), (5) et (7), nous pouvons voir que l'équation (10) peut être réécrite ainsi :

$$1=\mathrm{e}^{-\lambda}\ +(\lambda\mathrm{e}^{-\lambda})+\left[1\ -((1\ +\ \lambda)*\mathrm{e}^{-\lambda})\right]\ =\left[(1\ +\ \lambda)*\mathrm{e}^{-\lambda}\right]\ +\ \left[1\ -((1\ +\ \lambda)*\mathrm{e}^{-\lambda})\right]. \tag{11}$$

- Pour décomposer n (1_ajusté) en composantes appropriées relatives aux différentes compositions des ménages, on multiplie chaque terme de l'équation (11) par n(1_ajusté) et on obtient :

$$\begin{aligned}\mathrm{n(1\_ajust\acute{e})} &= \left[(1\ +\ \lambda)*\ \mathrm{e}^{-\lambda})\right]*\mathrm{n(1\_ajust\acute{e})}+\left[1-\left((1\ +\ \lambda)*\ \mathrm{e}^{-\lambda}\right)\right]*\mathrm{n(1\_ajust\acute{e})}\\ &= \left[\mathrm{A}*\mathrm{n(1\_ajust\acute{e})}\right]\ +\ \left[(1\ -\mathrm{A})*\mathrm{n(1\_ajust\acute{e})}\right].\end{aligned} \tag{12}$$

  Nous obtenons cette dernière expression en appliquant la définition de A donnée dans l'équation (8).

L'équation (12) rompt essentiellement n (1_ajusté) en une somme composite ayant deux parties données par A * n(1_ajusté) et (1 − A) n(1_ajusté).

De l'équation (9) ci-dessus, nous voyons que A = Pr (X = 0) + Pr (X = 1) et ainsi la première partie de l'équation (12) représente le nombre de ménages à visiter qui contiennent soit aucun, soit un seul enfant d'âge admissible.

De l'équation (7) ci-dessus, nous voyons que 1 − A = Pr(X ≥ 2) et ainsi la deuxième partie de l'équation (12) représente le nombre de ménages à visiter qui contiennent deux ou plus enfants d'âge admissible.

- L'objectif du déflateur est de réduire le nombre de ménages qui proviennent de la deuxième partie. Par conséquent, nous souhaitons diminuer de une moitié le nombre de ménages comptant deux enfants d'âge admissible et de réduire au  tiers le nombre de ménages comptant trois enfants d'âge admissible et ainsi de suite. Toutefois, nous pouvons supposer que le nombre de ménages ayant trois enfants d'âge admissible ou plus est négligeable, relativement parlant. Donc, par souci de simplicité, nous « les regroupons » avec les ménages ayant deux enfants d'âge admissible. Ce qu'on entend par « groupement », c'est que nous ne diminuons pas, par exemple, d'un tiers, le nombre de ménages ayant trois enfants d'âge admissible, parce que la complexité ajoutée au calcul ne vaut pas la différence négligeable que ça ferait. Au lieu de cela, nous diminuons le nombre de ces ménages jusqu'à la moitié. De même, nous diminuons jusqu'à la moitié le nombre de ménages ayant quatre enfants d'âge admissible.

- Ainsi, nous créons un nouvel ajustement appelé n(2_ajusté), où nous réduisons de moitié la deuxième partie de l'équation (12) :

$$\mathrm{n(2\_ajusté)} = [\mathrm{A} * \mathrm{n(1\_ajusté)}] + \frac{[(1 - \mathrm{A}) * \mathrm{n(1\_ajusté)}]}{2}. \tag{13}$$

- **N.B.** Ni n(1_ajusté), ni n(2_ajusté) ne devraient être utilisés pour les valeurs de λ ≥ 1,5.

  Sinon, il y aura une déflation globale de la taille initiale de l'échantillon n. Dans le scénario où λ ≥ 1,5, il est recommandé d'utiliser simplement n, la taille initiale de l'échantillon d'enfants et d'appliquer l'ajustement pour les ménages sans réponse anticipée, mentionné précédemment, mais d'omettre à la fois le facteur d'ajustement n(1_ajusté) et le facteur d'ajustement n(2_ajusté).