FANTA III
FOOD AND NUTRITION TECHNICAL ASSISTANCE

USAID
FROM THE AMERICAN PEOPLE

Addendum to FANTA Sampling Guide by Robert Magnani (1997): Correction to Section 3.3.1 Determining the Number of Households That Need to be Contacted

Diana Stukel
Megan Deitchler

March 2012

fhi360
THE SCIENCE OF IMPROVING LIVES

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

This Addendum is made possible by the generous support of the American people through the support of the Office of Food for Peace, Bureau for Democracy, Conflict, and Humanitarian Assistance, and the Office of Health, Infectious Diseases, and Nutrition, Bureau for Global Health, United States Agency for International Development (USAID), under terms of Cooperative Agreement No. AID-OAA-A-11-00014, through the FANTA Project, managed by FHI 360.

The contents are the responsibility of FHI 360 and do not necessarily reflect the views of USAID or the United States Government.

Published March 2012

Recommended Citation:

Stukel, Diana; and Deitchler, Megan. 2012. *Addendum to FANTA Sampling Guide by Robert Magnani (1999): Correction to Section 3.3.1 Determining the Number of Households That Need to be Contacted.* Washington, DC: FHI 360/ FANTA.

Contact information:

Food and Nutrition Technical Assistance III Project (FANTA)
FHI 360
1825 Connecticut Avenue, NW
Washington, D.C. 20009-5721
Tel: 202-884-8000
Fax: 202-884-8432
Email: fantamail@fhi360.org
Website: www.fantaproject.org

Table of Contents

1. Introduction 1
2. Calculating the initial required sample size 1
3. Inflating the initial required sample size to account for households with no eligible children 2
4. Deflating the adjusted sample size, n(adjusted_1), to account for households with two or more eligible children 3
5. Inflating n(adjusted_2) to account for household non-response 4

Appendix 1. Formula for Initial Required Sample Size (of Children) 5
Appendix 2. Derivation of n(adjusted_1) 7
Appendix 3. Derivation of n(adjusted_2) 8

Acknowledgments

The authors wish to express their appreciation for the contributions of the following members of the FHI 360/FANTA staff: Sandy Remancus, Gilles Bergeron, Vicky Michener, and Laura Glaeser.

1. Introduction

The FANTA Sampling Guide provides guidance on how to calculate the sample size for baseline and final evaluation surveys conducted by Food for Peace Title II (FFP/TII) programs. Typically, the sample size calculation is driven by an anthropometric indicator, such as that relating to stunting or underweight, which requires collecting data on children under 5 years of age. In such cases, the initial sample size calculated reflects the number of children in this age category for which data are required. However, most surveys use households (or dwellings) rather than children as the basis for sampling within clusters. Therefore, it is most typical for households to be sampled first and then for data to be collected on the eligible children residing in the sampled households. This creates a challenge for sample size calculation, because the correspondence between households and children is not always one-to-one.

To overcome this problem, the FANTA Sampling Guide suggests an approach (see Section 3.3.1) to translate the calculated child-level sample size into the number of households that need to be visited to ensure that the required number of sampled children is achieved. However, past field experience for some baseline and final evaluation surveys conducted by FFP/TII programs has shown that the approach given in the FANTA Sampling Guide can sometimes result in an underestimation of the number of households that should be visited. In this addendum, we recommend an alternative approach that will result in a household sample size that is greater than that suggested by the current FANTA Sampling Guide, but is more likely to result in the required sample size for children being achieved.

2. Calculating the initial required sample size

The formula for the required sample size, n (for children, if the calculation is based on an anthropometric indicator), given in Section 3.3.1 of the FANTA Sampling Guide is also given in **Appendix 1** of this addendum. The sample size that results from this formula reflects the number of units on which data for that indicator should be collected in both the baseline and final evaluation surveys. Therefore, if the indicator used for the sample size calculation is “prevalence of stunting among children 0–59 months of age,” the sample size calculated from the formula would reflect the required number of children 0–59 months of age that should be sampled in both the baseline and final evaluation surveys.

However, when carrying out a baseline or final evaluation survey, it is rare to have a complete list of children from which to sample. It is much more typical for surveys to use households (or dwellings) as the basis of sampling within clusters and to sample households rather than children. Data are then collected on the eligible children residing in those sampled households. This means that the correspondence between households and eligible children is not always one-to-one. Although some households will have exactly one eligible child, other households will have more than one eligible child and some household will have no eligible children at all. For sampling purposes, it is essential therefore to have not only an estimate of the number of eligible children that must be sampled, but also an estimate of the number of households that need to be visited to obtain the required sample of eligible children. (Note, however, that if the sample size calculation is based on a household level indicator, such as the Household Dietary Diversity Score or the Household Hunger Scale, there is no correspondence issue since, for these indicators, households are the ultimate sampling units. In this case, the inflation and deflation factors described in this addendum are not required.)

3. Inflating the initial required sample size to account for households with no eligible children

Although the required sample size calculation described above is given in terms of ultimate sampling units (e.g., children under 59 months of age, when based on anthropometric indicators related to stunting or underweight), one cannot know the current age composition of children in a sampled household until the household is contacted and such information is obtained through screening. Thus, the actual sample size of children that will be achieved by visiting a fixed number of households can never be predicted prior to the commencement of fieldwork. One way to help ensure that the required sample size for children will be met in advance of fieldwork is to inflate the number of children to be sampled by an amount that accounts for households with no eligible children.

The current guidance given in the FANTA Sampling Guide (Section 3.3.1) suggests inflating the required sample size by an amount equal to the inverse of the estimated *average* number of eligible children per household. An example is given in Section 3.3.1 of the guide, where the required sample size of children is n = 300. For the country in question, the average household size is 6 and the proportion of children in the population in the target age group for the key indicator (under 24 months in the example) is no more than 0.08 (equivalent to 8 percent).[1] Thus, the estimated average number of children under 24 months of age per household is $6 * 0.08$ or 0.48. We will refer to this factor, 0.48, the average number of children of the desired age group per household, as λ ("lambda"). To obtain the appropriate number of households that need to be sampled to ensure the required sample size of 300 children is achieved, the guide suggests dividing n = 300 by λ = 0.48, to obtain a sample size of n ÷ λ households. In this example, we calculate this as 300 ÷ 0.48 = 625 households.

However, past field experience for some baseline and final evaluation surveys conducted by FFP/TII programs has shown that this approach can underestimate the number of households that should be visited to obtain the required sample size of children. This in turn has resulted in surveys falling short of achieving the required sample size of children while conducting fieldwork. In such instances, some FFP/TII programs have opted to augment the number of households sampled and visited using "on the fly" non-probability-based sampling techniques.[2] Such strategies should be avoided, and it is preferable to appropriately approximate the sample size of households truly needed in advance of conducting fieldwork.

Therefore, we recommend an alternative approach that will more closely approximate the number of households that need to be sampled and visited to ensure the required child sample size. The approach involves inflating the required sample size by the inverse of the *proportion* of households that have at least one eligible child. This approach will result in a household sample size that is greater than that suggested in the FANTA Sampling Guide, but is more likely to result in the required sample size for children being achieved.

The alternative approach involves a sample size inflation factor that is approximated using the Poisson distribution.[3] Using this distribution, it can be estimated that the proportion of households having at least one eligible child (1 or 2 or 3 or …) is given by $1 - e^{-\lambda}$, when on average there are λ eligible children per household. Here, "e" refers to the exponential function, found on any scientific hand calculator under the symbol "exp" or "e^x." In the above example with λ = 0.48, we have that $1 - e^{-\lambda} = 1 - e^{-0.48} = 0.38$. Thus, if n is the original required sample size as calculated using the formula given in the FANTA

[1] Figures for both the average household size and the proportion of children in the target age group are typically obtained from the most recent national census or from some other national or internationally sponsored survey.

[2] Such "on the fly" techniques have included visiting additional adjacent households until the required sample size is achieved. For this particular technique, households are not drawn using a random mechanism and therefore the technique is not probability-based.

[3] The Poisson distribution is a discrete statistical distribution defined for integers 0, 1, 2, 3 … that gives the probability (or proportion) of the number of times (0,1, 2, …) a random variable occurs, when it is known to occur an average of λ times.

Sampling Guide (and in **Appendix 1**), then an adjusted sample size that takes into account this inflation factor is given by:

$$\mathrm{n(adjusted_1)} = \frac{\mathrm{n}}{(1 - e^{-\lambda})}.$$

The technical details of the derivation of n(adjusted_1) are given in **Appendix 2**. In the above example, recall that the original required sample size of children, n = 300, was adjusted to n ÷ λ = 300 ÷ 0.48 = 625 households using the inflation factor given in the FANTA Sampling Guide. However, under the new approach recommended here, the original required sample size of n = 300 children is adjusted to

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

households instead. Using the approach suggested in the FANTA Sampling Guide, it is assumed that 325 (or 625 − 300) of the households sampled will have no eligible children. On the other hand, under the new approach recommended here, it is assumed that 486 (or 786 − 300) of the households sampled will have no eligible children.

From a field operations point of view, the survey team can screen each sampled household (based on the inflated sample size) for eligible children using a household roster[4] obtained through an initial contact visit. If no eligible children are found within, these households can be screened out for the purposes of collecting data on the indicator relating to children of the target age group. However, in most instances, data supporting other indicators relating to other household members and the household in general will still be collected from these households (e.g., for such indicators as the Household Hunger Scale and the Household Dietary Diversity Scale).

4. Deflating the adjusted sample size, n(adjusted_1), to account for households with two or more eligible children

Although the above approach more correctly approximates the number of households with no eligible children relative to the original approximation given in the FANTA Sampling Guide, the adjusted sample size, n(adjusted_1), does not account for the fact that some households may have two or more eligible children. Furthermore, surveys can opt to collect information on either all or a sub-sample of one or more eligible children within a sampled household. However, for the baseline and final evaluation surveys conducted by FFP/TII programs, it is strongly recommended that the strategy of selecting *all* eligible children within a household be adopted, rather than sub-sampling one or more such children.[5] In light of this, n(adjusted_1) should be deflated slightly to account for households that contribute two or more children toward the overall required sample size of children.[6] Once again, the Poisson distribution is used to approximate the required deflation factor, and the sample size inflation from the previous section is

[4] A household roster is a listing of all household members, and typically includes details such as name, age, sex, relationship to head of household, and other relevant demographic information.

[5] The main advantage of selecting all children is to avoid sub-sampling within households. In doing so, there is no need to calculate and apply an additional sampling weight to the data during analysis to reflect this additional stage of sampling. (See Section 5.2 of the FANTA Sampling Guide for more details on sample weighting.) This advantage is particularly relevant for multipurpose surveys, such as those conducted by FFP/TII programs, where there is often an attempt to collect data in support of a number of indicators, each having different target age groups (for example, children under 6 months for exclusive breastfeeding, children aged 0–59 months for stunting and underweight, children aged 6–23 months for minimum acceptable diet, etc.). If a survey were to randomly select one eligible child per household within each of the above target age groups, there would need to be a separate “child weight” associated with each of the associated indicators. Such a sampling strategy would be overly complex to manage. The strategy of selecting all eligible children within a sampled household helps avoid this situation.

[6] It is important to note that the deflation factor described in this section relies on strict adherence to the strategy of sampling *all* eligible children in a sampled household. If, instead, a strategy of sub-sampling eligible children in a sampled household is adopted, then the final sample size calculation of households will not be accurate.

used as a starting point. The formula for the deflation adjustment is shown below, where n(adjusted_1) is the result of the earlier sample size inflation and n(adjusted_2) is the result of the deflation adjustment:

$$\mathrm{n(adjusted_2)} = [\mathrm{A} * \mathrm{n(adjusted_1)}] + \left[\frac{(1 - \mathrm{A}) * \mathrm{n(adjusted_1)}}{2}\right]$$

where:

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

The details of the derivation of n(adjusted_2) can be found in **Appendix 3**. Continuing the example from above where n(adjusted_1) = 786 and λ = 0.48, we obtain:

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

and finally:

$$\mathrm{n(adjusted_2)} = [0.92 * 786] + \left[\frac{(1 - 0.92) * 786}{2}\right] = 754 \text{ households.}$$

The deflation adjustment results in the sample size decreasing from 786 households to 754 households. This means that approximately 786 − 754 = 32 households are expected to contribute two or more children to the sample of children.

5. Inflating n(adjusted_2) to account for household non-response

The final step in calculating the appropriate household sample size to ensure data on the required number of eligible children is to apply a final inflation factor to the household sample size to account for anticipated household non-response.[7] As discussed in the FANTA Sampling Guide (Section 3.3.6), unless prior information from past surveys is known in advance of fieldwork regarding the household non-response level in the country or region of the country in question, a minimum household non-response of 10 percent should be assumed and accounted for in the final sample size calculation. The formula for the non-response adjustment is given below[8]:

$$\mathrm{n(final)} = \{\mathrm{n(adjusted_2)} + [\mathrm{n(adjusted_2)} * (0.1)]\} = \mathrm{n(adjusted_2)} * (1.1).$$

Continuing with the same example above, n(adjusted_2) = 754 is further inflated to n(final) = 754 * 1.1 = 829.4, or 830 households. This is the number of households on which the sample design for data collection should be based. Using the same example, it is interesting to contrast these results with those that would be obtained under the current strategy suggested in the FANTA Sampling Guide, where the adjusted sample size of 625 would be further inflated to account for household non-response as 625 * 1.1 = 687.5 or 688 households. Under the current strategy, it is anticipated that 688 − 625 = 63 households will not respond to the survey. On the other hand, under the recommended new strategy that encompasses the adjustment factors n(adjusted_1) and n(adjusted_2), it is anticipated that 830 − 754 = 76 households will not respond to the survey.

[7] It is assumed that some residual non-response at the household level will remain despite any concerted effort to contact and conduct interviews in all sampled households. The non-response may be due to refusals, absences, language barriers, or other issues.

[8] Note that the rate of 10 percent is considered a minimum. If it is known that the non-response rate in a particular country or region in the country is higher than 10 percent, the higher rate should be used in the formula for n(final) instead.

Summary of various adjustment factors to ensure the appropriate number of households are visited in order that the required sample size of children is achieved

1. Calculate the initial sample size of children required (n) as prescribed in the FANTA Sampling Guide (1997) in Section 3.3.1 (and also given in **Appendix 1** of this addendum).

2. To translate this number to the number of households that need to be sampled, inflate n to n(adjusted_1) to account for households that have no eligible children. The translation from number of required children, n, to number of required households, n(adjusted_1), is given by:

$$\mathrm{n(adjusted_1)} = \frac{\mathrm{n}}{(1 - e^{-\lambda})}.$$

   Here λ represents the average number of eligible children per household, calculated by computing the average household size multiplied by the proportion of eligible children in the population. Figures for both the average household size and the proportion of eligible children are typically obtained from the most recent national census or some other national or internationally sponsored survey.

3. Deflate n(adjusted_1) to n(adjusted_2) to account for households that contribute two or more eligible children to the sample. This adjustment is made using the formula:

$$\mathrm{n(adjusted_2)} = [\mathrm{A} * \mathrm{n(adjusted_1)}] + \left[\frac{(1 - \mathrm{A}) * \mathrm{n(adjusted_1)}}{2}\right]$$

   where:

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

   Note that Step 3 assumes that the strategy of sampling *all* children in a sampled household is employed.

4. Finally, inflate n(adjusted_2) to n(final) to account for anticipated household non-response. Assuming an overall household non-response rate of 10 percent, this can be calculated using:

$$\mathrm{n(final)} = \mathrm{n(adjusted_2)} * (1.1).$$

   If it is known that the non-response rate in a particular country or region of the country is higher than 10 percent, the higher rate should be used in the formula for n(final) instead.

   In conclusion, although n(final) households are sampled initially, some will not respond, some will have no eligible children, and some will have two or more eligible children. After taking these factors into account through approximate adjustments to the household sample size, the final number of completed interviews with eligible children should be very close to n.

   Note: If λ ≥ 1.5, it can be shown that Steps 2 and 3 combined will result in an overall *deflation* from the original sample size n. Therefore, in the case where λ ≥ 1.5, Steps 2 and 3 should be omitted and only Steps 1 and 4 should be applied.

Appendix 1. Formula for Initial Required Sample Size (of Children)

The formula for the required sample size, n (of children, if the calculation is based on an anthropometric indicator) given in the FANTA Sampling Guide (Section 3.3.1) is:

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

where:

- n = required sample size of children;
- D = design effect (we assume D = 2 for most FFP/TII programs);
- p_1 = the value of the key indicator at baseline (or a proxy value), expressed as a proportion between 0 and 1;
- p_2 = the planned target value of the key indicator at the end-line/final evaluation, expressed as a proportion between 0 and 1;
- Z_{1-a} = the z-score corresponding to the desired confidence level (typically, we set a = .05, thus $Z_{0.95}$ = 1.645); and
- Z_{1-b} = the z-score corresponding to the desired power level (typically, we set b = 0.2, thus $Z_{0.8}$ = 0.840).

This sample size formula is based on a statistical test of the difference of proportions (or prevalence) for an indicator (e.g., from baseline to final evaluation), controlling for inferential error. The statistical test is applied at the time of the final evaluation to see if the targets set by FFP/TII programs (in collaboration with the United States Agency for International Development [USAID]) have been achieved (although the achievement may or may not be attributable to the program). For instance, if the test is based on the stunting indicator, it is of interest to see if there has been a statistically significant drop in stunting over the duration of the program commensurate with the target set at baseline.

Appendix 2. Derivation of n(adjusted_1)

- To derive the first inflation factor, we use the Poisson distribution, which is a discrete distribution defined for integers 0, 1, 2, 3, … and which gives the probability (or proportion), denoted by Pr, of the number of occurrences (x) of a particular event (X), given that it is known that the *average* number of times the event occurs is λ. The distribution looks as follows:

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

  where "x!" is called "x factorial" and is defined as x! = x ∗ (x − 1) ∗ (x − 2) ∗ … ∗ 1. Note that 0! = 1.

- If, for example, we define the event, X, as "children under 5 years of age in a household," and we define λ as "the average number of children under 5 years of age per household," then the Poisson distribution gives the probability (or proportion) of the number of children under 5 years of age in a given household.

- For example, if we want the probability that there are 0 (or no) children under 5 years of age in a household, using equation (1) with x = 0 (and noting that $\lambda^0 = 1$), we compute:

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

- Assuming that we wish completed interviews on n children under 5 years of age, we need to know how many households to visit, including those where there are no children of eligible age. Therefore, we wish to know the probability (or proportion) of households that will have *at least one* (i.e., one or more) child under 5 years of age. Using equation (2), we can see that this is given by:

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

- To obtain the number of households to visit, we should inflate n (the sample size calculated for children under 5 years of age) by the inverse of the proportion given in equation (3). Therefore, we have:

$$\mathrm{n(adjusted_1)} = \frac{\mathrm{n}}{(1 - \mathrm{e}^{-\lambda})}. \tag{4}$$

- **Note:** The Poisson distribution spreads probabilities across mass points that range in value from 0 to infinity. However, there are not an infinite number of children under 5 years of age within a household. Therefore, to be most technically correct, this derivation should be based on a "truncated Poisson" distribution that does not permit values greater than some reasonable number of children of eligible age per household (e.g., 5) and that defines the distribution for discrete values 0, 1, 2, 3, 4, and 5 only. However, it can be shown that, for small values of λ (say, λ < 1.5), Pr (X > 5) is close to 0 and thus is negligible. So, it was deemed that the added accuracy in using the truncated Poisson distribution did not warrant the additional complexity in the formula. Therefore, the usual Poisson distribution was used instead of the truncated Poisson distribution in the above derivation.

Appendix 3. Derivation of n(adjusted_2)

- The adjusted sample size, n(adjusted_1), from Appendix 2 gives the number of households to sample to achieve the required sample size of children, n, taking into account households with no eligible children. Therefore, n(adjusted_1) includes households that have exactly one child of eligible age. But n(adjusted_1) also includes households with two or more children of eligible age. In cases where only one child of eligible age is sampled per household, this latter group would not be a concern. However, FFP/TII programs are advised to sample *all* children of eligible age within a selected household, and therefore are likely to achieve the overall desired sample size of children, n, by visiting fewer than n(adjusted_1) households. This is because some households included in n(adjusted_1) will contain two or more children of eligible age, all of whom will be sampled. To account for households with two or more children of eligible age, we must deflate n(adjusted_1) accordingly.

- To derive the deflator for this, we use equation (1) with x = 1 and note that:

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

Furthermore:

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

using equations (2) and (5). Combining terms, we have:

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

where:

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

Using equations (6) and (7), it is useful to note that:

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

- Next, we use the tautology:

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

Equation (10) is true because the sum of the Poisson (or any other discrete) distribution across all possible values is equal to 1. Using equations (2), (5) and (7), we can see that equation (10) can be rewritten as:

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

- To decompose n(adjusted_1) into appropriate component pieces relating to the differing household compositions, we multiply each term in equation (11) by n(adjusted_1) and obtain:

$$\begin{aligned} \mathrm{n(adjusted_1)} &= \left[(1 + \lambda) * \mathrm{e}^{-\lambda})\right] * \mathrm{n(adjusted_1)} + \left[1 - \left((1 + \lambda) * \mathrm{e}^{-\lambda}\right)\right] * \mathrm{n(adjusted_1)} \\ &= [\mathrm{A} * \mathrm{n(adjusted_1)}] + [(1 - \mathrm{A}) * \mathrm{n(adjusted_1)}]. \end{aligned} \tag{12}$$

We obtain this last expression by applying the definition of A given in equation (8).

Equation (12) essentially breaks n(adjusted_1) into a composite sum with two component parts given by A ∗ n(adjusted_1) and (1 − A) ∗ n(adjusted_1).

From equation (9) above, we see that A = Pr(X = 0) + Pr(X = 1), and so the first component part of equation (12) represents the number of households to be visited that contain either no children or one child of eligible age.

From equation (7) above, we see that 1 − A = Pr(X ≥ 2), and so the second component part of equation (12) represents the number of households to be visited that contain two or more children of eligible age.

- The aim of the deflator is to reduce the number of households that comes from the second component. We therefore wish to diminish to a half the number of households containing two children of eligible age and to diminish to a third the number of households containing three children of eligible age, and so on. However, we can assume that the number of households having three or more children of eligible age is negligibly small, relatively speaking. Therefore, for simplicity, we "bundle them" with households having two children of eligible age. What is meant by "bundling" is that we do not, for instance, diminish to a third the number of households having three children of eligible age, because the added complexity of the computation is not worth the negligible difference this would make. Instead, we diminish the number of such households to a half. Similarly, we diminish to a half the number of households having four children of eligible age.

- Thus, we create a new adjustment called n(adjusted_2), where we halve the second component of equation (12):

$$\mathrm{n(adjusted_2)} = [\mathrm{A} * \mathrm{n(adjusted_1)}] + \frac{[(1 - \mathrm{A}) * \mathrm{n(adjusted_1)}]}{2}. \tag{13}$$

- **Note:** Neither n(adjusted_1) nor n(adjusted_2) should be used for values of λ ≥ 1.5. Otherwise, there will be an overall *deflation* from the initial sample size of n. In the scenario where λ ≥ 1.5, it is recommended to simply use n, the initial sample size of children and to apply the adjustment for anticipated household non-response discussed earlier, but to omit both adjustment factors n(adjusted_1) and n(adjusted_2).